Agent-Relativity and the Foundations of Moral Theory

Matthew Hammerton

May 2017

A thesis submitted for the degree of Doctor of Philosophy of the Australian National University

Acknowledgments

There are many people who have made contributions to this thesis that I would like to acknowledge. First, I would like to thank Daniel Nolan, the Chair of my supervisory panel. Daniel has been the ideal supervisor in several respects. He has consistently provided incisive critical feedback on my work, pushing me to make my arguments stronger and express my views with greater clarity. He has also offered excellent strategic advice, helping me decide what direction to take my project and how to deal with various challenges. Finally, he has been very supportive, offering sympathetic understanding when I have faced personal setbacks, and helping me grow my professional confidence.

Second, I thank Nic Southwood who has taken up the unofficial role of 'vice-captain' on my supervisory panel. Like Daniel, Nic has contributed to this project with incisive critical feedback and invaluable strategic advice. His enthusiasm for my project has been very encouraging, and I am also grateful for his help navigating several complex professional decisions.

Third, I thank Ryan Cox who, although himself only a PhD candidate, has acted as a de facto member of my supervisory panel, reading all my work and providing sharp critical feedback that has pushed me to better develop my ideas. Ryan and I had the good fortune of being asked to share an office shortly after commencing our PhD programs, and our similar philosophical interests and temperament have led to a fruitful friendship.

I am also grateful to Philip Pettit for his supervisory support. Philip's part time status at ANU (being here for only four months each year) has limited how much supervision he has been able to offer. However, when he has had the opportunity to engage with my work he has been tremendously helpful, combining astute critical feedback and strategic nous with a generosity of spirit that is both encouraging, and provides an excellent ideal to aspire to when engaging philosophically with others.

I would also like to thank the two other members of my supervisory panel, Daniel Stoljar and Seth Lazar, who have aided me by reading some of my work and offering their feedback.

In addition to those I have acknowledge above, a long list of philosophers have assisted me by giving feedback on my work via conferences, seminars, email correspondence, and casual chats over coffee. This list includes Jamin Asay, Derek Baker, Alma Barner, Christian Barry, Ben Blumson, Rachel Briggs, Mark Colyvan, Garrett Cullity, Tom Dougherty,

Abstract

This thesis is concerned with agent-relativity and its importance in understanding and evaluating moral theories. Many would agree with Thomas Hurka that the distinction between agent-relativity and agent-neutrality is one of "the greatest contributions of recent ethics".[1] Among other things, it is important because it has allowed moral philosophers to better explore the logical space of moral theories, enabling them to see promising moral theories that were previously absent from the debate. This thesis continues this project of exploring the logical space of moral theories, and then applies the insights gained to various debates in moral theory. It contains a preface and five chapters. Each chapter is an independent essay that can be read without reference to any of the other chapters. Yet the arguments made in the five chapters overlap in various ways and are united by their focus on agent-relativity.

Taken together, here is roughly what these chapters say. There is an important distinction between moral rules and theories that are agent-relative and moral rules and theories that are agent-neutral. Despite some difficulties, this distinction can be made with formal precision. There is also an important related distinction between moment-relative rules and theories and moment-neutral rules and theories. Finally, there is a third important distinction between patient-relative rules and theories and patient-neutral rules and theories. Together, these three distinctions give us a precise way of understanding the structure of moral theories. Traditionally, deontology has been associated with agent-relativity and consequentialism with agent-neutrality. Several philosophers hold that consequentialism can also be agent-relative. For this to be true there must be agent-relative moral values. Some argue that such values are implausible, however, they can be defended by showing that commonsense morality is already committed to them. Thus, agent-relative consequentialism appears plausible. Other philosophers hold that deontology can be agent-neutral. I argue this is false—deontology is necessarily agent-relative. One of the main attractions of deontology is its ability to explain agent-centered constraints. However, when we think through cases involving agent-relativity and moment-relativity we see that deontology faces a trilemma with respect to agent-centered constraints. Either, (i) it is really a version of agent-relative consequentialism, or (ii) it fails to be uniquely action-guiding in cases where commonsense morality requires it, or (iii) it endorses a counterintuitive anti-aggregation principle. This trilemma poses a serious problem for the deontologist and suggests that agent-relative consequentialism is the better theory when it comes to accommodating commonsense morality.

---

[1] Thomas Hurka, 'Moore in the Middle', Ethics 113 (2003), p. 628.

Contents

Preface

This thesis is about moral theories—theories like consequentialism and deontology that normative ethicists construct in order to systematically answer first-order moral questions. One important research programme in normative ethics focuses on the structure of moral theories, looking at them from an abstract perspective and positioning them in logical space. One of the main attractions of this programme is its potential to help us make progress in old debates (such as the longstanding debate between consequentialists and deontologists) by casting those debates in a new, more logically rigorous, light. A second related attraction is that, by exploring the logical space of moral theories, new theoretical possibilities may be revealed, enriching our menu of options in normative ethics. This thesis aims to make a contribution to this research programme. It does this by clarifying in various ways the structure of moral theories and using some of the insights gained to evaluate deontology and consequentialism. One of the main structural features I focus on is the distinction between agent-relativity and agent-neutrality. This distinction has played a leading role in recent work on the structure of moral theories, and is widely recognized as one of the most important distinctions in normative ethics. Its importance to this thesis can be seen from the fact that it plays a key role in each of the five chapters. Another related structural feature is the distinction between moment-relativity and moment-neutrality. This distinction has not received the same attention as the agent-relative/agent-neutral distinction, yet I think it is also important for understanding the structure of moral theories. Thus, when I make original arguments in this thesis they have usually come from thinking more carefully about moment-relativity and its relation to agent-relativity.

I have chosen to write this thesis as a set of five essay chapters. Each chapter is thematically related to the others, and together they are complementary, often providing helpful background, or supporting premises, for each other. Nonetheless, despite these links, each chapter is an individual piece of work that stands up by itself without the need to refer to any of the other chapters. There are two reasons why I chose to write my thesis this way rather than as a more traditional monograph. First, although I found myself having many interesting ideas related to agent-relativity and moral theory, at no point did I find these ideas coalescing into one big idea. Thus, it seemed better to present my ideas separately rather than forcing them into a single complex argument. Second, I liked the option of writing independent essays as it allowed me to simultaneously write thesis chapters and articles that could be submitted to philosophy journals. In the end, this strategy proved

fruitful as chapter 2 has been accepted for publication by the journal *Ethics*, and two other chapters form the basis of papers currently under review.

I have arranged the five chapters in the order that I think best brings out their connections. However, given their independence, it is possible to read them in any order without losing coherence. One of the downsides of writing independent essay chapters on closely connected topics is that there may be some repetition. I have tried to keep this to a minimum, however, one instance of repetition deserves comment. The opening chapter presents my most considered view on the agent-relative/agent-neutral distinction. In several other chapters, there is a brief section introducing the distinction in a way that tailors it to the purposes of the chapter. At times the reader may find these sections repetitive and thus might prefer to skim, or skip them.

I will now give a brief overview of the chapters that follow. The opening chapter introduces the agent-relative/agent-neutral distinction, defending a version of the distinction for moral rules based on the work of David McNaughton and Piers Rawling. It then extends the distinction to moral theories and reasons, and makes a related distinction between moment-relative and moment-neutral rules.

Chapter two examines how agent-relativity and moment-relativity interact and argues that this leads to a third important distinction—the distinction between patient-relative and patient-neutral moral rules. It then shows how this distinction can be applied to various parts of moral theory, including being used to provide a novel reply to Derek Parfit's 'appeal to full relativity' argument.

Chapter three examines arguments that deontology can be agent-neutral offered by Francis Kamm, Tom Dougherty, and John Broome. It gives grounds for rejecting each of these arguments and then presents a general argument showing that deontology is necessarily agent-relative. This general argument helpfully clarifies the nature of deontology.

Chapter four poses a trilemma for deontologists. All deontological theories endorse agent-centered constraints, yet many deontologists fail to specify the exact form these constraints take. I show that there are four relevant forms for constraints and argue that this leads to one of three problems for the deontologist. First, her theory may turn out to be a version of agent-relative consequentialism and not a version of deontology at all. Second, her theory may fail to be uniquely action-guiding in cases where commonsense morality requires it. Third, her theory may contain a counterintuitive anti-aggregation principle. Each of these problems appears to count strongly against deontology.

The closing chapter defends agent-relative value against the objection that it is implausible because it has no precedent in ordinary moral thought and talk. I argue that both commonsense morality and moral theory connect rules requiring the maximization of states of affairs to moral value. I then argue that commonsense morality is committed to certain agent-relative maximizing rules, and conclude from this that commonsense morality requires us to postulate agent-relative values.

Looking through this chapter overview, you may notice that an appeal to commonsense morality pervades the latter chapters. This is not because I regard commonsense morality as irreproachable (for the record, I think it is an important but fallible means of testing our moral theories). Rather, it is because appeals to moral commonsense are prominent among the critics of consequentialism. As someone who finds consequentialism appealing, I aim to show that one form of consequentialism (agent-relative consequentialism) fits with commonsense morality as well as, if not better than, rival non-consequentialist approaches. This still leaves open the question of whether, taking for granted that some form of consequentialism is correct, an agent-relative version that most closely matches moral commonsense is better than all other varieties. This is not a question I can address in this thesis, although I hope to address it in future work.

M.H.

1

Agent-Relativity and Agent-Neutrality[2]

1. Introduction

Contemporary moral philosophers often distinguish agent-neutrality and agent-relativity. They apply this distinction to several different domains including reasons, rules, theories, and values. In this chapter I will defend an account of the agent-relative/agent-neutral distinction. My account is based on the account defended by McNaughton and Rawling (1991). However, I make two significant revisions to their account and thus the account I defend is substantially different from their original proposal. These revisions are made in response to two objections against McNaughton and Rawling's account that I develop here. The original McNaughton and Rawling account applies the agent-relative/neutral distinction to moral rules, and my revised account will do the same. However, I will show in later sections how it can be extended to reasons and theories. I will also show how the account can be adapted to make a related distinction between moment-relativity and moment-neutrality. The work I do in this chapter is important because the concept of agent-relativity plays a central role in all of the chapters that follow this one. Thus, this chapter aims to provide a state-of-the-art account of the distinction, marking it with formal precision and defending its significance in moral theorizing.

Here is a map of how this chapter will proceed. I begin in §2 by giving a brief history of the development of the agent-relative/neutral distinction in the late 20th century. Next, in §3 I look at some commonly cited examples of agent-neutral and agent-relative rules. I follow this in §4 with a list of key desiderata that an account of the distinction ideally should meet. In §5 I review McNaughton and Rawling's account of the distinction. Then, in §6 and §7, I raise two objections against their account, and explain and defend two corresponding revisions. Following this I extend the account I have defended in various ways: to permissions (§8), theories and reasons (§9), and moments (§10). I conclude in §11 by discussing connections between the formal distinction I defend and the intuitive account of the distinction defended by Derek Parfit.

---

[2] A version of this chapter was published as Hammerton (2019).

2. A Brief History

The terms 'agent-relative' and 'agent-neutral' were coined by Derek Parfit in his 1979 article 'Prudence, Morality and the Prisoner's Dilemma'. However, the distinction predates Parfit, and Thomas Nagel appears to have been the first to clearly identify it in his 1970 work *The Possibility of Altruism*. In that work Nagel makes a distinction between 'subjective' and 'objective' reasons. According to Nagel, a subjective reason is a reason that, when put into its correct general form, makes reference to the agent whose reason it is. On the other hand, an objective reason is a reason that, when put into its correct general form, makes *no* reference to the agent whose reason it is.[3] Nagel took this formal distinction to correspond to a distinction we can make at the intuitive level between different kinds of reasons, those that have a special kind of personal element built into them and those lacking this element.

In his 1979 article, and then later in his monograph *Reasons and Persons*, Derek Parfit developed Nagel's distinction further, coining new terms to name it, and applying it to rules and theories and not just reasons. In contrast to Nagel's original formal account of the distinction, Parfit instead offers an intuitive account, saying that agent-relative rules, reasons, and theories give different ultimate aims to different agents, whereas agent-neutral rules, reasons, and theories give the same ultimate aim to all agents. The distinction plays a key role in several arguments in *Reasons and Persons*, and Parfit appears to hold that, when made at the level of theories, it corresponds to the distinction between consequentialism and non-consequentialism.[4]

Interestingly, several decades prior to Nagel and Parfit's contributions, C.D. Broad, in a critique of G.E. Moore's arguments against ethical egoism, contrasts ethical egoism with what he calls 'ethical neutralism'. According to Broad:

> Ethical neutralism assumes that there is a certain one state of affairs—"the sole good"—at which everyone ought to aim as an ultimate end. Differences in the proximate ends of different persons can be justified only in so far as the one ultimate end is best secured in practice by different persons aiming, not directly at it, but at different proximate ends of a more limited kind.[5]

Broad's account of ethical neutralism appears to anticipate some of the ideas of Nagel and Parfit. Indeed, it is probable that they had both had read Broad's critique of Moore and

---

[3] See Nagel (1970), pp. 47-92. Nagel's view is complicated by the fact that he held a controversial, and somewhat baroque, view about what the correct general form of a reason is. He later simplified the view in Nagel (1986), pp. 152-153, and replaced his original terms 'subjective' and 'objective' with Parfit's terms 'agent-relative' and 'agent-neutral'.

[4] See Parfit (1984), p. 27.

[5] Broad (1942), p. 46.

were influenced by the above passage. However, Broad falls short of making a distinction between agent-relativity and agent-neutrality because he categorizes ethical neutralism as the class of theories directing agents to bring about an impersonal good (which are only one kind of agent-neutral theory), and contrasts it with ethical egoism (which is only one of many possible agent-relative theories).

Following on from the work of Nagel and Parfit, many moral philosophers have drawn a distinction between agent-relative and agent-neutral reasons, rules, theories, and values. The distinction has been employed in many different contexts and has become one of the most important distinctions in contemporary normative ethics. It is, for example, considered essential to understanding debates around deontological constraints, special duties, and personal prerogatives. And many have come to regard it as the central fault line in moral theory, marking two rival approaches to ethics.[6] As a result of this influence, Thomas Hurka has described the distinction as one of "the greatest contributions of recent ethics".[7] However, despite its prominent role in contemporary moral theorizing, there is still no agreement on how best to draw the distinction. Nagel's (1970) and (1986) formal accounts of the distinction are widely criticised for not correctly picking-out the intuitive distinction that he was attempting capture.[8] Parfit's (1984) account in terms of ultimate aims has been widely used. However, some find it limited because it lacks the precision of a formal account. Furthermore, it is controversial whether all moral rules literally give agents aims, and thus there is a worry that Parfit's talk of the 'ultimate aim' given by a moral rule is, at best, metaphorical.[9] As a result, alternative accounts of the distinction have been proposed by Pettit (1987), McNaughton and Rawling (1991), Skorupski (1995), and Ridge (2005a). The account of McNaughton and Rawling has probably received the most support of all these proposals. However, there is still much controversy about how to make the distinction, and no account has reached anything approaching a consensus.

3. The Intuitive Distinction

Although there are several competing accounts of the agent-relative/agent-neutral distinction, there is nonetheless general agreement on how the distinction should classify various cases. Looking at some of these cases is a good place to start when examining the distinction. As my focus here is primarily on the distinction applied to moral rules, the

---

[6] See especially Dreier (1993).
[7] Hurka (2003), p.628.
[8] For criticisms see McNaughton and Rawling (1991), Schroeder (2007) pp. 276-278, Ridge (2005a), and Portmore (2014).
[9] This point is made by Portmore (2014).

cases I examine will be cases of moral rules. Furthermore, they will be cases of rules that are requirements (as opposed to permissions, or recommendations). Consider the following ten moral rules:

(1) Each agent must not kill innocent people.

(2) Each agent must minimize the killing of innocent people.

(3) Each agent must care for her children.

(4) Each agent must ensure that children are cared for by their parents.

(5) Each agent must maximize her own wellbeing.

(6) Each agent must maximize general wellbeing.

(7) Each agent must maximize the wellbeing of her friends.

(8) Each agent must ensure that no one is disloyal to the Supreme Leader.

(9) Each agent must minimize her violations of rule (1).

(10) Each agent must minimize general violations of rule (1).

Each of these rules is universal in the sense that it applies to all moral agents. However, some of these rules are classified as agent-relative and some as agent-neutral. In fact, all the odd numbered rules (1,3,5,7,9) are agent-relative, and all the even numbered rules (2,4,6,8,10) are agent-neutral. Exactly why the rules are classified this way is difficult to say without taking a stand on substantive issues in dispute between different accounts. However, it is worth reviewing how Parfit's intuitive account of the distinction classifies these cases to see what might be at stake. According to Parfit, an agent-relative rule gives different aims to different agents, whereas an agent-neutral rule gives all agents the same aim.

Let's start by applying Parfit's idea to rules (1) and (2). Rule (1) requires of each agent that she does not kill innocent people. In doing so it gives each agent a special concern with her own actions, for it requires her not to kill innocent people even if doing so is the only way to prevent more killings of innocent people by others. Therefore, (1) appears to give different aims to different agents, by giving each agent the aim that *she* does not kill innocent people. For example, it gives Alfred the aim that *Alfred does not kill innocent people*, and Sita the aim that *Sita does not kill innocent people*. On the other hand, rule (2) requires each agent to minimize the killing of innocent people and thus gives each agent a general concern with acts of killing committed by anyone rather than, as (1) did, a special concern with her own acts. It requires an agent to kill an innocent person in any case where doing so is the only way to prevent more killings of innocent people being committed by others. Given this result, (2) appears to give all agents the same aim—the aim that the killing of innocent people is minimized.

We can now apply Parfit's idea to the remaining rules. Rule (3) is agent-relative because it gives each agent a special concern with her own acts of caring, giving her the aim that she cares for her children. On the other hand, (4) is agent-neutral because it gives all agents the same aim—that children are cared for by their parents. Likewise, (5) is agent-relative because it gives each agent the aim that *her* own wellbeing is maximized, whereas (6) is agent-neutral because it gives all agents the same aim—that everyone's wellbeing is maximized. Rule (7) is agent-relative because it gives each agent the different aim that the wellbeing of *her* friends is maximized, whereas (8) is agent-neutral because it gives all agents the same aim—that no one is disloyal to the Supreme Leader. Finally, (9) is agent-relative because it gives each agent a special concern with her own violations of (1), giving her the aim that violations made by *her* are minimized. By contrast, (10) is agent-neutral because it gives all agents the same aim—that violations of (1) committed by any agent are minimized.

Parfit's intuitive account of the distinction appears to get the correct verdict in all these examples. However, as I noted in the previous section, it lacks formal precision because it relies on our intuitive judgements about what aim each rule gives, rather than identifying a formal feature present in the underlying structure of each rule. Furthermore, it controversially assumes that all moral requirements give agents aims. I will discuss some of these issues later in §11.

Having a list of paradigm examples like the one above demonstrates how hard it can be to account for the agent-relative/agent-neutral distinction. For example, you might suppose that agent-neutrality is connected to maximization and agent-relativity connected to non-maximizing rules. However, rules (5), (7), and (9) demonstrate that a maximizing rule can be relative, and rules (4) and (8) demonstrate that you can have neutrality without maximization. Alternatively, you might associate agent-relativity with the presence of a personal pronoun in the rule, and agent-neutrality with the absence of personal pronouns. However, rule (1) demonstrates that you can have relativity without the presence of a personal pronoun, and rule (4) demonstrates that personal pronouns can occur in neutral rules.

Finally, it is worth considering a particular challenge to the standard classifications of rules like those above. None of the rules above contain any clauses stating conditions under which the rule does, or does not, apply to an agent. However, rules often do contain clauses stating such preconditions. For example, many philosophers believe that *ought-implies-can*, which is standardly interpreted as saying that a moral requirement only applies to an

agent if she has the ability, and the opportunity, to obey the requirement. If this thesis is accepted, then (1) and (4) (to take two examples) would need to be revised as follows:

> (1′) Each agent, whenever she has the ability and opportunity to do so, must not kill innocent people.

> (4′) Each agent, whenever she has the ability and opportunity to do so, must ensure that children are cared for by their parents.[10]

Some (such as Postow 1997) have interpreted all rules with these kinds of clauses as agent-relative on the grounds that such clauses are making a kind of special reference back to the agent bound by the rule. Rønnow-Rasmussen (2009) has gone further, arguing that all moral reasons (and, by extension, all moral rules) have a 'personalizability-feature' like this built into them because they are rules or reasons for someone. He claims that it follows from this that all moral reasons (and rules) are agent-relative, and thus the distinction is of limited importance to moral theory. However, such views miss the theoretically interesting difference between the odd and even numbered rules we have been considering. Thus, we should reject views like that of Rønnow-Rasmussen that reduce the distinction to a triviality,[11] and views like that of Postow that take the agent-neutral status of a rule like (4) to change when a preconditional clause is added. However, the fact that moral rules can have such 'personalizability-features' should remind us that, whatever account of the distinction we do settle on, it must account for these features. Thus, it is not good enough that an account correctly classifies rules like (1)-(10) as agent-neutral and agent-relative. To be satisfactory, it must also correctly classify rules like (1′) and (4′) that contain clauses limiting when, and to whom, the rule applies.

4. Desiderata

I am now in a position to state what I take to be the key desiderata that any account of the agent-relative/agent-neutral distinction must satisfy. I take there to be four key desiderata.

First, as was suggested by the discussion in the previous section, an account of the agent-relative/agent-neutral distinction should correctly classify paradigm cases, such as (1)-(10) above, as relative or neutral. In other words, the classifications given by an account of the distinction should match the intuitive distinction that Nagel and Parfit pointed out

---

[10] It is important that the ought-implies-can clause appears before the deontic operator 'must' and not after. In other words, the deontic operator should be read as having narrow scope and not wide scope. For if the operator is read with wide scope then the resulting rule implausibly allows an agent to avoid transgressing inconvenient moral requirements by making it the case that she lacks the ability or opportunity to do the required action. See Portmore (2001) and McNaughton and Rawling (2002) for discussion.

[11] This point is made by Ridge (2005b) and Portmore (2014).

between various kinds of rules (or reasons, or theories). The account of Rønnow-Rasmussen (2009) can be ruled out on these grounds as it classifies all rules as agent-relative. Likewise, the accounts offered by Nagel (1986) and Pettit (1987) appear to fail here because they incorrectly classify rules like (4) as agent-relative.[12]

Second, an account of the distinction should be theoretically neutral. By this I mean that it should not assume any controversial ethical or metaethical claims. For example, in the previous section we considered the *ought-implies-can* thesis and noted that it is controversial. Given this controversy, an account of the agent-relative/agent-neutral distinction should be able to classify both rules that incorporate the *ought-implies-can* clause, and those that do not. An account that only classifies rules without the clause because it presupposes that the *ought-implies-can* thesis is false, fails on neutrality. The motivation for theoretical neutrality is as follows. The agent-relative/agent-neutral distinction is of fundamental importance to moral theory, in part, because it appears to be independent of most substantive moral disputes and thus gives us a way of categorizing moral theories that does not beg the question in favour of some theories and against others. Thus, in searching for an account of the distinction, we should look for an account that is compatible with various stances taken in moral theory. Of course, it may ultimately turn out that the distinction cannot be made in a completely theoretically-neutral way.[13] But even then it seems desirable to make it in a way that minimizes its theoretical assumptions, making it as broad in its application as possible.

Various accounts of the distinction have been rejected because of they lacked theoretical neutrality. For example, the account of Nagel (1970) has been criticised for making several controversial assumptions about the nature of moral reasons, including the assumption that moral reasons are always teleological reasons.[14] Also, Michael Ridge (2005a) has criticised some accounts that apply the distinction to moral reasons (such as the accounts of Parfit (1984) and Nagel (1986)) for being incompatible with moral particularism. Ridge criticises these accounts not because he thinks that particularism is true, but because he thinks that an adequate account of the distinction should not beg the question against it.

A third desideratum is that the distinction is drawn by appealing to properties that are relevant and explanatory, rather than properties that are ad hoc and arbitrary. Properties

---

[12] See McNaughton and Rawling (1991) for this criticism of the accounts of Nagel (1986) and Pettit (1987).

[13] For example, obviously if we are to apply the distinction to moral rules, moral values, or moral reasons we must make the theoretical assumption that each of these moral concepts is coherent. However, to take one example of someone disputing this, Thomson (1993) has argued that any non-attributive use of 'good' is incoherent.

[14] For criticism see Skorupski (1996), p. 236, Ridge (2005a), and Schroeder (2007), pp. 276-278

associated with agent-relativity are relevant and explanatory when they are appropriately connected to Nagel and Parfit's motivating idea that some moral rules (reasons, theories, and values) have a special kind of personal element built into them. Properties associated with agent-neutrality are relevant and explanatory when they are appropriately connected to the converse idea that some moral rules (reasons, theories, and values) lack this special personal element. This desideratum rules out making the distinction with gerrymandered properties picked out only because they correctly classify the paradigm cases. For example, suppose there is an account of the distinction that marks agent-neutrality by the presence of a logical property P which happens to be relevant and explanatory. Furthermore, suppose that many paradigm agent-neutral rules possess P, but all agent-neutral rules containing preconditional clauses lack P, and thus the account appears inadequate. It would be a mistake to address this problem by revising the account to say that rules with preconditional clauses count as agent-neutral if, when the clauses are removed, they possess P. This is a mistake because, although it may lead to an account that correctly classifies the paradigm cases, it links agent-neutrality to an indirect property that appears irrelevant and unexplanatory.

Finally, an account of the distinction should be simple and elegant. I take this desideratum to be defeasible, as there are various circumstances where it is reasonable to accept a complicated, inelegant account of the distinction. However, others things being equal, we prefer an account with more simplicity and elegance over one with less.

5. McNaughton and Rawling's Distinction

McNaughton and Rawling (1991) offer a formal account of the agent-relative/agent-neutral distinction that applies it to moral rules.[15] The first step in their account is to take all moral rules and transform them into semiformal statements with the following form: '$(x)(x$ must ensure that [...])'. Thus, consider rules (1)-(6) from above:

> (1) Each agent must not kill innocent people.
> (2) Each agent must minimize the killing of innocent people.
> (3) Each agent must care for her children.
> (4) Each agent must ensure that children are cared for by their parents.
> (5) Each agent must maximize her own wellbeing.
> (6) Each agent must maximize general wellbeing.

---

[15] They discuss and defend their account further in McNaughton and Rawling (1992), (1993), (1995a), (1995b), (1998), and (2002).

McNaughton and Rawling's approach transforms these rules into semiformal notation as follows:

> (1*) $(x)(x$ must ensure that $[x$ does not kill innocent people$])$
>
> (2*) $(x)(x$ must ensure that $[(y)(y$ does not kill innocent people$)])$
>
> (3*) $(x)(x$ must ensure that $[(y)(y$ is x's child $\rightarrow x$ cares for $y)])$
>
> (4*) $(x)(x$ must ensure that $[(y)(z)(y$ is a child of $z \rightarrow z$ cares for $y)])$
>
> (5*) $(x)(x$ must ensure that $[x$ has maximal wellbeing$])$
>
> (6*) $(x)(x$ must ensure that $[(y)(y$ has maximal wellbeing$)])$

The second step is to point out that when the rules are transformed in this way then all agent-relative rules have an occurrence of 'x' in the square brackets that is bound by the initial universal quantifier (e.g., (1), (3), (5)), whereas all the agent-neutral rules have no such occurrence (e.g., (2), (4), (6)). Thus, McNaughton and Rawling's account can be stated as follows:

> A rule expressing a moral requirement is agent-relative *iff*, when the rule is represented with the form '$(x)(x$ must ensure that [...]$)$', there is an ineliminable occurrence of 'x' in the square brackets that is bound by the initial universal quantifier.
>
> A rule expressing a moral requirement is agent-neutral *iff* it is not agent-relative.

In the next section I will raise two objections against this account of the agent-relative/agent-neutral distinction. However, before proceeding, it is worth considering how McNaughton and Rawling's account does reasonably well by the desiderata outlined above. First, as we have seen with the six rules above, their account does well at correctly classifying paradigm cases as either agent-relative or agent-neutral. In particular it correctly classifies rules like (1) and (4) that other accounts struggle with.[16] It also correctly classifies rules containing preconditional clauses such as (1′) and (4′) because such clauses will appear outside the square brackets when rules like (1′) and (4′) are translated into McNaughton and Rawling's semiformal notation and thus the presence of these clauses will not alter a rule's classification as neutral or relative.[17] Thus, McNaughton and Rawling's account does very well at satisfying the first desideratum. It also does well on the second desideratum as it does not appear to make any morally controversial assumptions. For example, it appears compatible with different views about the nature of

---

[16] McNaughton and Rawling (1991) show how several other accounts of the distinction do not correctly classify rules like (1) and (4).

[17] To see that preconditional clauses must appear outside the square brackets consider the point I make about wide vs. narrow scope in footnote 8.

moral rules, and the validity of *ought-implies-can*. Regarding the third desideratum, it also does well. It represents moral rules as containing propositional content and ties agent-relativity to the appearance of a variable bound by the initial universal quantifier in this content. This feature appears to be both relevant and explanatory. For the initial universal quantifier is the part of the rule that applies it to each agent, and when a variable bound by this quantifier appears in the rule's propositional content it is a kind of reference back to the agent bound by the rule. Thus, the formal property McNaughton and Rawling use to mark agent-relativity is appropriately connected to the intuitive idea that some moral rules have a special kind of personal element built into them. Finally, their account is reasonably simple and elegant, and so does well at meeting the fourth desideratum. These points together explain why McNaughton and Rawling's account of the distinction has been fairly well received, and has perhaps received more support than any other account of the distinction proposed.[18]

Finally, we should note two important clarifications of McNaughton and Rawling's account. First, McNaughton and Rawling (1995, p. 34) state that they take '*x* must ensure that ...' to be shorthand for '*x* must ensure, to the best of *x*'s abilities, and insofar as there is no conflicting duty of greater weight, that ...'. As we will see later, this qualification is important because helps them address a certain line of criticism.

Second, we should explain the important work done by the term 'ineliminable' in McNaughton and Rawling's definition of agent-relativity. The inclusion of this term is required because there is a trivial way in which any moral rule transformed into the canonical form can be made to have an occurrence of '*x*' in the square brackets. For example, the agent-neutral (2) could be transformed as follows:

(2**) $(x)(x$ must ensure that $[x$ ensures that $((y)(y$ does not kill innocent people$))])$

However, it would obviously be wrong to claim that (2) was agent-relative because of the appearance of '*x*' in the square brackets in (2**) as this appearance of '*x*' is eliminable. It is eliminable because there is a translation of (2) into the canonical form (namely (2*)) that has the same meaning as (2**) and yet does not contain the variable '*x*' in the square brackets. By contrast, there is no valid translation of the agent-relative (3) into the canonical form where the variable '*x*' does not appear in the square brackets. This clarification is important because it is sometimes claimed that maximizing rules like (2) are actually better translated into the canonical form as follows:

---

[18] McNaughton and Rawling's distinction is made use of by Howard-Snyder (1996), Louise (2004), Dougherty (2013), and Portmore (2014).

(2***) (x)(x must ensure that [x minimizes the killing of innocent people])

However, if (2) is best translated this way then it appears agent-relative as 'x' occurs in the square brackets. McNaughton and Rawling's response to this concern is the same as their response to rules like (2**).[19] Rule (2***) is not an agent-relative rule because, although there is an occurrence of 'x' in the square brackets that is bound by the initial universal quantifier, this occurrence is eliminable. It is eliminable because they take (2***) to be equivalent to (2*) and thus hold that (2*) demonstrates that the rule can be formulated without an occurrence of 'x' in the square brackets. Perhaps this response is adequate. However, in §7 I will defend a revision of McNaughton and Rawling's account that, as a side-result, resolves this problem in a different way.

6. Two Objections

McNaughton and Rawling offer a promising account of the agent-relative/agent-neutral distinction. However, there are two problems with their account that I think make it unsatisfactory as they present it. Both problems come back to the first of their two steps— the step where they transform moral rules into the semiformal notation. As McNaughton and Rawling acknowledge, these transformations must preserve the content of the rules that are transformed.[20] In other words, the transformations must be accurate translations from ordinary English into the semiformal notation. For if the transformation of a rule changes its content, then showing that the transformed rule has a logical property that we associate with agent-relativity fails to explain why the original rule (with its different content) might count as agent-relative. The two objections I raise in this section stem from two ways in which McNaughton and Rawling's proposed transformations fail to give us accurate translations.

The first objection concerns the step from (1) to (1*). Recall:

(1) Each agent must not kill innocent people.

McNaughton and Rawling translate this rule into their canonical form as follows:

(1*) (x)(x must ensure that [x does not kill innocent people])

If this translation is correct then *requiring an agent not to kill* must be equivalent to *requiring an agent to ensure that she does not kill*. More generally, if direct-rules like (1)

---

[19] For their response see: McNaughton and Rawling (1991), footnote 7, and (1993), pp. 83-86.
[20] McNaughton and Rawling (1993), p. 85, say: "Transmutation to form F must be transmutation without alteration in ethical content".

can be translated as ensuring-rules like (1*) then every rule requiring an agent not to perform a certain act must be equivalent to a corresponding rule requiring that agent to ensure that she does not perform that act. However, these things are not equivalent, as there are counterexamples where an agent does not perform an action without ensuring that she does not perform it. In these cases the agent obeys a direct-rule without obeying the corresponding ensuring-rule. For example, suppose that Lee aims his rifle at the President's head and pulls the trigger with the intention and expectation that in doing so he will assassinate the President. Furthermore, suppose that, by chance, the President turns his head just as Lee fires, fortuitously moving it out of the path of the oncoming bullet and narrowly escaping the assassination attempt. In such a scenario, it is true that Lee has not done a certain act (the act of killing the President), yet it seems false that Lee has ensured that he did not do this act. After all, he has done various things in an attempt to make it the case that he *did* do this act. And it seems odd to describe a circumstance where an agent tries to make it the case that she φ's, but fails due to bad luck, as a case where 'she ensures that she does not φ'.

The above example is sufficient to prove that '*x* does not φ' and '*x* ensures that *x* does not φ' are not equivalent. However, it is helpful if we add a second example showing further reasons why we might reject this equivalence claim. I have not performed the action of killing Napoléon Bonaparte. However, it is false that I have *ensured* that I have not killed him. For Napoléon Bonaparte died in 1821 more than 150 years before my birth, and thus I have never had the opportunity (or even the physical possibility) of killing him. No act or omission of mine has ever made it more or less likely that I will not kill him.[21]

These examples show us that McNaughton and Rawling's translation of (1) as (1*) fails. Further reflection shows us that a rule like (1) cannot be transformed into the canonical form without altering its content. For there is no alternative to (1*) that both fits the canonical form and is an accurate translation of (1). Thus, McNaughton and Rawling's account of the agent-relative agent-neutral distinction fails. To draw the distinction with formal precision they assume that all moral requirements can be translated as rules with the form: '(*x*)(*x* must ensure that [...])'. However, a rule like (1) cannot be successfully translated in this way, and thus their account fails to classify (1) as agent-relative.

Before moving on it is worthwhile considering two responses to this objection and seeing why they do not work. First, you might attempt to defend McNaughton and Rawling's

---

[21] I am grateful to Daniel Nolan for suggesting this example to me. I should also add that I am assuming in this example, for the sake of argument, that time travel is nomologically impossible.

account by pointing out that although (1) cannot be translated as (1*), a different rule that is similar to (1) is correctly translated as (1*):

> (11) Each agent must ensure that she does not kill innocent people.

Like rule (1), rule (11) is an agent-relative rule. Furthermore, these two rules appear to be rivals. If you think there is an agent-relative constraint on killing innocent people then rules (1) and (11) are two possible forms that constraint might take. Thus, you might attempt to defend McNaughton and Rawling's account by arguing that (11) is better than (1) at capturing such a constraint. More generally, you might argue that ensuring-rules like (11) are always superior to their corresponding direct-rule (i.e. rules like (1)). From this you may conclude that it does not matter that McNaughton and Rawling's account does not classify rules like (1), as it does correctly classify rules like (11) and the best moral theories will endorse rules like (11) rather than rules like (1) (i.e. they will endorse ensuring-rules and eschew direct-rules). Perhaps there is some merit to the claim that ensuring-rules are preferable to direct-rules (although I will not take a stand on this). Nonetheless, this argument cannot save McNaughton and Rawling's account against the present objection. For, as we saw in §4, one desideratum that a successful account of the agent-relative/agent-neutral distinction must satisfy is theoretical neutrality. An account of the distinction must not take a stand on substantive moral questions, especially those that are controversial. The argument we have just considered fails this desideratum because it requires the controversial assumption that agent-relative constraints are best formulated as ensuring-rules rather than direct-rules, and thus abandons theoretical neutrality.

Second, you might attempt to defend McNaughton and Rawling's account by granting that direct-rules like (1) are plausible but arguing that such rules entail ensuring-rules like (11) because of the following general principle:

> If some act or omission is required of agent A, then A is also required to ensure that she does the relevant act or omission.

McNaughton and Rawling's account can then be used to classify (1) as agent-relative by appealing to the fact that the above general principle and rule (1) jointly entail rule (11), and (11) meets the formal criteria for agent-relativity. However, as in the previous defence, this defence fails the desideratum of theoretical neutrality because it takes for granted the controversial general principle mentioned above. It also fails the third desideratum of explanatory relevance by linking the agent-relativity of (1) to an indirect property that fails to explain it (i.e. the property of, in conjunction with the general principle, entailing a rule that meets the criteria for agent-relativity).

We can now move on to the second objection against McNaughton and Rawling's account. This objection concerns their translations of maximizing/minimizing rules. For example consider rules (2), (5), and (6) from above:

(2) Each agent must minimize the killing of innocent people.

(5) Each agent must maximize her own wellbeing.

(6) Each agent must maximize general wellbeing.

McNaughton and Rawling transform these rules into the canonical form as follows:

(2*) $(x)(x$ must ensure that $[(y)(y$ does not kill innocent people)$])$

(5*) $(x)(x$ must ensure that $[x$ has maximal wellbeing$])$

(6*) $(x)(x$ must ensure that $[(y)(y$ has maximal wellbeing)$])$[22]

The objection is that (2*), (5*), and (6*) are not adequate translations of (2), (5), and (6). More fundamentally, the objection is that there is no satisfactorily way to translate maximization rules into McNaughton and Rawling's canonical 'ensure that' form. The objection can initially be pressed by pointing out that rules like (2) and (2*) give different deontic verdicts in certain cases and thus cannot be equivalent. For example, consider the following case:

> **The super crime-fighter**: Clark Kent devotes his life to preventing the killing of innocent people, often performing superhuman feats and pushing himself to his limits. His efforts are a great success and innocent people are rarely killed in his society. However, occasionally he finds himself unable to prevent all such killings. For instance, sometimes supervillains in different locations are simultaneously killing random innocent people and although Clark can reach one location in time to prevent the killing, it is impossible for him to reach both. In these situations he prioritizes the location where the most killings of innocents can be prevented, allowing the killings at the other location to occur.

In this example Clark appears to obey rule (2). For, at each decision point he faces, he always takes the available option that will lead to the fewest killings of innocent people, and this is what it means to *minimize* such killings. However, Clark appears to disobey rule (2*). For, despite his tremendous efforts, innocent people are occasionally killed and when this happens he has failed to ensure otherwise. Therefore, rule (2*) is much more demanding than rule (2).

This example demonstrates that (2) and (2*) have different content. However, McNaughton and Rawling can respond by pointing to their qualification (mentioned

---

[22] To be clear, they only explicitly consider rule (2), translating it as (2*). However, they intend their account to apply the distinction to all other moral rules, including rules like (5) and (6). Rules (5*) and (6*) are my best attempt at fitting (5) and (6) into their framework.

earlier in §5) that an agent is only required to ensure *to the best of her ability*. Thus, they can say that the correct translation of (2) is not (2*) but the closely related:

> (2′) (x)(x must ensure, to the best of x's ability, that [(y)(y does not kill innocent people)])

In the above example, although Clark Kent does not prevent every killing of an innocent, he does prevent such killings to the best of his ability, and thus obeys (2′). In this respect (2′) appears to be a reasonable translation of (2). However, a problem remains. Rule (2) does not only let Clark off the hook when there is nothing he can do to prevent an innocent person being killed, it also instructs him on how to make choices between options that involve preventing different killings; telling him to take the option that will prevent the most killings. However, it is not clear that (2′) provides these further instructions. Of course, McNaughton and Rawling might stipulate that it does, stating that the 'best' in their formulation can be read as requiring this kind of maximization. However, as Skorupski (1996, p. 240) notes, this leads to a further problem. McNaughton and Rawling say that all rules are qualified with the 'to the best of *x*'s ability' clause. However, if this clause means the agent is required to maximize the relevant quantity mentioned in the square brackets then, although this may get the correct result with rules like (2), it appears to get the incorrect result with other rules. For example, it interprets rule (1) as requiring each agent to minimize her killings of innocent people, essentially making it equivalent to rule (9). However, this is contrary to how (1) is normally read. In its standard reading, (1) requires agents to never kill and classes them as doing the wrong thing any time they kill, even if they are killing to minimize their total killings. Thus, McNaughton and Rawling appear to face a dilemma; their canonical form either fails to accurately translate maximizing rules, or it successfully translates these rules but at the expense of mistranslating constraints like (1).

We can add to the above concern a second problem for McNaughton and Rawling's translation of maximizing rules—a problem related to the first objection above. It is possible for an agent to 'accidently' maximize the non-killing of innocent people by performing an action that results in non-killing being maximized while not intending or foreseeing that her action has this result. In such cases, although the agent maximizes, she does not appear to *ensure* that she maximizes. Thus it seems possible to obey (2), and yet disobey (2′), which demonstrates that they are not equivalent.

7. Revising McNaughton and Rawling

Above we have seen that McNaughton and Rawling's canonical 'ensure that' form is unable to adequately accommodate direct-rules and maximizing rules. Some may be tempted to conclude from this that McNaughton and Rawling's method for capturing the difference between agent-neutral and agent-relative rules does not work and that a fundamentally different account of the distinction is needed.[23] However, I think this is a mistake. For although their canonical form is inadequate, one of their key ideas is very promising. They realized that, if all moral rules are expressible such that they contain propositional content, then the difference between agent-relative and agent-neutral rules might be discernible in that content.[24] More precisely, they noted that the difference may be marked by the presence or absence of a certain kind of variable in a rule's content. They then erred in thinking that all rules can be expressed as ensuring-rules, and thereby shown to be expressible with propositional content. Some rules can be expressed as ensuring-rules but others cannot. However, if alternative translations involving propositional content can be found for those rules that are not expressible as ensuring-rules, then McNaughton and Rawling's proposal, with appropriate modifications, may ultimately succeed. In this section I will attempt this task; showing how to translate direct-rules and maximizing rules into semiformal notation such that they contain propositional content, avoid the objections raised above, and have the appropriate formal properties to be correctly classified as agent-neutral or agent-relative.

Let's start with direct-rules. We saw above that 'x does not ϕ' and 'x ensures that x does not ϕ' are not equivalent, and so (1) cannot be translated as (1*). However, if we can find an alternative verb to 'ensure' that both takes propositional content, and can be used to generate an accurate translation of (1), then we will have a way of revising McNaughton and Rawling's account to accommodate direct-rules. In other words, we are looking for a rule with the form 'x {verb} that x does not ϕ' that is equivalent to 'x does not ϕ'. One possibility is using the verb 'making it the case'. Perhaps 'x makes it the case that x does

---

[23] For example, Skorupski (1996), pp. 239-240, presents a version of the second objection and rejects McNaughton and Rawling's account on this ground. He then goes on to argue that his own version of the distinction is better. I think that Skorupski is too quick in dismissing McNaughton and Rawling's account as he fails to consider ways of revising their account (such as the revisions I offer below) that resolve his objection. I also think that Skorupski's alternative account is implausible because it misclassifies several paradigm cases (failing the first desideratum). See, Broome (1995) and McNaughton and Rawling (1995b) for more elaboration on this problem with Skorupski's account.

[24] It should be noted that when I talk of moral rules containing propositional content I am not taking a stand on whether 'ought' statements contain a relation between an agent and a proposition (as argued by Broome 1999, and Wedgewood 2006), or an agent and an action (as argued by Schroeder 2011). For an 'ought' statement may contain propositional content directly by taking a proposition as one of its relata, or indirectly by taking as one of its relata an action that contains propositional content (e.g. the action of 'ensuring that p').

not φ' is equivalent to '*x* does not φ'. To test this equivalence claim let's consider the two examples from above. In the chancy-failed-assassination case, 'make it the case that' appears to do better than 'ensure'. For, although Lee is *attempting* to make it the case that he assassinates the President, it seems plausible to say that by firing his bullet at that exact moment, in that exact way, he is (by accident) making it the case that he does *not* assassinate the President at that moment. However, despite doing well on the first example, 'make it the case that' fails when it comes to the Napoléon example. For, it is true that I do not kill Napoléon Bonaparte and yet false that I make it the case that I do not kill him. The latter is false because 'making it the case' suggests that I have done some act (or perhaps some omission) that causes the relevant state of affairs to obtain. Yet, given the physical impossibility of me killing Napoléon Bonaparte, it seems false to say that my acts (or omissions) have caused this negative state of affairs to obtain. It follows from this that direct-rules are not equivalent to their corresponding make-it-the-case-that rules.

Two other possible verbs are 'bring it about that' and 'see to it that'.[25] Perhaps '*x* brings it about that *x* does not φ' or '*x* sees to it that *x* does not φ' are equivalent to '*x* does not φ'. However, because of the causal connotations they carry, both these verbs fail in the Napoléon example for the same reason 'make it the case that' failed. I do not kill Napoléon Bonaparte and yet it is false that I bring it about that, or see to it that I do not kill him. Thus, we are looking for a verb that is appropriately connected to an agent not performing an act without implying that the agent did something that caused herself not to perform that act.

A promising suggestion is the verb 'act such that'. Perhaps '*x* does not φ' is equivalent to '*x* acts such that *x* does not φ'. Let's test this proposal against the two examples considered above. In the chancy-failed-assassination case, it is true that Lee does not assassinate the President, and also true that he acts such that he does not assassinate the President. For firing his bullet at that exact moment, in that exact way, is one way of acting such that he does not assassinate the President. Likewise, in the Napoléon example, it is true that I do not kill Napoléon Bonaparte, and also true that I act such that I do not kill him. This is because 'acting such that' does not carry the causal connotations that 'making it the case that' carries. Instead it means something weaker along the lines of 'performing acts or omissions that, given the state of the world, are consistent with it being the case that'.[26]

---

[25] 'See to it that' has pedigree in deontic logic. For example, see Belnap and Perloff (1988).

[26] This is the reading I find most natural for the locution 'act such that'. If others read this locution differently this does not matter as my argument does not require this piece of English to have this exact meaning. All I need to show is that there is a possible verb (perhaps one that I define by stipulation) that has the relevant properties.

And, at each moment when I do not kill Napoléon Bonaparte it is also true that I am performing acts or omissions that, given the state of the world, are consistent with me not killing him. Of course, this is true in a trivial way because any act or omission I might perform is consistent with me not killing him, but it is true nonetheless.

The results we have just seen, together with reflection on further cases, appears to show that 'x acts such that x does not φ' is true if, and only if, 'x does not φ', and thus they are equivalent.[27] It follows that although (1) cannot be accurately translated as (1*) it can be translated as follows:

> (1°) (x)(x must act such that [x does not kill innocent people])

The square brackets in rule (1°) contain an ineliminable occurrence of the variable 'x' that is bound by the initial universal quantifier, and so (1°) can be classified as agent-relative on these grounds. Thus, we appear to have resolved the first objection.

We can now look at the problem of fitting maximizing rules into McNaughton and Rawling's account. We saw above that translating (2) as either (2*), or (2′) was problematic for various reasons. Thus, we need a better translation of (2) that adequately captures the notion of maximizing but, following McNaughton and Rawling's key insight, does so with a verb that takes propositional content. I suggest the following translation:

> (2°) (x)(x must bring it about to the greatest degree x can that [In this world there is minimal killing of innocent people])

This rule adequately captures the notion of 'maximizing' because to maximize a quantity appears equivalent to bringing it about to the greatest degree you can that the world has the maximal amount of that quantity. Thus, consider the case of the Super Crime-Fighter discussed above in §6. Clark Kent obeys rule (2) in this case yet does not obey (2*) or (2′).

---

[27] I was once worried by the following putative counterexample based on a famous example from the *ought-implies-can* literature: The bride has promised the groom to be at their midday wedding, but has second thoughts and boards a flight at 11:30am to escape her pending marriage. Once in the air it is no longer possible for her to keep her promise to the groom. Now, granting that rules give deontic verdicts indexed to times, we can say the following. If the bride is bound by a direct-rule prohibiting promise-breaking, then she appears to disobey this rule at midday when she is not at her wedding (and not beforehand), as that is when she breaks her promise. However, if she is bound by a rule prohibiting 'acting such that you break a promise' then she appears to disobey this rule at 11:30am as that is when she performs an action (boarding the plane) that is inconsistent with her being at the midday wedding. This difference in the timing of wrongdoing given by each rule appears to show that they are not equivalent. However, I now think there is a good response to this counterexample. To get equivalence we just need to temporally index the act-such-that rule as follows: 'Each agent, at each moment, must act such that she does not break her promises at that moment'. This rule gives the same temporally located deontic verdict as the direct-rule in the runaway bride case. (I thank Ben Blumson for suggesting to me that temporally indexing the act-such-that rule can resolve this counterexample).

If (2°) is an adequate translation of (2), then whenever Clark Kent obeys (2) he must also be obeying (2°). In the Super Crime-Fighter this is indeed the verdict we get. When Clark Kent prevents as many killings of innocent people as he can, and yet fails to prevent every such killing, he is bringing it about to the greatest degree that he can that there is minimal killing of innocent people in this world. Reflection on further cases shows that (2) and (2°) consistently give the same verdicts.[28] Furthermore, like (2*) and (2′), (2°) contains a proposition in square brackets that can be used to explain (2)'s status as agent-neutral. Thus, we appear to have found a way of translating maximizing rules into semiformal notation such that they can correctly be classified as agent-neutral or agent-relative by a revised version of the McNaughton and Rawling account.

Given the revised translations I have defended above for direct-rules and maximizing rules, I can now present a revised version of McNaughton and Rawling's account of the agent-relative/agent-neutral distinction:

> A rule expressing a moral requirement is agent-relative *iff*, when the rule is represented with the form '$(x)(x$ must {verb} that [...])', there is an ineliminable occurrence of 'x' in the square brackets that is bound by the initial universal quantifier.
>
> A rule expressing a moral requirement is agent-neutral *iff* it is not agent-relative.

We can apply this revised account of the distinction to the ten paradigm rules from §3. Here are the ten rules again for reference:

> (1) Each agent must not kill innocent people.
>
> (2) Each agent must minimize the killing of innocent people.
>
> (3) Each agent must care for her children.
>
> (4) Each agent must ensure that children are cared for by their parents.
>
> (5) Each agent must maximize her own wellbeing.
>
> (6) Each agent must maximize general wellbeing.
>
> (7) Each agent must maximize the wellbeing of her friends.
>
> (8) Each agent must ensure that no one is disloyal to the Supreme Leader.
>
> (9) Each agent must minimize her violations of rule (1).

---

[28] One possible exception is cases discussed at the end of §6 where an agent unintentionally maximizes some quantity. Some may interpret 'bring it about' as an intentional concept and thus claim that an agent who unintentionally maximizes a quantity does not bring it about to the greatest degree she can that there is a maximal amount of that quantity. I do not find this reading of 'bringing it about' intuitive (I think it is possible to bring something about without intending to do so). Nonetheless, for those who disagree with me, the account I offer can be modified to accommodate their intuitions. Rule (2) can be translated as: $(x)(x$ must, to the greatest degree she can, act such that [In this world there is minimal killing of innocent people]).

(10) Each agent must minimize general violations of rule (1).

Here are my suggested translations of these rules into the semiformal notation:

(1°) $(x)(x$ must act such that [x does not kill innocent people])

(2°) $(x)(x$ must bring it about to the greatest degree x can that [There is minimal killing of innocent people])

(3°) $(x)(x$ must act such that $[(y)(y$ is x's child $\rightarrow$ x cares for $y)]$)

(4°) $(x)(x$ must ensure that $[(y)(z)(y$ is a child of $z \rightarrow$ z cares for $y)]$)

(5°) $(x)(x$ must bring it about to the greatest degree x can that [There is maximal wellbeing in x's life])

(6°) $(x)(x$ must bring it about to the greatest degree x can that [There is maximal wellbeing in the world])

(7°) $(x)($ x must bring it about to the greatest degree x can that $[(y)(y$ is x's friend $\rightarrow$ there is maximal wellbeing in y's life)])

(8°) $(x)(x$ must ensure that $[(y)(y$ is not disloyal to the Supreme Leader)

(9°) $(x)(x$ must bring it about to the greatest degree x can that [x does not violate rule (1)])

(10°) $(x)($ x must bring it about to the greatest degree x can that $[(y)(y$ does not violate rule (1))])

In these translations all the agent-relative rules contain an ineliminable occurrence of 'x' in the square brackets that is bound by the initial universal quantifier. By contrast, all the agent-neutral rules do not contain an occurrence of 'x' in the square brackets. Furthermore, unlike rules (1*)-(6*) considered above, rules (1°)-(10°) are accurate translations of the ordinary language rules (1)-(10) that they are based on. Thus, this revised account of the distinction avoids the two objections I raised in §6. It also does well at satisfying the four desiderata outlined in §4. At first glance, it may not appear to satisfy the fourth desideratum of simplicity as well as the original McNaughton and Rawling account. For, it uses several verb phrases where their account got by with one. However, simplicity is only desirable when all else is equal, not when it is purchased at the expense of accuracy. Thus, we cannot hold it against the revised account that, in order to make more accurate translations, it introduces more complexity. From this, I conclude that the revised version of McNaughton and Rawling's original proposal is the best account of the agent-relative/agent-neutral distinction for rules expressing moral requirements.

8. Rules Beyond Requirement

So far, I have shown how to apply the agent-relative/agent-neutral distinction to rules that express moral requirements. However, some rules express moral permissions rather than

moral requirements. Such rules permit agents in certain circumstances to perform actions that are otherwise barred by general moral requirements. By analogy with agent-relative requirements, some hold that moral permissions can be agent-relative. For such permissions often recognize an agent's special concern with her own actions and interests, allowing her to give them priority over the actions and interests of others. When permissions do this it seems apt to describe them as agent-relative.[29]

McNaughton and Rawling (1995, p. 36) show how to extend their account of agent-relative requirements to agent-relative permissions. They take moral permissions to be riders on moral requirements. For example, perhaps there is an agent-neutral requirement to maximize wellbeing, along the lines of (6) above:

> (6) Each agent must maximize general wellbeing.

A possible moral permission related to (6) might be:

> (12) Each agent is permitted to not maximize general wellbeing in circumstances where maximizing it would lead to her interests or projects being seriously compromised.

Rule (12) appears to be agent-relative in the sense that it allows each agent to give priority to her own interests and projects over and above the interests and projects of others. McNaughton and Rawling suggest translating (12) into semiformal notation as follows:

> (12*) (x)([x ensures that [(y)(y's wellbeing is high)] → x's interests/projects are seriously compromised] → [It is not the case that x must ensure that [(y)(y's wellbeing is high)]])

More generally, they argue that all moral permissions can be put into the following canonical form:

> (x)([x ensures that [...] → x's interests/projects are seriously compromised] → [It is not the case that x must ensure that [...]]).

With this translation at hand they then suggest that a rule expressing a moral permission is agent-relative if and only if, when it is put into the canonical form, the consequent of the first conditional sentence contains an ineliminable occurrence of the variable '*x*' that is bound by the initial universal quantifier. This suggestion appears promising, as the consequent of the first conditional contains a clause specifying the circumstances where the general obligation is overridden by the permission. Thus, if this clause contains an

---

[29] See especially Scheffler (1982) and Kagan (1989) for discussion of the idea that these kinds of moral permissions have the same kind of special agent focus as agent-relative requirements.

ineliminable occurrence of '*x*' that is bound by the initial universal quantifier, then it makes essential reference to the agent who is bound by the moral requirement, and in doing so appears agent-relative. However, there are two revisions that I think make their account of agent-relative permission more plausible. First, as we saw in §7, not all rules can be represented as ensuring-rules and thus the canonical form should stay neutral on this part of the form, replacing 'ensure' with the placeholder '{*V*}'. Second, in presenting their general form, McNaughton and Rawling take it for granted that all moral permissions are related to demands on an agent's interests and projects. This understanding of permissions is widely held. It also makes it trivially true that all moral permissions are agent-relative. However, to preserve theoretical neutrality it is best to avoid such assumptions, leaving open what content is contained in the relevant clause. With these revisions in mind I suggest revising McNaughton and Rawling's canonical form as follows:

$$(x)((x \ \{V\} \ \text{that} \ p \rightarrow \ [...]) \rightarrow \ (\text{It is not the case that} \ x \ \text{must} \ \{V\} \ \text{that} \ p)).$$

All moral permissions can be translated into this form. We can then define agent-relativity/neutrality for moral permissions as follows:

> A rule expressing a moral permission is agent-relative *iff*, when the rule is represented with the form '$(x)((x \ \{V\} \ \text{that} \ p \rightarrow [\ ... \ ]) \rightarrow (\text{It is not the case that} \ x \ \text{must} \ \{V\} \ \text{that} \ p))$', there is an ineliminable occurrence of '*x*' in the square brackets that is bound by the initial universal quantifier.

> A rule expressing a moral permission is agent-neutral *iff* it is not agent-relative.

Rule (12) above counts as agent-relative because, when translated into the canonical form, there is an ineliminable occurrence of the variable '*x*' in the square brackets:

> (12°) $(x)((x$ brings it about to the greatest degree x can that (There is maximal wellbeing in the world)) $\rightarrow [x$'s interests/projects are seriously compromised]) $\rightarrow$ (It is not the case that x must bring it about to the greatest degree x can that (There is maximal wellbeing in the world))).

By contrast, let's consider another moral permission that is also a rider on (6):

> (13) Each agent is permitted to not maximize general wellbeing in circumstances where maximizing it would lead to the interests or projects of the Supreme Leader being seriously compromised.

Rule (13) is an agent-neutral moral permission because it does not allow an agent to give special priority to her own actions or interests.[30] This rule can be translated into the relevant form as follows:

> (13°) $(x)((x$ brings it about to the greatest degree x can that (There is maximal wellbeing in the world)) → [The Supreme Leader's interests/projects are seriously compromised]) → (It is not the case that x must bring it about to the greatest degree x can that (There is maximal wellbeing in the world))).

We can see that there is no occurrence of the variable 'x' in the square brackets and thus this rule is classified as agent-neutral by the definition above.

Finally, let's consider two more possible moral permissions that are also riders on (6), and yet are substantially different from (12) and (13):

> (14) Each agent is permitted to not maximize general wellbeing in circumstances where maximizing it would require her to kill an innocent person.

> (15) Each agent is permitted to not maximize general wellbeing in circumstances where maximizing it would result in an innocent person being killed.

These two rules are similar. However, (14) only gives an agent the permission to not comply with (6) in circumstances where compliance requires that *she* gets her hands dirty and thus gives her a special focus on her own actions. By contrast, (15) gives an agent permission to not comply with (6) in *any* circumstances where compliance results in an innocent person being killed. Thus, (15) gives each agent an equal concern with acts of killing committed by anyone rather than a special focus on her own acts. On this basis, (14) can be classed as agent-relative and (15) as agent-neutral. This is born out when each rule is translated into the canonical form.

> (14°) $(x)((x$ brings it about to the greatest degree x can that (There is maximal wellbeing in the world)) → [x kills an innocent person]) → (It is not the case that x must bring it about to the greatest degree x can that (There is maximal wellbeing in the world))).

> (15°) $(x)((x$ brings it about to the greatest degree x can that (There is maximal wellbeing in the world)) → [An innocent person is killed]) → (It is not the case

---

[30] Of course, sometimes an agent's interests or projects will correspond to the content of an agent-neutral clause. For example, perhaps Mark happens to have a personal interest in the wellbeing of the Supreme Leader and so rule (13), in giving him permission to prioritize the Supreme Leader's wellbeing, is giving him permission to prioritize his own personal interest. However, the latter permission is only contingent and thus does not count as the right kind of special priority. By contrast, rule (12) gives Mark permission to prioritize his own interests and projects necessarily, regardless of what those interests are, and thereby counts as agent-relative.

that x must bring it about to the greatest degree x can that (There is maximal wellbeing in the world))).

Rule (14°) has an ineliminable occurrence of 'x' in the square brackets bound by the initial universal quantifier, whereas rule (15°) does not have such an occurrence. Thus, by the account defended here, (14°) is agent-relative and (15°) is agent-neutral.

9. Extending the Distinction to Reasons and Theories

Above I have defended a revised version of McNaughton and Rawling's account of the agent-relative/agent-neutral distinction that applies the distinction to moral rules. I will now examine whether their account can also be extended to moral theories and reasons. A simple idea for extending the distinction to moral theories is to take a theory's status as agent-relative or agent-neutral to depend on the status of the rules it contains. This idea can be drawn from the discussion of Parfit (1984) (although Parfit's preferred way of applying the distinction to theories is to appeal to the ultimate aims a theory gives). The idea can be developed as follows:

> A moral theory is agent-neutral *iff* all the moral rules it contains are agent-neutral.

> A moral theory is agent-relative *iff* it contains at least one agent-relative rule.

When these biconditionals are combined with the account of agent-relative/neutral rules that I have defended above we have a complete account of what it is for a moral theory to be agent-neutral or agent-relative. Furthermore, the classification of moral theories produced by this account match those that are generally accepted. For example, classical utilitarianism is widely regarded as an agent-neutral theory, and this account classifies it as agent-neutral because its sole rule—'maximize total net happiness'—is agent-neutral. On the other hand, deontological moral theories are widely regarded as agent-relative theories, even though they often contain agent-neutral rules such as a general duty of beneficence. This account classifies them as agent-relative because, although they contain agent-neutral rules, they also contain at least one agent-relative rule—namely at least one agent-centered constraint. However, one might question why the standard classifications are asymmetrical in this way. Why not instead make the distinction symmetrical, holding that: (i) an agent-neutral theory contains only agent-neutral rules/reasons, (ii) an agent-relative theory contains only agent-relative rules/reasons, and (iii) a theory that mixes neutral and relative rules/reasons is a 'hybrid theory'? When Parfit (1984) first applied the agent-relative/neutral distinction to moral theories he made the distinction asymmetrically but never explained why he was doing it this way rather than the

alternative. However, I think a good answer is available. It is widely accepted that any plausible moral theory will endorse at least some agent-neutral rules. What moral philosophers tend to disagree on is whether the best candidate moral theories will also contain agent-relative rules. Thus, applying the distinction to theories in this asymmetrical way has the virtue of capturing the distinction that philosophers are most interested in when it comes to examining moral theories.

Applying the agent-relative/agent-neutral distinction to reasons is much more complicated than applying it to theories. Indeed, McNaughton and Rawling (1991) developed an account of the distinction that applies it to moral rules because they found various accounts applying the distinction to reasons to be unsatisfactory. I tend to agree with McNaughton and Rawling, which is one of the reasons I have followed them in focusing on the rule-based distinction. Furthermore, in the various chapters of this thesis, while I often refer to agent-relative rules, theories, and values, I do not find it necessary to appeal to agent-relative reasons. Thus extending the account to reasons is not something I need to do for the purposes of this thesis. However, it is worth briefly reviewing a recent attempt by Douglas Portmore to extend McNaughton and Rawling's account to reasons.[31] Portmore (2014) argues that, letting '$x\varphi cEp$' stand for 'x's φ-ing in circumstances c would ensure that p', we can define agent-relative and agent-neutral reasons as follows:

> $(x)(\varphi)(c)(p)$(if the fact that $x\varphi cEp$ constitutes a reason for x to φ in c, then this fact constitutes an agent-relative reason for x to φ in c if and only if p contains an essential reference to x, and any reason for x to φ in c that doesn't constitute an agent-relative reason for x to φ in c constitutes an agent-neutral reason for x to φ in c).[32]

Portmore's suggestion appears promising and deserves further discussion and analysis. However, doing so would take me too far afield from my present purpose of explaining and defending my use the agent-relative/agent-neutral distinction in this thesis.

Finally, I should point out that I will not extend McNaughton and Rawling's account of agent-relative/agent-neutral rules to moral values. In fact, I do not think it is possible to 'extend' their account in this way. Instead, agent-relative/neutral values need to be

---

[31] One reason why we might want a reasons-based account of the distinction is that such an account can be used even by a moral particularist who rejects all generalized moral rules (see Ridge 2005a). While I can appreciate this motivation, the question still remains of how to apply the distinction to moral rules, and that is the question I find it most useful to focus on.

[32] Portmore further clarifies that "The statement p contains an essential reference to x if and only if there is no non-x referring statement q such that, for all x, the world in which x ensures that q is identical to the world in which x ensures that p." This use of 'essential' matches my use of 'ineliminable' above.

accounted for separately. I will offer such an account in the final chapter of this thesis where I defend agent-relative value.

10. Applying the Apparatus to Moments

So far, I have defended a revised version of McNaughton and Rawling's account of the agent-relative/agent-neutral distinction and extended it in various ways. In the literature discussing agent-relativity and agent-neutrality, it is common to make a further distinction between moment-relativity and moment-neutrality.[33] Like the agent-relative distinction, this further distinction can also be applied to rules, reasons, theories, and values. I will now show how to use the formal apparatus I explicated above to make a distinction between moment-relativity and moment-neutrality. As I did above, I will focus on applying the distinction to rules expressing moral requirements. Consider the following two rules:

> (16) Each agent must, at each moment in time, not break her promises at that moment.

> (17) Each agent must minimize her overall promise-breaking.

Each of these rules is concerned with the act of promise-breaking, and each is an example of an agent-centered constraint on promise-breaking. Such constraints prohibit an agent from performing an action even when it is the only way to prevent more instances of that action being performed by others. Thus, both rules give each agent a special concern with her own promise-breaking, over and above the promise-breakings of others. However, rule (16) does something extra that rule (17) does not do. It also gives each agent a special concern with the present-moment. By contrast rule (17) gives each agent a general concern with all moments. On these grounds (16) is typically classed as moment-relative whereas (17) is typically classed as moment-neutral. To see this difference more clearly consider a case where an agent must choose between breaking one promise at the present-moment and breaking several promises at some future moment:

> Sita has made comparable promises to three colleagues—Ash, Bart, and Chang. The promise owed to Chang is due now whereas the promises owed to Ash and Bart are due tomorrow. However, if Sita fulfils the promise made to Chang now, she will be unable to keep the promises made to Ash and Bart tomorrow. Only by breaking the present promise owed to Chang can she keep those other two promises.

---

[33] For example, see Parfit (1984), Broome (1991), Kamm (1992/1996), Ridge (2001), Smith (2003/2009), Louise (2004), and Portmore (2011). The terminology 'moment-relative' and 'moment-neutral' comes from Broome (1991).

What should Sita do in this case? According to (16), at each moment she must keep whatever promises are due at that moment. Thus, she must keep her promise to Chang at the present-moment regardless of the consequences this may have for her future promise-keeping. According to (17), Sita must minimize her total promise-breaking. In the above example she does this by breaking her promise to Chang which will allow her to keep her promises to Ash and Bart, as this will result in less promise-breaking by her overall.

We can draw this distinction between the moment-relative (16) and moment-neutral (17) more formally using the apparatus from above. To do this we need to introduce another formal tool—a time variable 't' ranging over all possible moments in time. We then translate all rules into the following canonical form:

$(x)(t)(x$ must at t $\{v\}$ that [...])

When translated into this form, rules that are moment-relative will have an ineliminable occurrence of the time variable 't' in the square brackets that is bound by the initial universal temporal quantifier. By contrast, moment-neutral rules will not have such an occurrence. For example, (17) is accurately translated into the canonical form as follows:

(17*) $(x)(t)(x$ must at t act such that [x does not break a promise at t])

In (17*) there is an ineliminable occurrence of the variable 't' in the square brackets that is bound by the initial universal temporal quantifier, and thus (17*) is counted as moment-neutral by our definition. By contrast, (16) is accurately translated into the canonical form as follows:

(16*) $(x)(t)(x$ must at t bring it about to the greatest degree x can that [There is minimal promise breaking by x])

In (16*) an instance of the time variable 't' does not occur within the square brackets and thus this rule is moment-neutral. Importantly, (16*) is equivalent to an alternative formulation of rule (16) that eschews time variables altogether:

(16**) $(x)(x$ must bring it about to the greatest degree x can that [There is minimal promise breaking by x])

These two translations are equivalent because requiring an agent to minimize her promise-breaking in general is the same as requiring her to minimize her general promise-breaking at each moment when she acts. Generalising this, a rule requiring 'A to φ simpliciter' is equivalent to a rule requiring 'A to φ at all times'. Thus, all rules (including all those that are moment-neutral) can be translated into the canonical temporal form

without their content changing. In this way, we can extend the revised McNaughton and Rawling account to cover moment-relative moral requirements. We can also extend this account of moment-relativity to moral permissions, moral theories, and moral reasons by following the methods employed above in §8 and §9.

11. Parfit's Distinction

Now that I have defended a formal account of the agent-relative/agent-neutral distinction, and extended this account in various ways, I want to go back to the intuitive account of the distinction presented by Parfit (1979/1984). I commented earlier that some find Parfit's intuitive account limited because it lacks the precision of a formal account. Parfit's accounts relies on our intuitions about what the ultimate aim of any particular rule is to determine its status as agent-relative or agent-neutral. By contrast, a formal account, like the one defended here, relies on the presence or absence of some specified property in a formal representation of a rule to determine whether that rule is agent-neutral or agent-relative. Even if it is true that formal accounts are generally preferable to intuitive accounts, it still possible to complement a formal account with an intuitive account. This is especially helpful if the intuitive account is easy to grasp and apply to complex cases, as I think Parfit's intuitive account is.

Nonetheless, Parfit's intuitive account is controversial because it relies on the idea that moral rules (and moral theories) give agents aims, and many wonder in what sense, if any, this may be true. For example, Portmore (2014) says:

> But this talk of aims is at best metaphorical… Moral theories provide criteria for determining whether an act has a certain deontic status (e.g., permissible, impermissible, obligatory, or supererogatory), but they don't explicitly state which aims agents ought to have. And it's not clear how we are to move from talk about an act's having a certain deontic status to there being a certain aim that its agent ought to have. Suppose, for instance, that a moral theory holds that Smith's giving at least ten percent of his income to charity is obligatory and that Smith's giving more than ten percent is supererogatory. Does such a moral theory give Smith the aim of giving ten percent, giving more than ten percent, or something else? It's not clear.

I think Portmore is correct that Parfit's talk of aims is at best metaphorical. For many moral rules do not appear to give an agent an aim in any literal sense. To see this let's consider rule (1) from above:

(1) Each agent must not kill innocent people.

Recall from §3 that Parfit's account classifies this rule as agent-relative because he takes it to give each agent the aim that *she* does not kill, thereby giving different aims to different agents. However, consider the following case:

> **Unconscientious conformity:** Lennie has never considered the possibility that morality requires him to not kill innocent people. On many occasions during his life, Lennie has both the ability and the opportunity to kill an innocent person. However, Lennie does not kill on any of these occasions, not because of any intention he has not to kill, but simply because on each occasion the thought of killing never appeals to him.

In the case described, Lennie does not possess the aim that he does not kill innocent people. However, despite lacking this aim he still obeys rule (1) because he does not kill an innocent person at any point during his life. If an agent can obey rule (1) without aiming at not killing innocent people then it is difficult to see how (1) can be said to 'give' that agent the aim of not killing. For there appear to be no plausible senses of 'give' and 'aim' that have this result. Thus, if we read Parfit's talk of 'aims' literally, then we must reject his assumption that rules like (1) give agents aims.[34]

Nonetheless, despite this shortcoming, Parfit's account of the distinction might still be pragmatically useful. For when a moral rule fails to literally give an agent an aim, we tend to metaphorically ascribe a rule-based aim to the agent by looking for the state of affairs that the rule is concerned with and treating that as its aim. Of course, different moral rules are concerned with states of affairs in different ways. Some rules are about bringing about states of affairs, others about ensuring that they obtain, and others about acting such that they obtain. Nonetheless, if we always treat such states of affairs as the rule's 'aim' then, following Parfit's method, we get an intuitive account of the distinction which gives the same classifications of cases as the formal account I defended above. For, when a rule translated into the semiformal notation contains an ineliminable occurrence of the variable '*x*' in the square brackets, the state of affairs represented in the square brackets will vary from agent to agent as the '*x*' is filled in with each agent's identity. And when there is no such occurrence of the variable, the state of affairs picked out will be the same for each agent.[35] Of course, this only works with rules that express moral requirements.

---

[34] Interestingly, some moral rules fit better with Parfit's idea that they 'give agents aims'. For example consider rule (11) from above: "Each agent must ensure that she does not kill innocent people". On certain understandings of 'ensure' an agent cannot ensure that p without explicitly intending, or aiming at p. Thus, according to one interpretation of 'ensuring' it is true that (11) gives agents the aim of not killing.

[35] Or, at least, there will be a way of picking out that state of affairs such that it is the same for each agent. Recall from §5 that any rule can trivially be formulated such that the state of affairs it is concerned with varies from agent to agent.

Portmore is correct that when we move to rules concerning other deontic properties it makes no sense to ask (even metaphorically) what aim that rule gives to agents. However, I think I have shown how Parfit's account of the distinction can be useful if we want an intuitive and easy to grasp method for applying the agent-relative/agent-neutral distinction to moral requirements. Thus, I will sometimes employ Parfit's account of the distinction in this thesis when I am classifying various moral requirements as agent-relative or agent-neutral and it is too tedious to use the formal account of the distinction I have defended above.

References

Belnap, N. & Perloff, M. (1988). Seeing to it That: A Canonical Form for Agentives. *Theoria*, 54: 175-199.

Broad, C.D. (1942) Certain Features in Moore's Ethical Doctrines. In P.A. Schilpp (ed.) *The Philosophy of G.E.Moore*, La Salle: Open Court.

Broome, J. (1995). Skorupski on Agent-Neutrality. *Utilitas*, 7: 315.

Broome, J. (1999). Normative Requirements. *Ratio*, 12: 398–419.

Dougherty, T. (2013). Agent-Neutral Deontology. *Philosophical Studies*, 163: 527-537.

Dreier, J. (1993). Structures of Normative Theories. *The Monist*, 76: 22-40.

Hammerton, M. (2019). Distinguishing Agent-Relativity from Agent-Neutrality. *Australasian Journal of Philosophy*, 97: 239-250.

Howard-Snyder, F. (1996). A New Argument for Consequentialism? A Reply to Sinnott-Armstrong. *Analysis*, 56: 111-115.

Hurka, T. (2003). Moore in the Middle. *Ethics*, 113: 599-628.

Kagan, S. (1989). *The Limits of Morality*. Oxford: Clarendon Press.

Kamm, F. M. (1992). Review: Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status. *Philosophy and Public Affairs*, 21: 354-380.

Kamm, F. M. (1996). *Morality, Mortality: Volume II: Rights, Duties, and Status*. Oxford University Press.

Louise, J. (2004). Relativity of Value and the Consequentialist Umbrella. *Philosophical Quarterly*, 54: 518-536.

McNaughton, D. & Rawling, P. (1991). Agent-Relativity and the Doing-Happening Distinction. *Philosophical Studies*, 63: 167-185.

McNaughton, D. & Rawling, P. (1992). Honoring and Promoting Values. *Ethics*, 102: 835-843.

McNaughton, D. & Rawling, P. (1993). Deontology and Agency. *The Monist*, 76: 81-100.

McNaughton, D. & Rawling, P. (1995a). Value and Agent-Relative Reasons. *Utilitas*, 7: 31-47.

McNaughton, D. & Rawling, P. (1995b). Agent-Relativity and Terminological Inexactitudes. *Utilitas*, 7: 319-325.

McNaughton, D. & Rawling, P. (1998). On Defending Deontology. *Ratio*, 11: 37-54.

McNaughton, D. & Rawling, P. (2002). Conditional and Conditioned Reasons. *Utilitas*, 14: 240-248.

Nagel, T. (1970). *The Possibility of Altruism*. Oxford: Clarendon Press.

Nagel, T. (1986). *The View from Nowhere*. New York: Oxford University Press.

Parfit, D. (1979). Prudence, Morality and the Prisoner's Dilemma. *Proceedings of the British Academy*, 65: 539-564.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Pettit, P. (1987). Universality without Utilitarianism. *Mind*, 72: 74-82.

Portmore, D. (2001). McNaughton and Rawling on the Agent-relative/Agent-neutral Distinction. *Utilitas*, 13: 350-356.

Portmore, D. (2014). Agent-relative vs. Agent-neutral. In Hugh LaFollette (ed.), *International Encyclopedia of Ethics*, Wiley Blackwell.

Postow, B. C. (1997). Agent-Neutral Reasons: Are they for Everyone? *Utilitas*, 9: 249-257.

Ridge, M. (2001). Agent-Neutral Consequentialism from the Inside-Out: Concern For Integrity without Self-Indulgence. *Utilitas* 13: 236-254.

Ridge, M. (2005a). Agent-Neutral vs. Agent-Relative Reason. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (Ed.), URL = http://plato.stanford.edu/archives/win2005/entries/reasons-agent/

Ridge, M. (2005b). Universalizability for Collective Rational Agents: A Critique of Agent-relativism. *Philosophy and Phenomenological Research*, 70: 34-66.

Rønnow-Rasmussen, T. (2009). Normative Reasons and the Agent-Neutral/Relative Dichotomy. *Philosophia*, 37: 227-243.

Scheffler, S. (1982). *The Rejection of Consequentialism*. Oxford: Clarendon Press.

Schroeder, M. (2007). Teleology, Agent-relative Value, and 'Good'. *Ethics*. 117: 265-295.

Schroeder, M. (2011). Ought, Agents, and Actions. *Philosophical Review*, 120: 1-41.

Skorupski, J. (1995). Agent-neutrality, Consequentialism, Utilitarianism: a Terminological Note. *Utilitas*, 7: 49-54.

Skorupski, J. (1996). Neutral Versus Relative: A Reply to Broome, and McNaughton and Rawling. *Utilitas*, 8: 235-248.

Smith, M. (2003). Neutral and Relative Value after Moore. *Ethics*, 113: 576-598.

Smith, M. (2009). Two Types of Consequentialism. *Philosophical Issues*, 19: 257-272.

Thomson, J. J. (1993). Goodness and Utilitarianism. *Proceedings of and Addresses of the American Philosophical Association*, 67: 145-159.

Wedgwood, R. (2006). The Meaning of 'Ought'. *Oxford Studies in Metaethics*, 1: 127–160.

2

Patient-Relativity in Morality[1]

1. Introduction

The distinction between an agent-relative reason, rule, or value, and an agent-neutral reason, rule, or value, is widely recognized as one of the most important distinctions in normative ethics. A second related distinction is that between a moment-relative reason, rule, or value and a moment-neutral reason, rule, or value. Although this distinction is not as prominent as the agent-neutral/relative distinction, it is also important and has been discussed by several philosophers. In this chapter I introduce a third distinction that stands alongside these two distinctions—the distinction between a patient-relative reason, rule, or value and a patient-neutral reason, rule, or value. This distinction is based on the observation that the numerical identity of a moral patient may be a morally significant factor. In some contexts the moral significance of patient identity has already been noted. However, a sustained and systematic account of this phenomenon has not been provided. I offer such an account in this chapter. Furthermore, I show how this account leads to a distinction between patient-relativity and patient-neutrality. Recognising the possibility of patient-relativity is important because it plays a key role in several moral theories, gives us a better understanding of how agent-relativity and moment-relativity interact, and provides a novel objection to Derek Parfit's 'appeal to full relativity' argument.

Here is how the rest of this chapter will proceed. In §2 I will explain the agent-neutral/agent-relative distinction and in §3 I will explain the moment-neutral/moment-relative distinction. This provides a helpful background for the discussion that follows in §4. In §4.1 I introduce the patient-relative/patient-neutral distinction by presenting several cases where a patient's numerical identity appears to make a difference to our

moral obligations. I follow this in §4.2 with a discussion of the logical structure of patient-relativity and its relation to agent-relativity and moment-relativity. Finally, in §4.3 I show how several moral theories may be committed to patient-relative rules. I then show how the possibility of patient-relativity provides a response to Derek Parfit's 'appeal to full relativity' argument.

2. Agent-Neutrality and Agent-Relativity

The agent-neutral/agent-relative distinction was brought to the attention of contemporary philosophers by Nagel (1970), although it was not until Parfit (1979/1984) that the distinction came to be described using the terminology 'agent-neutral' and 'agent-relative'. Since Nagel's initial discussion, many different accounts of the distinction have been put forward.[2] Debates about how to formulate the distinction are complicated by the fact that it can be applied to at least four different categories—reasons, rules, theories, and values. Which of these categories, if any, should take priority in explicating the distinction is often disputed. Importantly, the participants in these debates agree they are talking about the same distinction and give roughly the same assessment of various cases as agent-neutral or agent-relative. What they disagree about is the optimal way of cashing the distinction out.

In this chapter I will not take a stand on which account of the distinction is best. In the previous chapter I defended a revised version of McNaughton and Rawling's account of the distinction. However, the arguments I offer in this chapter have the virtue of being compatible with all the leading accounts of the distinction. Thus, to proceed with these arguments, I do not need to assume that the account I defended in chapter 1 is correct. Nonetheless, in order to present the ideas of this chapter, it is helpful to make use of an account of the distinction. I will use Parfit's intuitive account because it is the easiest to follow when considering complex arguments. Furthermore, I will focus on the distinction made in terms of moral rules. When I make claims in this chapter about certain rules being agent-relative or agent-neutral the reader should keep the following two things in mind. First, although I might explain these claims by appeal to Parfit's account, other leading accounts also support them. Second, although I only make claims about rules, corresponding claims about reasons, theories, and values can also be made.

With this in mind, let's review Parfit's account. According to Parfit (1984) an agent-relative rule is a rule that gives different agents different aims, whereas an agent-neutral rule is a

---

[2] Ridge (2005), and Portmore (2014) provide useful overviews of different accounts of the distinction. See also chapter 1 of this thesis.

rule that gives all agents the same aim. As an example, consider a moral rule that prohibits the killing of innocent people which I will call 'DK', short for 'Do not kill':

DK: Each agent must not kill innocent people.

DK is an agent-relative rule because it gives each agent a different aim—the aim that he or she does not kill. Thus, DK gives Alfred the aim that 'Alfred does not kill', and Sita the aim that 'Sita does not kill'. DK can be usefully contrasted with a rule that instructs us to *minimize* the killing of innocent people. Let's call this rule 'MK', short for 'minimize killing':

MK: Each agent must minimize the killing of innocent people.

MK is agent-neutral because it gives each agent the same aim. Alfred's aim is that there is as little overall killing as possible, and Sita's aim is the same.[3] This contrast is most vivid in cases where, by killing an innocent person, an agent can prevent more killings by others. Suppose the Mafia make a credible threat to Alfred that if he does not kill a random innocent person they will kill several innocent people. What should Alfred do? According to DK, Alfred must *not* kill the innocent person even though that will mean there is more killing overall. This is because DK gives Alfred a special concern with his own killing, requiring him to give it priority over the killings that others might commit. According to MK, Alfred *must* kill the innocent person. For, in this situation, doing so will result in fewer killings overall and MK requires Alfred to give equal consideration to any killings that may occur whether or not he is responsible for them.

Whatever account of the agent-neutral/agent-relative distinction one accepts, there appears to be widespread agreement on its importance. Thomas Hurka describes it as one of "the greatest contributions of recent ethics".[4] Several others have claimed that it gives a precise and illuminating account of the distinction between consequentialism and non-

---

[3] Two clarifications are necessary here. First, when Parfit talks of agents having the same aim, he is referring to "ultimate aims" and not "derived aims". If Alfred and Sita are in different circumstances, they might find that minimizing killing requires them each to do different things. Thus, MK gives them different aims. However, these different aims are derived from the same ultimate aim that they each share—the aim of minimizing killing. Second, one might interpret MK as giving different agents different ultimate aims. Perhaps it gives Alfred the ultimate aim that "Alfred minimizes killing" and Sita the ultimate aim that "Sita minimizes killing". However, as McNaughton and Rawling (1991), and Portmore (2014) note, the distinction can still be made even if such an interpretation of MK is legitimate. For, although this interpretation counts MK as giving different agents different ultimate aims, there is also an equally valid interpretation of MK where it gives different agents the same aim (i.e., the aim of minimizing killing). However, there is no plausible interpretation of DK where different agents have the same ultimate aim. Thus, strictly speaking, the distinction is stated as follows: A rule is agent-neutral if and only if there is a valid interpretation of it that gives all agents the same ultimate aim, and is otherwise agent-relative.

[4] Hurka (2003), p.628.

consequentialism.[5] Finally, James Dreier has argued that it is distinct from, yet more fundamental to moral theory than, the distinction between consequentialism and non-consequentialism.[6]

3. Moment-Neutrality and Moment-Relativity

The next distinction goes by many different names. Broome (1991) uses the terms 'moment-neutral' and 'moment-relative' to mark it. Ridge (2001) employs the terms 'diachronic' and 'synchronic'. Louise (2004) prefers the terms 'temporal-neutral' and 'temporal-relative'. Finally, Kamm (1992) describes this kind of relativity as the 'agent-centered and time-slice approach to persons' and 'deontology of the present moment'. I will use Broome's terminology of 'moment-neutral' and 'moment-relative', although nothing important hangs on the terminology used. The best way to see this distinction clearly is to consider an alternative version of the Mafia case described above. Suppose that Alfred is not in a situation where the Mafia will kill several if he fails to kill one. Instead, he is in a situation where it is his own future self who will kill several if his current self fails to kill one. Perhaps Alfred finds his mind slowly being overtaken by homicidal desires and correctly judges that, unless he kills one now, his weak will, and morally corrupted desires will soon lead him to kill several. According to the agent-relative rule DK, Alfred must not kill innocent people. However, in this scenario where he has to choose between killing one now and killing several later what should he do? One possible moral rule (let's call it DK′) requires him to give equal consideration to all of his potential killings and prevent as many of them as possible. This would require Alfred to kill the one now to prevent his later killing of several. Such a rule is moment-neutral because it does not give priority to any particular moment in time such as the present. An alternative moral rule (let's call it DK″) requires Alfred to give special priority to any killing he is about to perform, and thus to not kill the one now. Such a rule is moment-relative because it gives special priority to the present moment. We can also state the difference between these two rules using Parfit's talk of 'aims' (although Parfit himself never does this). DK′ is moment-neutral because it gives different time-slices of the same agent the *same* aim (the aim that the agent kills as little as possible). DK″ is moment-relative because it gives different time-slices of the same agent *different* aims (the aim that the time-slice does not kill at the moment it inhabits).

---

[5] For example, see McNaughton and Rawling (1992), and Pettit (1997).
[6] See Dreier (1993), p.39.

In the example described above the moment-relative rule was also an agent-relative rule (the rule both gives different agents different aims, and different time-slices of the same agent different aims). However, moment-relative rules can also be agent-neutral. For example, Broome (1991, p.8) describes a possible moral rule that appears to be agent-neutral yet moment-relative. According to this rule when a miner is trapped in a mine everyone is required to perform actions (e.g. contributing resources to the rescue effort) that will maximize the miner's chance of being saved. Stated thus the rule appears agent-neutral (it gives all agents the same aim). However, Broome's rule also has a caveat requiring agents to perform any action which contributes to saving miners who are currently trapped, even if doing so will mean that more trapped miners in the future will go unsaved. Thus, we must not hold resources back from current efforts even if doing so would allow more miners to be saved in the future. This caveat does not alter the rule's agent-neutral status (everyone has the same aim at any one time) but does appear to make the rule moment-relative (our different time-slices are given different aims).

So it appears that the moment-neutral/moment-relative distinction is logically independent of the agent-neutral/agent-relative distinction. There are four possible ways that these two distinctions can combine.[7] First, a rule might be both agent-neutral and moment-neutral (e.g. the 'maximize total net pleasure' rule of classical utilitarianism). Second a rule might be both agent-relative and moment-relative (e.g. the traditional deontological constraint on killing). Third, a rule might be agent-neutral but moment-relative (e.g. the rule specified in Broome's miner example). Fourth, a rule might be agent-relative yet moment-neutral (e.g. a rule that normally prohibits killing to prevent more killing but permits it when current and future killers are the same agent).

Although the moment-neutral/moment-relative distinction has not received as much attention as the agent-neutral/agent-relative distinction, it is nonetheless an important distinction in value theory. For example, Kamm (1992) uses moment-relativity to argue that deontological constraints are best explained by the inviolability of victims. Portmore (2011, pp.103-108) uses it to argue that agent-relative consequentialism does better than its deontological rivals. Finally, Broome (1991, p. 9) argues that moment-relativity is more important than agent-relativity when it comes to understanding deontological side-constraints.

4. Patient-Neutrality and Patient-Relativity

---

[7] See Louise (2004) p. 535.

The two distinctions I discuss in §2 and §3 are logically independent and yet appear to be closely related. Both distinctions are recognized by moral philosophers and applied to important debates in normative ethics. However, there is a third distinction that appears to belong to this family of distinctions and yet has so far gone unrecognized. I will label this distinction the 'patient-neutrality/patient-relativity distinction'. The idea behind this distinction is simple. The existence of agent-relative moral theories shows us that the numerical identity of the *agent* acting might make a difference to the deontic status of the act. The existence of moment-relative moral theories shows us that the numerical identity of the *time-slice* acting might make a difference to the deontic status of the act. A third factor that can be variable in such cases is the identity of a moral patient. In many scenarios when an agent faces a morally significant choice there is a moral patient who is affected in some way by the choice the agent makes. Perhaps, just as the numerical identity of the agent or the time-slice can make a difference, the numerical identity of the moral patient can make a difference. When patient identity makes a difference to the deontic status of an act then we have a case of patient-relativity. When it makes no difference we have patient-neutrality.

This is a reasonable first attempt at explaining patient-relativity. However, much more needs to be said to explain this distinction and demonstrate its importance to moral theory. I will begin this task in §4.1 below by presenting a case which, according to our commonsense moral judgments, appears to be an instance of patient-relativity. In §4.2 I will offer a precise account of the patient-relative/patient-neutral distinction. I will also look at different ways that patient-relativity can combine with agent-relativity and moment-relativity. Finally, in §4.3 I will show some useful applications in moral theory for patient-relativity.

4.1 An Initial Case of Patient-Relativity

Some moral theories endorse an agent-relative rule prohibiting promise-breaking. Such a rule requires agents not to break a promise even if doing so is the only way to prevent more promise-breaking by others. However, for any theory that endorses such a rule there is a further question we can ask. Should the rule be interpreted as moment-neutral or moment-relative? To answer this we ask whether an agent is permitted to break a promise now if that is the only way she can ensure that she keeps several equally weighty promises in the future. For example, suppose Alfred has made a promise to his colleague Marvin to send a report to Marvin by tonight. However, he has also made five separate promises to five other colleagues that require him to complete various tasks by tomorrow (which are

each as important as sending Marvin the report). Alfred has failed to make a start on any of these tasks and realises that, unless he works on these tasks now, thereby breaking the promise to Marvin, he will break all five of tomorrow's promises. Thus he must choose between keeping his promise to Marvin now and tomorrow breaking promises to five other colleagues, or breaking his promise to Marvin now and keeping his promises to five others tomorrow. In these kinds of cases I suspect that some will endorse a moment-neutral interpretation of the agent-relative rule prohibiting promise-breaking, while others will favour a moment-relative interpretation. Those who favour moment-neutrality might reason as follows:

> Those who make incompatible promises have done wrong and owe compensation to any promisee who fails to get her due. However, in deciding which promises to keep and which to sacrifice, it is best that the agent minimizes her overall promise-breaking because, other things being equal, it is better that a lesser number rather than a greater number are wronged by the bad faith of promise-breaking. Furthermore, temporal order does not seem to be a morally significant factor. Why should the fact that the promise made to you will be fulfilled before the promise made to me mean that the fulfilment of your promise counts for more than the fulfilment of mine? But, if temporal order is not morally significant, then it is morally unjustified to favour keeping the promise that will be broken first simply because of its temporal order.

By contrast, those who favour the moment-relative account may reason as follows:

> The person who makes incompatible promises still has an obligation, at each moment, to keep any promise that it is in her power to keep. To fail to do so for the gain of more personal promise-keeping overall is to use one of the promisees and the commitment she is owed as a means to the end of fulfilling other promises, and this is morally objectionable. Furthermore, my moment-neutral rival is already committed, via her agent-relativity, to the claim that maximizing general promise-keeping is not a good enough reason for breaking a promise. But if she is committed to this then how can she say that maximizing personal promise-keeping justifies breaking a promise? To think that the general case does not count but the personal case does is to be overly partial to oneself, and such partiality is morally objectionable.

Perhaps many will find the moment-relative account the more convincing of the two in this debate. However, consider a variation of the case where the one proximate promise and the five equally weighty future promises all have the same promisee (i.e. the same moral patient). For example, perhaps Alfred's one promise due tonight and his five promises due tomorrow are all promises he made to Marvin. Does the promisee being the same person rather than a different person change things? Many will be inclined to say that it does. For, it is one thing to say that I must not break a promise to you now in order to keep (equally weighty) future promises I have made to five *others*. It is another thing to say that I must not break a promise to you now if that is the only means available to me to

keep five (equally weighty) future promises I also made to *you*. In the former case you will be worse off if I break my promise. However, in the latter case you will be better off overall. For, from your perspective as the promisee, five promises kept and one broken is preferable to one kept and five broken (assuming that all promises are equally weighty). Furthermore, in the former case you might reasonably claim that you have been treated unfairly by having the promise owed to you neglected so that others can gain some advantage. However, in the latter case such claims of unfairness fail because the promise owed to you now is being neglected not to advantage others at your expense, but rather to your advantage. On these grounds, many who endorse the moment-relative interpretation of the constraint on promise-breaking will agree that the constraint is best given a moment-neutral interpretation when we are making a trade-off between present and future promise-breaking involving the same patient. Thus, in this case it appears that the identity of the moral patient makes a difference to moral requirements. When the patients involved in a trade-off are numerically identical then the constraint is moment-neutral. However, when the patients are not numerically identical then the constraint is moment-relative.

The example above involves promising. However, other moral duties also appear to have this patient-relative structure. For example, consider the special duties we owe to friends. Such duties are agent-relative because they require us not to neglect our friends even if doing so will cause less neglecting of friends by others. They also appear moment-relative because they require us not to neglect a friend now even if doing so will result in our neglecting other friends less in the long run. However, what about a case where the short-term and long-term trade-offs involve the same friend? Suppose that Sita has the options of either giving support to her friend Ruth now, or withholding support at this moment. However, Sita learns that if she gives her support now then this will make her unable to give her support on many future occasions (perhaps giving support at this moment will lead to a deep resentment which will make it hard for her to support Ruth in the future). On the other hand, if Sita withholds support now she will be able to offer more support to Ruth in the future and will end up being a better friend in the long run. Many will think that, although duties of friendship are generally moment-relative, in cases like this where short-term and long-term trade-offs involve the same friendship (i.e. the same moral patient), the moment-relative constraint no longer applies. Thus, duties of friendship appear plausible candidates for patient-relativity.

The examples considered so far (promissory duties and duties of friendship) are both special duties. However, some natural duties (i.e. duties that are not 'special' or

'associative' duties) are also candidates for patient-relativity. For example, consider a natural duty prohibiting stealing. Such a duty appears agent-relative because it requires each agent not to commit an act of theft even when doing so is the only way to prevent several comparable thefts being committed by others. Furthermore, many will hold that this duty is generally moment-relative because we must not steal from one person at the present-moment even if doing so is the only way to prevent our future selves from stealing from several others. However, consider a case where the current and future thefts are to be committed against the same person. It is not easy to construct a plausible version of such a case, but here is an attempt. Suppose that Rudi has set up a scam to steal one hundred dollars from Suki's bank account on five separate occasions over the next month. The scam is difficult to undo and the only way for Rudi to prevent it from going ahead is by stealing one hundred dollars from Suki's account now (perhaps doing so will set off a warning system that will prevent the future thefts). In such a case many will claim that the agent should steal from the patient now to prevent several comparable thefts against that same patient in the future. However, they will insist that the agent must not steal from one patient now in order to prevent several comparable thefts against different patients in the future. Thus, duties not to steal property are another candidate for patient-relativity.[8]

At this point it is worth clarifying what these initial examples of patient-relativity are intended to show. I am not arguing that any of the patient-relative rules considered above are in fact true. I am merely pointing to the possibility of such rules to highlight patient-relativity as a possible feature that a moral theory can possess. Many will find patient-relative rules of the kind described above intuitively plausible and even compelling. However, others will reject them. Agent-neutral consequentialists will reject them because they reject all agent-relative constraints. Others will reject them because they prefer the moment-neutral interpretation of such constraints (even when the trade-off is between different patients). Finally, someone who endorses the moment-relative interpretation of such a constraint might insist that this interpretation holds even in same-patient cases. For example, it might be claimed that it is wrong to break a promise to one patient in the present-moment, even if doing so is the only way to prevent a greater amount of future promise-breaking to that same patient. For perhaps doing so is to disrespect the institution of promising and this is wrong regardless of whether it is harmful or unfair to any particular individual. It is not my concern here to offer replies to each of these critics.

---

[8] Reflection on these examples raises an interesting question about whether there is a general principle that predicts patient-relativity (I thank an anonymous referee from *Ethics* for suggesting this possibility). I will not propose such a principle in this chapter. However, in §4.3 I will show how some moral theories are committed to patient-relativity, and this discussion will illuminate several directions we might take in formulating a general principle.

However, §4.3 will indirectly address the issue of whether we should endorse any patient-relative moral rules by showing how some normative theories might entail, or provide strong support for, patient-relativity.

4.2 The Logical Structure of Patient-Relativity

The cases we considered above give us an initial picture of what a patient-relative moral rule might look like. However, more work needs to be done to get a clearer picture of the nature and scope of patient-relativity. In the cases we have already considered, the identity of the patient acted as a switch that turns on and off the moment-relative requirement. When the proximate patient was also the future patient affected by the action, then moment-relativity was turned off. When the proximate patient was distinct from any future patients, then moment-relativity was turned on. However, the same process may also occur in relation to agent-relativity. To see how consider the following example. Suppose that if I break a promise to you I will prevent several promises made by other agents to other patients from being broken. The agent-relative rule will prohibit promise-breaking in such circumstances. However, let's modify the example and suppose that the promises made by other agents are also promises made to you. I have the options of either keeping my promise to you or breaking it. I know that five (equally weighty) promises made to you by others will only be kept if I break my promise. What should I do? An obvious answer is that I might ask you to cancel my promise (a power that you have as the promisee) so that you can benefit from your future promises being honoured. However, suppose that this is not possible (perhaps you are unaware of the situation and I am unable to inform you). What should the agent do in such a circumstance? One possible answer is that the agent ought to break the promise because this will make the promisee better off overall. This answer involves a form of patient-relativity. However, some may have reservations about this answer. For, unlike the initial case where the agent was minimizing her own promise-breaking, this case requires an agent to get her hands dirty by breaking a commitment that she otherwise would have kept, thereby committing a prima facie wrong. It might be argued that this is an unfair burden to place on the agent.[9] Nonetheless, some moral theories may endorse this kind of patient-relativity. Therefore, although patient-relativity may look most plausible in the case where it acts as a switch on moment-

---

[9] This complaint may have even more intuitive force in the friendship example. For, unlike the good of promise-keeping, the good of friendship may be equally good for both the patient and the agent and thus a duty to sacrifice friendship for the greater good of the friend may be an unfair burden on the agent.

relativity, it could also act as a switch on agent-relativity, cancelling the agent-relative constraint when the trade-off involves the same patient.

This raises the question of how patient-relativity is related to the other distinctions we have discussed. In §3 we saw that agent-relativity and moment-relativity are logically independent of one another and thus there are no limits on how agent-neutrality/relativity and moment-neutrality/relativity might be combined in a moral theory. Is the same true of patient-relativity? The talk above of it being a switch that turns on and off other kinds of relativity suggests that patient-relativity lacks logical independence. However, the switch metaphor is a little misleading here. For each of the cases we have considered so far involve not *one* patient-relative rule but rather *two* rules. For example, consider the initial promise-breaking example. One interpretation of the agent-relative constraint on breaking promises is moment-neutral. According to this interpretation the constraint on breaking promises requires agents to never break a promise unless doing so will lead to her breaking less promises overall. Such a rule can be succinctly stated as follows:

> (3) Each agent must minimize her own promise-breaking.

Another interpretation of the agent-relative constraint on promise-breaking is moment-relative. According to this interpretation the constraint on breaking promises requires agents to never break a promise in the present even when doing so will result in her breaking fewer promises overall. This rule can be stated as follows:

> (4) Each agent, at each moment, must not break a promise at that moment.[10]

Finally, the patient-relative option requires agents to follow (3) in same-patient cases and follow (4) in different-patient cases. I submit that such a theory is best represented as containing two distinct rules:

> (5) Each agent must minimize her own promise-breaking to the same patient.

> (6) Each agent, at each moment, must not break a promise to one patient to bring about less promise-breaking to other patients.

Rules (5) and (6) are each patient-relative because each rule makes reference to the numerical identity of moral patients. Rules (3) and (4) are patient-neutral because each rule makes no reference to the numerical identity of moral patients. Thus, we can say that

---

[10] An alternative way of stating the moment-relative rules is as follows: '(4′) Each agent, at each moment, must minimize her promise-breaking at that moment'. For my purposes here it does not matter which version of the moment-relative rule we use. However, later in chapter 4 of this thesis we will see that the difference between rules like (4) and (4′) can be very important.

a moral rule is patient-relative if and only if, when the rule is stated fully, it contains an ineliminable reference to the numerical identity of a moral patient, and is otherwise patient-neutral.[11]

Although it seems natural for a moral theory to combine rule (5) with rule (6), these two rules are logically independent as there is nothing contradictory in a moral theory containing one of these rules without containing the other. More generally, a moral theory can contain a rule proscribing an action in same-patient cases and yet contain no rule pertaining to what is permitted or prohibited in the corresponding different-patient cases (and vice versa). However, it seems reasonable to conjecture that any plausible moral theory containing a rule covering same-patient cases will also contain a rule covering different-patient cases (and vice versa). For, it would be strange for a moral theory to require one course of action in same-patient cases if it didn't require a different course of action in different-patient cases. Thus, we can expect that, in the moral theories that interest us, patient-relative moral rules will generally come in pairs (covering same and different patient cases) whereas patient-neutral rules will not.

With this understanding of patient-relativity in place we can now see that there are various possible combinations of agent, moment, and patient relativity/neutrality. The cases we have considered so far have only covered a few of these possible combinations. Table 1 summarizes all eight possible combinations.[12]

---

[11] This intuitive account of the patient-neutral/patient-relative distinction can be adapted to the formal approach of McNaughton and Rawling discussed in chapter 1 of this thesis. According to that approach, all rules expressing moral requirements can be expressed in the canonical form '$(x)(t)(x$ must {*verb*} at t that [. . .])'. Now note the following: (i) in some formulas in this canonical form an identity operator '=' or '≠' will appear in the square brackets, (ii) in some of these cases, the variables that stand in the identity relation will take a moral patient position in the formula. We can then define patient-relativity as follows: A rule expressed in the canonical form is patient-relative just in case there is an ineliminable occurrence of an identity operator in the square brackets, and the variables that take a position in the identity operator also take a moral patient position in the formula. Using this idea, we might translate (6) above into the canonical form as: '(6′) $(x)(t)(x$ at t must act such that $[(y)(z)(\text{If } y{\neq}z$ then it is not the case that x breaks a promise to y at t in order to keep more promises overall to $z)])$'. Notice that (6′) contains an ineliminable occurrence of an identity operator in the square brackets, and that the variables connected by this operator take patient positions in the formula. Thus, according to the formal account of patient-relativity, (6′) is correctly categorized as patient-relative.
[12] I thank Philip Pettit for urging me to consider all eight possible combinations summarized in this table.

Table 1

Combinations of Agent, Moment, and Patient-Relativity/Neutrality

	Agent	Time	Patient	Example
1	N	N	N	Minimize general promise-breaking
2	N	N	R	Minimize promise-breaking to the same patient
3	N	R	N	Minimize present-moment promise-breaking
4	N	R	R	Minimize present-moment promise-breaking to the same patient
5	R	N	N	Minimize your promise-breaking
6	R	N	R	Minimize your promise-breaking to the same patient
7	R	R	N	Never break a promise, no matter the future consequences
8	R	R	R	Never break a promise to a patient, no matter the future consequences for that patient

Note.—N = neutral; R = relative.

It should be noted that any rule that combines agent and moment-relativity with patient-neutrality (i.e., combination 7) will entail a rule that combines relativity in all three domains (combination 8). For example, the rule "Never break a promise, no matter the future consequences" entails the rule "Never break a promise to a patient, no matter the future consequences for that patient" (meaning that the former rule prohibits everything that the latter rule prohibits).

No other entailments hold between the different combinations in this table. For example, it is not the case that rules that are neutral in every domain (combination 1) entail rules that are agent and moment-neutral yet patient-relative (combination 2). One might think that such an entailment holds but this is a mistake. For example, one might think that if you are required to minimize general promise-breaking then you are required to minimize promise-breaking to the same patient. However, this is not the case. In fact, these two rules are contraries. For the latter rule (combination 2) requires an agent to break a promise to a patient when that is the only way to bring about more promise keeping to that patient overall. However, this requirement conflicts with the former rule (combination 1) because there will be cases where breaking a promise will minimize promise-breaking to the promisee and yet cause more promise-breaking overall. In such cases the former rule requires promise-keeping whereas the latter rule requires promise-breaking.

4.3 Applications of Patient-Relativity

Above I motivated patient-relativity by presenting several cases where commonsense moral intuitions appear to require moral rules that make reference to the numerical identity of moral patients. However, we might ask whether there are moral theories that lend support to patient-relativity independently of these commonsense intuitions. In this

section I will consider three different moral theories that appear to provide a theoretical basis for patient-relativity. I will then show how patient-relativity provides a reply to Derek Parfit's 'appeal to full relativity' argument.

First, consider the *humanity formula* of Kant's categorical imperative:

> FH: Always treat people as ends in themselves and never as mere means.

The humanity formula is a very influential moral principle and there are many different accounts of how to interpret it. On some, but not all, interpretations of this principle it gives the patient-relative verdict in the promise-keeping examples considered above. For breaking a promise now to one promisee to keep several promises in the future to other promisees might be classified as treating the present promisee as a 'mere means' to fulfilling promises owed to several others. Furthermore, breaking a promise now to one promisee to keep several promises in the future to that same promisee might be classified as treating that promisee as an 'end in herself' because it is out of concern for the promisee, qua rational agent, that the present promise is broken. It might be thought that this possibility demonstrates that we can account for the promise-keeping examples without appealing to patient-relative rules. For the humanity formula does not appear, at face value, to be a patient-relative rule (it makes no reference to the numerical identity of patients). However, I believe it shows the opposite. For the concepts used in the humanity formula of treating a person as a 'mere means', and as an 'end in herself', are imprecise and ambiguous in various ways. Indeed, the principle has many interpretations because different thinkers have taken these concepts to mean different things. Thus, when the principle is used in a particular example to derive a moral verdict we should ask: what interpretation of these key concepts is required to produce the claimed verdict? Perhaps there are several possible interpretations of 'mere means' and 'end in themselves' that allow the humanity formula to explain the promise-keeping examples. However, here is one interpretation worth considering:

> To treat someone as a 'mere means' is to perform an action with the motive of using that person to serve, or bring about, a goal that is not that person's own goal but rather the goal of some numerically distinct person. To treat someone as an 'end in themselves' is to perform an action with the motive of using that person to serve, or bring about, a goal that is that person's own goal.

On this interpretation, the humanity formula turns out to be a patient-relative rule as, when fully stated, it makes reference to the numerical identity of moral patients. Several other obvious ways of interpreting the imprecise concepts in the humanity formula also produce a rule that is patient-relative. Thus, I suspect that all plausible interpretations of

the formula of humanity that allow it to explain the promise-keeping examples will be interpretations that make the principle patient-relative. However, arguing for this claim would be a long and difficult task and I will not attempt it here. Showing that, on some candidate interpretations of the formula of humanity, it is a kind of patient-relative rule that explains the promise-keeping examples is itself a significant conclusion. Thus, one theoretical basis for patient-relativity comes via certain interpretations of Kant's humanity principle.

As a second example, consider the moral contractualism of T.M. Scanlon. According to Scanlon:

> An act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behaviour that no one could reasonably reject.[13]

Now, consider again the same-patient promise-keeping example. In this case there does not appear to be anyone who could reasonably reject the trade-off that is made on the patient-relative rule. For, there appears to be no one who can claim that they are being treated unfairly or unjustly. The patient involved has no reason to reject the trade-off, and, in fact, has reasons to welcome it as she is better off in the long run by it (e.g., fewer promises owed to her are broken). The agent involved also has no reason to reject it and also appears to be better off as she minimizes her own wrong-doing. Finally, there appear to be no reasons for any third party to object to the trade-off being made. This can be contrasted with the 'different-patient' case where the thing owed to one patient is withheld in order to give several other patients what they are owed. In this case the patients who benefit in the future, and the agent who performs the action, may be better off. However, the patient who loses in the present-moment is worse off and can claim that she has been treated unfairly. On these grounds it could be reasonable to reject a principle that allowed this kind of trade-off. Thus, Scanlon's contractualism can provide a theoretical grounding for patient-relativity.

As a third example consider the distinction between interpersonal aggregation and intrapersonal aggregation. Interpersonal aggregation is aggregation of the good across the lives of different patients. Intrapersonal aggregation is aggregation of the good in the life of a single patient. Some take this distinction to be morally significant.[14] They typically endorse the following principle:

---

[13] Scanlon (1998), p. 153

[14] See Nagel (1970) p. 138, Scanlon (1998) pp. 229-240, and Kelleher (2014) pp. 404-405. Nagel appeals to the separateness of persons to explain the moral significance of the distinction between intrapersonal and interpersonal aggregation.

PA: It is morally permissible to harm a patient in order to maximize that patient's total welfare. However, it is morally impermissible to harm a patient in order to maximize the total welfare of other patients.[15]

PA is often motivated by appeal to the following considerations. It is normally rational for us to bear risks or accept certain harms to ourselves when doing so maximizes our overall welfare. Given this fact it seems permissible for others to harm us in these ways in order to maximize our welfare. Yet many have the intuition that it is morally impermissible to harm someone to bring about a greater benefit to others. PA seems to adequately capture this set of intuitions.

PA is a patient-relative rule because it makes reference to the numerical identity of moral patients. Furthermore, PA entails the patient-relative verdict in the promise-keeping example. For making the trade-off in the different-patient case appears to be a form of interpersonal aggregation and thus is prohibited by PA. On the other hand, making the trade-off in the same-patient case appears to be a form of intrapersonal aggregation and thus is permitted by PA. However, although PA can explain patient-relativity in the promise-keeping case, it cannot explain all candidate cases of patient-relativity. For some moral rules that are unrelated to the aggregation of welfare can also be classed as patient-relative or patient-neutral. As an example, consider a rule that says the guilty ought to be punished. What does such a rule require in a case where the only options are: (i) punish one guilty person now, which will allow several people guilty of comparable crimes to escape punishment later; (ii) let one guilty person escape punishment now, which will ensure that several people guilty of comparable crimes get punished later? What about a case where the trade-off is between letting a guilty person get punished for one crime now but escape punishment for several comparable crimes later, or letting that person escape punishment for a crime now to ensure that she is punished for several later crimes? Some will endorse a rule that requires the trade-off in both kinds of cases because they believe we must maximize the punishment of the guilty. Such a rule is patient-neutral. Others will endorse a rule that prohibits the trade-off when there are different guilty patients but requires the trade-off with respect to the same guilty patient. Such a rule is patient-relative. Yet, on several prominent theories of retributive punishment, neither of these rules concerns the aggregation of patient welfare. Therefore, we might endorse patient-relativity in certain cases that cannot be explained by the principle PA.

---

[15] Those who endorse this principle may qualify it in various ways. For example, they may agree that, above a certain threshold, harming in order to interpersonally aggregate is morally permitted. However, such a threshold is only applied to interpersonal aggregation; intrapersonal aggregation is taken to be permissible simpliciter.

It is also worth considering more generally how the distinction between interpersonal and intrapersonal aggregation relates to the patient-relative/patient-neutral distinction. Any moral rule concerned with interpersonal or intrapersonal aggregation will be a patient-relative rule as it will make reference to the numerical identity of moral patients. By contrast, any moral rule concerned only with aggregation simpliciter will be a patient-neutral rule as it will not make reference to the numerical identity of moral patients. An example of the former kind of rule is PA above. An example of the latter kind of rule is the aggregation rule endorsed by classical utilitarianism:

> UA: The morally right action is that which maximizes total net welfare.

UA is an aggregation rule, yet in contrast to PA it is a patient-neutral rule as it makes no reference to the numerical identity of moral patients. It might be tempting to conclude from these observations that the patient-neutral/patient-relative distinction is equivalent to the distinction between moral rules that concern aggregation simpliciter and moral rules that concern interpersonal or intrapersonal aggregation. However, this is not the case, as moral rules that do not require any form of aggregation can also be classed as patient-relative or patient-neutral. For example, some deontologists advocate the 'Equal greatest chance principle' in rescue cases involving interpersonal trade-offs.[16] They argue that if we find ourselves having to choose between saving the life of one stranger and saving the life of five strangers, and other than numbers all else is equal, then the morally correct thing to do is not to favour the larger number. Instead we must give all strangers the greatest possible chance of being saved consistent with every stranger having an equal chance of being saved. In the case described we might do this by flipping a coin to decide whether to save the one or save the five, thereby giving each stranger a one in two chance of being saved. Deontologists who endorse this principle in rescue cases might also apply it to trade-offs involving promise-breaking as follows:

> EGC: When there are comparable promises owed by one promisor to different promisees, the promisor must give each promisee the greatest possible chance of having her promises honoured, consistent with every promisee having an equal chance of having her promises honoured.[17]

EGC is a patient-relative rule because it makes reference to the numerical identity of moral patients. However, it is not a rule concerned with either interpersonal or intrapersonal aggregation. Thus, although all rules requiring interpersonal or intrapersonal aggregation

---

[16] For example, see Taurek (1977).
[17] EGC would typically be combined with a rule requiring promisors to minimize their promise-breaking with respect to comparable promises owed to the same promisee.

are patient-relative, some patient-relative relative rules do not require either kind of aggregation. Therefore, the patient-relative/patient-neutral distinction is broader and more fundamental than the distinction between rules that require aggregation simpliciter and rules that require interpersonal or intrapersonal aggregation.

Finally I will examine how recognizing patient-relativity can provide a response to Derek Parfit's 'appeal to full-relativity argument'. In chapter nine of *Reasons and Persons*, Parfit develops an argument against rational egoism he calls 'the appeal to full relativity'. As we saw above in §3, there are four possible ways a rule can combine agent and moment neutrality/relativity. Two of these four possibilities combine relativity in one domain with neutrality in another domain. Parfit describes such rules as 'incompletely relative'. The other two possibilities were either neutral in both domains or relative in both domains. Parfit describes these respectively as 'fully neutral' and 'fully relative'. According to Parfit, we should reject on theoretical grounds rules that are incompletely relative, and thus only accept rules that are fully neutral or fully relative. Parfit claims this because he thinks there is an important analogy between 'I' and 'now':

> This claim can appeal to the analogy between oneself and the present, or what is referred to by the words 'I' and 'now'. This analogy holds only at a formal level. Particular times do not resemble particular people. But the word 'I' refers to a particular person in the same way in which the word 'now' refers to a particular time. And when each of us is deciding what to do, he is asking, 'What should *I* do *now?*' Given the analogy between 'I' and 'now', a theory ought to give to both the same treatment.[18]

Parfit (1984) directs the appeal to full-relativity against the theory of rational egoism. However, he notes that this appeal is intended to apply to all normative theories and thus can be used to defeat any incompletely relative theory.[19] This has led some to argue that there are only two plausible kinds of moral theories, those that are agent- and moment-neutral and those that are agent- and moment-relative.[20] However, the most promising cases of patient-relativity appear to challenge this. For example, take the promise-breaking case discussed in §4.1. In this case we saw that there are strong intuitive and theoretical grounds for thinking that an agent-relative constraint on promise-breaking must be moment-neutral with respect to the same patient. However, any moral rule that secured this result would be incompletely relative because it would combine agent-relativity with moment-neutrality. Thus, the intuitive and theoretical grounds in favour of this kind of patient-relativity are in conflict with Parfit's theoretical considerations against in-

---

[18] Parfit (1984), p. 140.
[19] See Parfit (1984), p. 148
[20] See Louise (2004), p. 535

completely relative rules. Given that the considerations in favour of patient-relativity appear strong and Parfit's analogy argument appears weak, we should accept the possibility of incompletely relative moral theories. Importantly, once we accept an incompletely relative theory in any part of our normative theorizing, we no longer have grounds for rejecting other theories for their incomplete relativity. Thus, recognizing there are strong arguments for patient-relativity not only demonstrates that there can be incompletely relative moral theories, it also undermines Parfit's use of the appeal to full-relativity against rational egoism.

5. Conclusion

In this chapter I have presented a new kind of moral relativity that stands alongside agent-relativity and moment-relativity. All moral rules can be classified as either agent-relative or agent-neutral. They can also be classified as either moment-relative or moment-neutral. This chapter shows that there is a third type of classification that needs to be added to this list. A moral rule can also be classified as either patient-relative or patient-neutral. Being aware of this third kind of relativity improves our understanding of the main structural features of moral theories. It also helps us to better understand how agent-relativity and moment-relativity interact. Finally, it gives us plausible examples of normative theories that are incompletely relative and, as such, provides us with a response to Derek Parfit's 'appeal to full relativity' argument.

References

Broome, J. (1991). *Weighing Goods*. New York: Blackwell.

Dreier, J. (1993). Structures of Normative Theories. *The Monist*, 76: 22-40.

Hammerton, M. (2016). Patient-Relativity in Morality. *Ethics*, 127: 6-27.

Hurka, T. (2003). Moore in the Middle. *Ethics*, 113: 599-628.

Kamm, F. M. (1992). Review: Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status. *Philosophy and Public Affairs*, 21: 354-380.

Kelleher, J. P. (2014). Relevance and Non-consequentialist Aggregation. *Utilitas*, 26: 386-408.

Louise, J. (2004). Relativity of Value and the Consequentialist Umbrella. *Philosophical Quarterly*, 54: 518-36.

McNaughton, D. & Rawling, P. (1991). Agent-Relativity and the Doing-Happening Distinction. *Philosophical Studies*, 63: 167-185.

McNaughton, D. & Rawling, P. (1992). Honoring and Promoting Values. *Ethics*, 102: 835-843.

Nagel, T. (1970). *The Possibility of Altruism*. Oxford: Clarendon Press.

Parfit, D. (1979). Prudence, Morality and the Prisoner's Dilemma. *Proceedings of the British Academy*, 65: 539-564.

Parfit, D. (1984) *Reasons and Persons*. Oxford: Clarendon Press.

Pettit, P. (1997). The Consequentialist Perspective. In M. Baron, P. Pettit, and M. Slote (eds.), *Three Methods of Ethics*, Oxford: Blackwell.

Portmore, D. (2001). Can an Act-consequentialist Theory be Agent-relative? *American Philosophical Quarterly*, 38: 363-77.

Portmore, D. (2014). Agent-Relative vs. Agent-Neutral. In Hugh LaFollette (ed.), *International Encyclopedia of Ethics*, Wiley Blackwell.

Ridge, M. (2001). Agent-Neutral Consequentialism from the Inside-Out: Concern for Integrity without Self-Indulgence. *Utilitas*, 13: 236-254.

Ridge, M. (2005). Agent-Neutral vs. Agent-Relative Reason. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (Ed.), URL = http://plato.stanford.edu/archives/win2005/entries/reasons-agent/

Scanlon, T.M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Taurek, J. (1977). Should the Numbers Count? *Philosophy and Public Affairs*, 6: 293-316.

3

Agent-Neutral Deontology Refuted[1]

1. Introduction

It is commonly held that all deontological moral theories are agent-relative in the sense that they give each agent a special concern that *she* does not perform acts of a certain type. Furthermore, this is often taken to be a conceptual truth about deontology. Thus, agent-neutrality is thought to be solely the domain of consequentialist moral theories and agent-neutral deontology is regarded as a conceptual impossibility. Recently, several authors have challenged this standard view and argued that agent-neutral deontology is possible.[2] Some have even argued that agent-neutral varieties of deontology are generally more plausible than agent-relative varieties. This chapter defends the standard view against such challenges by first reviewing and refuting the arguments of several prominent challengers, and then offering a new general argument for the thesis that deontology is agent-relative as a matter of conceptual necessity. In the remainder of this introduction I will offer some background on what is at stake in this debate. I will then proceed as follows. In §2 I explain in more detail the distinction between agent-neutrality and agent-relativity. In §3, §4, and §5 I review and refute arguments that agent-neutral deontology is possible by Frances Kamm, Tom Dougherty, and John Broome respectively. Finally, in §6 I offer a general argument against agent-neutral deontology. This argument employs insights from the discussion of the preceding sections to show that all deontological moral theories are agent-relative.

It should be noted at the outset that the claim that deontology is necessarily agent-relative is distinct from the claim that consequentialism is necessarily agent-neutral. Some have endorsed these two claims together.[3] When these claims are combined they lead to a

[2] Examples of philosophers claiming that agent-neutral deontology is possible include Dougherty (2013), Tradros (2011), p. 117, Kamm (1992), Broome (1991) p. 9, and Kuflik (1986).
[3] For example, McNaughton and Rawling (1991/1992), and Pettit (1997).

specific account of the distinction between consequentialism and deontology. This account draws the distinction in terms of the agent-neutral/agent-relative distinction. Yet despite this combination of claims having this attraction, many philosophers have endorsed one claim while rejecting the other. For example, some agent-relative consequentialists reject the claim that consequentialism is necessarily agent-neutral and yet endorse the claim that deontology is necessarily agent-relative.[4] This independence means that my arguments against the possibility of agent-neutral deontology have no bearing on the debate surrounding agent-relative consequentialism.

Finally, before going on, it is worth saying something about why it matters whether deontology can be agent-neutral. To some this may appear to be a mundane question of theory classification. However, there are several debates in moral theory were the possibility of agent-neutral deontology matters. Here are three examples of such debates:

(i) It is sometimes argued that agent-relative moral theories are implausible because they require different things from different agents, failing to give everyone a common moral purpose. Furthermore, agent-relative theories entail that an agent's numerical identity can make a difference to her moral obligations, and this is in tension with the universalist aspirations that many hold for morality.[5] Some find these concerns troubling while others believe that they are easily dismissed. However, both sides generally agree that what is at stake in this debate is the plausibility of deontology, for they assume that deontology is necessarily agent-relative. If agent-neutral deontologists are correct then this is a mistake and the success of deontology is not crucially tied up with the plausibility of agent-relativity.

(ii) Some have argued that agent-neutral deontology is superior to agent-relative deontology in several respects. For example, Dougherty (2013) argues that his purported example of agent-neutral deontology better captures the intuitive idea that a deontologist should take steps to ensure that moral constraints are obeyed by all agents and not just herself. However, Dougherty's argument would fail if agent-neutral deontology were impossible. Therefore, knowing whether it is possible has relevance to debates between deontologists about what form the best deontological theories should take.

(iii) Some theorists have used the possibility of agent-relative consequentialism to 'consequentialize' moral theory. One consequentializes moral theory by showing how any deontic verdict reached by a non-consequentialist moral theory can also be reached by a

---

[4] For example, see Portmore (2001), pp. 376-377.
[5] See especially Pettit (1997) and Ridge (2005b) on this point.

consequentialist moral theory. If agent-neutral deontology is possible then, depending on the details, it might be possible to 'deontologize' moral theory—showing how any deontic verdict reached by a non-deontological moral theory can also be reached by a deontological moral theory. Some argue that consequentializing increases the plausibility of consequentialism. If this argument has any merit, then using agent-neutral deontology to 'deontologize' moral theory might be an important countermove in these debates.[6]

2. Agent-Neutrality and Agent-Relativity

In order to reach a verdict on whether deontology is necessarily agent-relative we need a precise account of the distinction between an *agent-relative* and an *agent-neutral* moral theory. However, as we will see below, a theory's status as neutral or relative depends on the status of the rules it contains. Thus, it is necessary to start with an account of the distinction between an *agent-relative rule* and an *agent-neutral rule*. With such a distinction in place we can describe what it is for a *moral theory* to be agent-relative or agent-neutral.

There are two main accounts in the literature of the distinction between agent-relative and agent-neutral rules.[7] Parfit (1984) cashes out the distinction in terms of the aims the rule gives to agents. McNaughton and Rawling (1991) offer a formal account of the distinction that relates it to the presence or absence of certain bound variables. Although there are some cases where these two accounts give different verdicts they agree on a wide range of cases.[8] Furthermore, they give the same verdict in all the cases discussed in this chapter. For this reason nothing crucial hangs on which of these two accounts I employ here. I will go with the Parfit account because I believe it is more intuitive to grasp than the formal account of McNaughton and Rawling.

According to Parfit (1984), an *agent-relative rule* is a rule that gives different agents different ultimate aims. On the other hand, an *agent-neutral rule* is a rule that gives all agents the same ultimate aim. Thus, consider a rule that prohibits killing innocent people:

DK: Each agent must not kill innocent people

---

[6] See Hurley (2013), pp. 140-141 for more on 'deontologizing'.

[7] There are further accounts (beyond the two I discuss) that apply the distinction to reasons or values rather than rules. However, as my focus is on moral rules I will not consider these accounts. In any case, these accounts generally give the same verdicts as the Parfit account in all the cases discussed in this chapter. For more detail on accounts of the distinction offered in different domains see Portmore (2014).

[8] See McNaughton and Rawling (1991) for discussion of these different verdicts.

DK is agent-relative because it gives different aims to different agents. For example, it gives Alfred the aim that 'Alfred does not kill'. On the other hand, it gives Sita the aim that 'Sita does not kill'. Now contrast DK with a rule that requires agents to minimize the killing of innocent people:

> MK: Each agent must minimize the killing of innocent people

MK is agent-neutral because it gives all agents the same ultimate aim. For example, it gives Alfred the aim that there be as little killing as possible, and it also gives Sita the aim that there be as little killing as possible. The difference between agent-relative rules like DK and agent-neutral rules like MK is especially clear in cases where, by performing some action, you can bring about fewer occurrences of that act-type overall. For example, suppose that the only way Alfred can prevent several innocent people being killed by the Mafia is by killing one innocent person himself. DK prohibits Alfred killing in such circumstances because it requires him to have a special concern with any killing that *he* performs, over and above killing that others perform. By contrast, MK requires him to give equal concern to all killings regardless of whether they are performed by him or performed by others. Thus, MK requires him to kill in cases like the one described above where doing so minimizes overall killing.

Two clarifications of Parfit's account are helpful here. First, when this account talks of a rule giving an agent an aim, it is referring to *ultimate* aims and not *derived* aims. For MK might give Alfred and Sita different derived aims if, given their different circumstances, Alfred best minimizes killing by aiming at one thing while Sita best minimizes killing by aiming at something else. However, these different aims are derived from the same ultimate aim—the aim that killing is minimized—and hence MK gives them the same aim in this relevant sense.[9] Second, one might interpret MK as giving different ultimate aims to different agents. Perhaps it gives Alfred the ultimate aim that 'Alfred minimizes killing' and Sita a different ultimate aim that 'Sita minimizes killing'. But then it seems that MK is not agent-neutral on such an interpretation. However, as McNaughton and Rawling (1991), and Portmore (2001) note, the distinction can still be made even if this interpretation of rules like MK is legitimate. For although this interpretation counts MK as giving different agents different ultimate aims, there is also an equally valid interpretation of MK where it gives different agents the same aim (i.e. the aim of minimizing killing). However, there is no plausible interpretation of DK where different agents have the same aim. Thus, strictly

---

[9] See Parfit (1984) p. 24.

speaking, a rule is agent-neutral if and only if there is a valid interpretation of it that gives all agents the same aim, and agent-relative otherwise.

Using the above account of the distinction between an agent-relative rule and an agent-neutral rule, we can now draw the distinction between an *agent-relative moral theory* and an *agent-neutral moral theory*. A moral theory is agent-neutral if and only if all the rules it contains are agent-neutral rules. A moral theory is agent-relative if and only if at least one of the rules it contains is agent-relative. Another way of putting this is as follows. A moral theory is agent-neutral if and only if it gives all agents the same set of ultimate aims. By contrast a moral theory is agent-relative if and only if it sometimes gives different agents different ultimate aims.[10] Given these definitions we can interpret the claim that agent-neutral deontology is possible as follows. There are possible deontological theories that contain only agent-neutral moral rules. Or, alternatively, there are possible deontological theories that give all agents the same set of ultimate aims.

3. Kamm's Agent-Neutral Deontology

In several publications Frances Kamm has argued that it is possible to give an agent-neutral account of deontological constraints, and thus it is a mistake to think that deontology is necessarily agent-relative.[11] For Kamm a deontological constraint is any constraint on performing an action that applies even when performing that action brings about a greater good overall.[12] Kamm appears to offer two distinct arguments for agent-neutral deontology. Below I will review each argument separately and argue that each is based on a mistake.

Kamm's first argument for agent-neutral deontology is based on a distinction she makes between two different ways of explaining or justifying deontological constraints—*victim-based explanations* and *agent-based explanations*.[13] In a victim-based explanation, constraints are explained and justified by facts about the victims they protect. For example, this explanation may appeal to the victim's inviolability or natural rights. In an agent-based-explanation, constraints are explained and justified by facts about the agents they constrain. For example, this explanation may appeal to the importance of agents never intending evil (even for a good end), or of agents maintaining personal integrity.

---

[10] See Parfit (1984), p. 27 and Dreier (1993), p. 22 for the second way of expressing it.

[11] This argument is most clearly stated in Kamm (1992). However, versions of it are also present in Kamm (1989), and Kamm (1996) chapter 9.

[12] See Kamm (1996), p. 189. Kamm makes it clear here that 'good' includes the good of there being less violations of that constraint overall. Thus, a deontological constraint on killing must prohibit killing even when it is the only way to prevent more killing overall.

[13] Kamm often calls the former 'victim-focused' and the latter 'agent-focused' or 'agent-centered'.

After making this distinction Kamm notes that deontological constraints that are given agent-based explanations may be agent-relative. However, she claims that when a constraint is given a victim-based explanation then it cannot be agent-relative and therefore must be agent-neutral.[14] It is difficult to reconstruct her argument for this claim because parts of it are left implicit. However, the main idea appears to be that victim-based explanations appeal to facts about victims rather than agents and thus must lack a special focus on the agent, making them ineligible for agent-relativity.

This argument for agent-neutral deontology is based on a fundamental mistake. It confuses the question of *what logical properties a moral rule possesses* with the question of *what explains or justifies a moral rule*. As we saw in §2, agent-relativity is a logical property that some moral rules possess. According to Parfit's account it is the property of giving different agents different ultimate aims. According to McNaughton and Rawling it is the property of there being a reference to a specific agent in the rule's content. Thus, as a consequence of its logical form, a rule will be either agent-relative or agent-neutral. Furthermore, it will maintain this status irrespective of what kind of explanation is offered for that rule. Thus, whether a moral rule is agent-relative or agent-neutral is logically independent of whether it is best justified by an agent-based explanation or a victim-based explanation.

To see this clearly consider the constraint on killing mentioned above:

> DK: Each agent must not kill innocent people

DK is an agent-relative rule because it gives different agents different ultimate aims. This follows from the logical form of DK and thus is true independently of whatever facts explain or justify DK. For example, suppose DK is true because all innocent people are inviolable (a victim-based explanation). This makes no difference to the fact that DK gives different agents different ultimate aims, and thus no difference to DK's status as agent-relative.

This conflation of a logical property with a category of explanation is especially evident when Kamm defines 'agent-relative', saying:

> Nagel also believes that we must give an *agent-relative* account of a constraint, where this means that the constraint arises because an agent stands in a special relation to his own act of intentionally killing by comparison to the acts of other agents who will intentionally kill. (1996, p. 247)

---

[14] See especially Kamm (1992), p. 382.

When Kamm says 'the constraint arises *because*' she is using the language of explanation to describe agent-relativity. This telling turn of phrase shows how Kamm may have come to think that a constraint being agent-relative makes it incompatible with a victim-based explanation. Indeed, it may lead one to suspect that Kamm is idiosyncratically using the term 'agent-relative' to refer to a kind of explanation, rather than following the common philosophical usage. However, in the same chapter where the above quotation occurs, she gives an additional definition of 'agent-relative' that follows the standard usage (basically paraphrasing Parfit's definition):

> This does not mean the view is fundamentally *agent-relative*, i.e., gives each agent a different *basic* aim not derivable from an aim each agent has (e.g., gives Joe a duty that other agents do not have to see to it that Joe not kill). (1996, p. 238)

Thus, this does not appear to be a case where different theorists are confusingly using the same terminology to talk about two different distinctions. When Kamm claims that deontology need not be agent-relative she appears to be talking about the same thing that Parfit and others are talking about when they use the term.[15] Yet, as we have seen, the first argument she offers for this claim fails.

A second argument for agent-neutral deontology might be inferred from a brief comment in Kamm (1992):

> In this sense the inviolable status (against being harmed in a certain way) of any potential victim can be taken to be an agent-neutral value. This is a non-consequential value. It does not follow (causally or noncausally) upon any act, but is already present in the status that persons have. Ensuring it provides the background against which we may then seek their welfare or pursue other values. It is not our duty to bring about the agent-neutral value, but only to respect the constraints that express its presence. (1992, p. 386)

Here Kamm argues that a deontological constraint that is explained by the inviolability of victims could be based on an agent-neutral value, i.e. something that is morally valuable for all agents and not valuable only relative to some. Agent-neutral values typically lead to agent-neutral consequentialism because it is assumed that we ought to promote whatever is valuable. However, Kamm, following the work of Pettit (1991), points out that one might honour a value rather than promote it. To honour a value is to always exemplify it in your actions even when doing so will lead to less of that value overall. So we might state Kamm's view as follows. The non-violation of people is an agent-neutral value, yet the appropriate

---

[15] As further evidence for this claim consider that Kamm sets her claims about agent-relativity in opposition to the claims of Nagel, Scheffler, and Kagan. Yet each of these authors is clear that they are using the term in its common philosophical meaning. See: Scheffler (1982), Nagel (1986), and Kagan (1989).

response to this value is to honour it rather than promote it. Thus, from this agent-neutral value we can derive deontological constraints on violating people such as DK. So far these are all reasonable things to say. However, one might then go further and reason as follows:

> Agent-neutral values entail agent-neutral rules and agent-relative values entail agent-relative rules. Therefore, if a deontological constraint is derived from an agent-neutral value then that constraint must be agent-neutral. Thus, deontological constraints on harming people who are inviolable, such as DK, are agent-neutral constraints.

It is easy to see how such an argument might appear sound. However, the first premise of the argument is false and thus the argument is unsound. For an agent-neutral rule follows from an agent-neutral value only if we take it for granted that *promoting* is the appropriate response to that value. If, instead, *honouring* is the appropriate response to that value then the rule we end up with is agent-relative. This follows because the honouring-relation, by its nature, gives different agents different ultimate aims. For example, consider the agent-neutral value of innocent people not being killed. To say this value is agent-neutral is to say that its moral value applies to all agents. If we decide that this value is best promoted then we end up with an agent-neutral rule like MK. However, if we decide that this value is best honoured then we end up with an agent-relative rule like DK.[16]

It is not clear if Kamm intended to suggest the unsound argument we have just considered when she mentioned deriving deontological constraints from agent-neutral values. However, such an argument is worth considering because if it were successful it would show that deontology can be agent-neutral. Yet, as we have seen above, the argument fails. Thus, there appears to be no credible way to derive agent-neutral deontology from the observation that deontological constraints might be explained by the inviolability of agents or some other victim-based explanation.

4. Dougherty's Agent-Neutral Deontology

Tom Dougherty (2013) offers an argument for agent-neutral deontology of a different kind from the argument found in Kamm. Dougherty attempts to show that agent-neutral deontology is possible by example. He produces a rule that he claims is an instance of a deontological constraint and yet is agent-neutral:

> DR: Each agent should ensure that no one kills to prevent more killings by others.[17]

---

[16] This link between 'honouring' and agent-relativity and 'promoting' and agent-neutrality is discussed at length by McNaughten and Rawling (1992) and Pettit (1997).

[17] Dougherty (2013) p.531

Let's call the act of *killing to prevent more killing by others* 'preventative killing'. Thus, DR requires agents to ensure that no one commits a preventative killing. According to Dougherty, DR is a deontological constraint because it prohibits agents performing an action even when doing so would result in there being fewer occurrences of that action overall. Thus, containing DR is sufficient for a moral theory to be deontological. However, DR also appears to be an agent-neutral rule because it gives all agents the same ultimate aim—the aim that *no one kills to prevent more killings by others*. Thus, if Dougherty's analysis of DR is correct then we have a proof by example that agent-neutral deontology is possible.

Dougherty also argues that the agent-neutrality of DR makes it preferable to standard agent-relative constraints on killing such as DK considered above. Agent-relative constraints like DK prohibit each agent from committing preventative killings. However, they do not give agents any further duties to prevent other agents from committing such killings. Yet Dougherty suggests that a deontologist opposed to preventative killings should want bystanders to intervene and prevent such killings whenever feasible. By being agent-neutral DR is able to achieve this, requiring that agents not only personally refrain from preventative killing but also that they stop others from committing such killings. Thus Dougherty sees DR as not only a possible agent-neutral deontological rule but also as a more plausible way of addressing the wrongness of killing than the traditional deontological constraint on killing.

Despite these claims, I believe that DR fails to deliver what it promises. For DR, as Dougherty formulates it, is ambiguous and has three possible readings. Each reading corresponds to a different way of interpreting the term 'ensure' that appears in the rule. Furthermore, I will argue that, although some interpretations of 'ensure' produce a rule that is deontological, and some produce a rule that is agent-neutral, no way of reading 'ensure' produces a rule that is both deontological and agent-neutral.

An initial option in interpreting DR is to think of 'ensure' as asking agents to guarantee that no one commits preventative killing. This leads to the following rule:

> GUARANTEE: Each agent should guarantee that no one kills to prevent more killings by others.

GUARANTEE faces a serious problem. It is too demanding as a moral rule. Even if an agent (or coalition of agents) devoted all her time and resources to stopping preventative killings it is still unlikely that she would be able to prevent every instance of preventive killing. Indeed, there will be cases where stopping multiple agents from committing preventative

killings is nomologically impossible and hence GUARANTEE violates the *ought-implies-can* principle. For these reasons GUARANTEE is implausible as a moral rule and is presumably not a rule that any deontologist would endorse. Furthermore GUARANTEE is not only implausible but may also be incoherent. For it is common to interpret the *ought-implies-can* principle as stating a conceptual necessity. Thus, a rule that violates the principle is conceptually incoherent.[18]

These problems with GUARANTEE suggest that how we read 'ensure that' needs to be sensitive to the fact that sometimes an agent is unable to stop every preventative killing and has to decide which killings to focus her limited resources on. For example, perhaps Tom, Mary and Lee are all independently about to commit preventative killings. Now suppose that: (1) you have the option of intervening to prevent Tom's preventative killing, (2) you have the option of intervening to prevent both Mary's and Lee's preventative killings, (3) you do not have the option of intervening to prevent all three preventative killings. What should you do? The answer that many will find intuitively appealing is that (all else being equal) you should choose the option of intervening to prevent Mary's and Lee's killings over the option of intervening to prevent Tom's killing. This answer is intuitive because the moral rule we are considering only appears to make sense if *someone killing to prevent more killings* is moral undesirable, and a natural way to respond to things that are morally undesirable is to minimize them (or to maximize their non-instantiation). Hence we may think of 'ensure' as asking agents to maximize or promote people not committing preventative killings. This leads to a second way of interpreting DR:

> PROMOTE: Each agent should promote no one killing to prevent more killings by others.

PROMOTE is an agent-neutral rule as it gives all agents the same ultimate aim—the aim of minimizing preventative killing. However, PROMOTE it is not a deontological rule. For, to take Dougherty's own suggested test of a deontological rule, it is not the case that PROMOTE prohibits an agent from performing a certain action for the good of there being fewer occurrences of that action overall. In fact, PROMOTE requires agents to commit a preventative killing whenever doing so is the only way to prevent more preventative killings from being committed by others. For example, consider the following case:

> PREVENTION: Tom and Mary are each about to commit a preventative killing. The only way for you to stop them is by committing one preventative killing

---

[18] See Vranas (2007), pp. 167-169 for a defence of the 'conceptual necessity' interpretation of the *ought-implies-can* principle.

yourself (perhaps, the horror of witnessing you performing a preventative killing will cause them to abandon their plans).

According to PROMOTE you must perform the preventative killing in PREVENTION because doing so will result in fewer preventative killings overall, thereby promoting it. This is in stark contrast to deontological approaches to killing which prohibit an agent from killing even for the good of less killing overall. Therefore, although PROMOTE is an agent-neutral rule, is not a deontological rule, and thus fails as an example of agent-neutral deontology.

There is a third way to interpret DR that avoids some of the problems with PROMOTE. To see this third option we need to recall the distinction Pettit makes between *promoting* and *honouring* discussed above in §3. According to Pettit, although we often promote things that are valuable, an alternative response to something valuable is to honour it. To honour a value is to exemplify that value in all your actions even when doing so will lead to less of that value overall. Thus, to honour the value of non-killing is to never kill even when doing so would prevent more killings by others. With this in mind we might disambiguate DR by interpreting 'ensure that *p*' as meaning 'honour the value of *p*'. This gives us the following rule:

> HONOUR: Each agent should honour the value of no one killing to prevent more killings by others.

Unlike PROMOTE, HONOUR appears to be a deontological rule. For, in cases like PREVENTION, HONOUR prohibits an agent from committing a preventative killing even though doing so is the only way to prevent more preventative killings by others. Thus, HONOUR is compatible with standard deontological constraints on killing such as DK. HONOUR also appears to satisfy Dougherty's goal of producing a deontological rule that not only prohibits an agent from performing preventative killings, but also requires bystanders to oppose such killings. For HONOUR requires agents to act in a way which exemplifies respect for the value of *no one killing to prevent more killings by others*. One way an agent exemplifies respect for this value is by not engaging in preventive killings herself. However, there seem to be other ways of respecting this value. For example, if an agent did not commit preventive killings herself, but nonetheless encouraged, supported, and assisted the preventive killings of other agents, she would seem to fail to exemplify respect for the value of no one committing preventative killings.

However, HONOUR is agent-relative rule and not an agent-neutral rule. For HONOUR requires each agent to exemplify respect for the value of non-preventative killings, in *her*

own actions. In this way, it gives different aims to different agents. For example, it gives Alfred the aim that: *Alfred exemplifies respect for the value of non-preventative killing in all his actions*, and gives Sita the different aim that: *Sita exemplifies respect for the value of non-preventative killing in all her actions*. Furthermore, it is this feature of HONOUR—its agent-relativity—that allows it to prohibit killing in cases like PREVENTION. Thus, although HONOUR might be the kind of deontological rule Dougherty was after, it is not an agent-neutral, deontological rule.

There is one final way to understand DR that is worth considering. Although HONOUR appears to require agents to oppose preventative killing, one might question whether it goes far enough in the demands it places on agents. HONOUR requires agents to 'exemplify respect'. But this phrase is loose and can be interpreted in different ways.[19] If I must exemplify respect for the value of *no one committing preventative killing* then what does this require of me when there are people around me committing such killings? It seems plausible that encouraging those about to perform such killings is failing to exemplify respect for this value. However, what about cases where I stand by and do nothing; neither encouraging nor discouraging a preventative killing. Have I failed to exemplify respect for the value? If the answer is 'yes' then HONOUR appears to be a satisfactory rule for the deontologist who endorses the agent-centered constraint on killing and also requires that bystanders intervene to prevent such killings. However, if the answer is 'no' then HONOUR does not go far enough for the aforementioned deontologist. In this case, the deontologist would need to supplement HONOUR with PROMOTE, making it clear that HONOUR has lexical priority. The result would be that whenever we can promote the goal of non-preventative killing without acting in a way that directly causes, or encourages a preventative killing then we should do so. This combination of HONOUR and PROMOTE may be an attractive option for the deontologist. However, importantly it does not result in a single rule that is both deontological and agent-neutral. Rather, it results in two rules—one of which is an agent-relative deontological rule, the other an agent-neutral consequentialist rule. There is nothing surprising about this result. Most deontologists accept that a plausible moral theory must contain some agent-neutral promoting rules. What makes their view deontological is that they think there are also deontological constraints that must not be violated even when doing so will bring about

---

[19] For more on the ambiguity of 'honouring a value' see Pettit (1989/1991/1997), McNaughton and Rawling (1992), and Louise (2004).

the best consequences. Therefore, even if Dougherty regards HONOUR as insufficient to cover our duties related to killing, and decides to supplement it with PROMOTE, he has still failed to produce an example of a moral rule that is both agent-neutral and deontological.

I conjecture that, apart from the three interpretations I discuss above, there are no further (minimally plausible) interpretations of Dougherty's rule available. As we have seen above, none of the three available interpretations gives us a case of agent-neutral deontology. Hence, Dougherty has failed to prove that deontology can be agent-neutral.

5. Broome's Agent-Neutral Deontology

John Broome argues that deontology can be agent-neutral in a brief passage in his *Weighing Goods*. Broome's argument, like Dougherty's, is a proof by example. However, despite similarities, there are several key differences between Broome's example and Dougherty's example, and thus they deserve separate treatment. Broome's argument is contained in the following passage:

> It is commonly believed that the ethics of side constraints is necessarily agent relative. But this is not so. Consider the view that whenever a miner is in mortal danger trapped in a mine, all available resources should be devoted to rescuing him. This will reduce the resources devoted to safety measures in mines, and so lead to the deaths of more miners in the future. Nevertheless it is what ought to be done. This is a side constraint view. But it is agent neutral. It says that all of us, equally, should contribute to saving the miner as far as we can.[20]

The view that Broome describes in this passage is a possible moral rule addressing what to do when a miner is trapped. This rule can be stated as follows:

> BR: Agents must devote all available resources to rescuing any miner who is currently trapped.

According to Broome's usage of terminology in *Weighing Goods* a side-constraint is a kind of deontological rule.[21] Thus, Broome counts BR as a deontological rule. If this assessment of BR is correct then containing BR is sufficient for a moral theory to be deontological. Broome also claims that BR is an agent-neutral moral rule. His basis for making this claim is that BR gives all agents the same ultimate aim—the aim that all available resources are

---

[20] Broome (1991), p.9.
[21] In the text Broome claims that side-constraints are a kind of nonteleological ethics. However in footnote 3, p. 20 Broome explains that his use of the phrase 'nonteleological ethics' corresponds to what others mean when they talk of 'deontology'.

devoted to any miner who is currently trapped. If Broome's assessment of BR is correct then BR proves by example that agent-neutral deontology is possible.

Importantly, in denying that BR is agent-relative, Broome accepts that BR is relative in a different, yet related, kind of way:

> [A] side-constraint theory might, perhaps, be called moment relative. Take the mining example... At present we have a reason to rescue the miner who is now trapped, and this reason fully determines what we should do now. When the next miner is trapped we shall have a reason to rescue him. That reason will fully determine what we should do then. But at present it does not count at all; it gives us no reason to save resources for the future rescue. The reason then applies at one time but not at another. So it might be called a moment-relative reason.

Thus, although Broome claims that BR gives all agents the same aim at each moment, he accepts that the ultimate aim that BR gives to all agents right now may differ from the ultimate aim BR gives to all agents at a later moment. For BR prohibits withholding resources from a trapped miner now even if that is the only way to ensure that there are enough resources in the future to save several trapped miners. It is this feature of BR that leads Broome to class it as a deontological rule. For, like Kamm and Dougherty, Broome thinks that deontological rules must prohibit performing an action even when doing so will lead to fewer occurrences of that type of action overall.

In this respect Broome's purported example of agent-neutral deontology is substantially different from Dougherty's example. For Broome's example employs moment-relativity and suggests the thesis that, although deontology might not be necessarily agent-relative, it is necessarily moment-relative (or necessarily either moment or agent-relative). By contrast Dougherty's example made no reference to moment-relativity and left no room for a connection between deontology and some kind of relativity.

The possibilities raised by Broome's example are intriguing. However, the question of whether BR succeeds as an example of a rule that is both agent-neutral and deontological remains. At first glance BR may appear to give all agents the same aim—*that all available resources are devoted to rescuing a trapped miner*. However, this impression may be based on a mistake. For we tend to think of rescue cases like Broome's as cases of cooperation where everyone acts in unison towards a common cause. We see the 'community' responding to the miner's plight and sacrificing shared resources to save her. Yet the crucial question for testing the status of BR is: *what does BR require of agents when there is an absence of cooperation?* To answer this question we need to know what BR requires in the following kind of case:

TRUCK: A miner is trapped and the only aid you can offer is your dump truck, which will be helpful in removing rubble from the mine. Two of your rivals also have dump trucks that are needed and have offered them to the rescue effort. However, you learn that, because of their professional jealousy, they will withdraw their offer to help if you proceed with your offer. Hence, if you give your dump truck to the rescue effort there will be fewer dump trucks overall to aid the rescue. Thus you must choose between the rescue effort being slightly less effective but containing your contribution, or lacking your contribution but being more effective overall.

What should an agent do in cases like TRUCK? One answer is that the agent must still make a full contribution to the rescue effort in such a case. The following version of BR gives this answer:

(BR′) Each agent must devote all her available resources to rescuing a trapped miner, even if doing so will lead to fewer resources overall being devoted to the rescue.

BR′ appears to be a deontological rule. However, BR′ is an agent-relative rule as it gives different agents different ultimate aims. For it gives each agent the aim that *she* devotes her available resources. Thus it is not an example of an agent-neutral deontological rule.

An alternative answer to the question above is that the agent should withdraw her aid in cases like TRUCK. The following version of BR gives this answer:

(BR″) Each agent must promote the general devoting of available resources to rescuing a trapped miner.

Unlike BR′, BR″ is an agent-neutral rule. For BR″ gives all agents the same ultimate aim— the aim that as many resources as possible are devoted to the rescue effort. However, BR″ is not a deontological rule. For BR″ requires agents to promote something. Furthermore, BR″ fails to contain a deontological constraint. For there is no action type that BR″ prohibits performing even when doing so is the only way to prevent more occurrences of that act-type.

Above I have shown how there are two distinct readings of Broome's rule BR. Perhaps some will prefer to read Broome's rule as the agent-relative BR′ and others will prefer to read it as the agent-neutral BR″. However, whichever reading is preferred, Broome has not given us an example of a rule that is both agent-neutral and deontological.

6. A Generalised Argument Against Agent-Neutral Deontology

In the previous three sections I reviewed arguments by Kamm, Dougherty, and Broome for the conclusion that deontology can be agent-neutral. I showed that all three arguments

were unsound and each failed on independent grounds. The failure of these three arguments may lead some to question whether any argument for agent-neutral deontology can succeed. In this section I will address this by presenting a generalised argument against agent-neutral deontology. This argument will demonstrate that any moral theory properly characterised as deontological cannot be an agent-neutral moral theory.

The obvious starting point for the generalised argument against agent-neutral deontology is an account of what it is for a moral theory to be 'deontological'. However, giving such an account is a difficult task as a wide range of moral views with different justifications can be found under the umbrella of deontology. As a result, accounts of deontology that initially look plausible may fail if they do not accommodate the many diverse moral theories that are taken to be instances of deontology. For example, it is sometimes said that deontological theories give priority to the right over the good or that they are non-teleological theories (in contrast to teleological theories like consequentialism).[22] It is true that many deontological theories can be characterised in this way. However, in §3 we saw that some deontological theories start with an account of what is intrinsically valuable but differ from consequentialist theories by arguing that agents ought to honour this value rather than promote it. Such theories appear to be teleological and give priority to the good and yet are non-consequentialist theories that give standard deontological moral verdicts. Accordingly, deontology cannot be characterised in terms of non-teleology or giving priority to the right.

Nonetheless, there is an account of deontology that adequately captures the diverse theories that fall under its umbrella. According to this account the essential feature that all deontological theories share is that of containing a constraint that prohibits performing actions of some act-type. This account can tell us what teleological and non-teleological versions of deontology have in common such that it is appropriate to place them in the same category. They count as deontological because they both place constraints on certain actions (for example, they might both contain a constraint on killing innocent people). By contrast, a consequentialist theory does not contain such constraints, and is thus distinguished from deontology. This account is promising, yet more needs to be said about the nature of the constraints that characterize deontology. For if a moral constraint is any prohibition on an act-type then consequentialist theories contain constraints because they prohibit the act-type of *failing to maximize the good*. Thus we need a thicker account of the moral constraints that deontological theories contain. One suggestion found in Kagan (1989) is that deontological constraints are constraints on promoting the good. All

---

[22] Classic statements of these views are found in Frankena (1973) and Rawls (1971).

deontological moral theories contain a constraint that prohibits promoting the good in some circumstance. On the other hand, all consequentialist theories lack constraints of this kind. This account is promising but it will not suit all versions of deontology. For, although many deontologists accept a prima facie duty to promote the good, some deontologists eschew the notion of good simpliciter, yet still appear to endorse a deontological constraint.[23] Furthermore, some deontologists reject the idea that deontology is a kind of afterthought to consequentialism. [24] An alternative to Kagan's formulation that avoids these problems takes deontological constraints to be constraints on performing an act-type even when performing it will bring about fewer occurrences of that act-type overall. This account of deontological constraints is found in Nozick (1974) and is widely endorsed by deontologists.[25] Indeed, those who employ Kagan's formulation, appeal to non-teleology, or appeal to the priority of the right, typically claim that these things result in agents being prohibited from performing an action (such as killing an innocent person) even if doing so is the only way to prevent several occurrences of that action. On the other hand, consequentialism appears to lack such constraints. Thus, the essential property that all deontological theories share is that of containing a constraint on performing an act-type even when performing it will bring about fewer occurrences of that act-type overall. [26]

Given this account of deontological constraints, here is a first attempt at characterising deontology:

> **DT**: A theory is deontological *iff* (i) it contains a constraint on performing acts of some act-type *T*, and (ii) this constraint prohibits performing acts of type *T* even when doing so is the only way to prevent more occurrences (either in general or below a relevant threshold) of acts of type *T*.

DT appears to adequately capture the notion of a deontological constraint. It preserves the central idea that a deontological constraint must prohibit some action even for the goal of fewer occurrences of that action overall. It also allows for a deontologist to do this

---

[23] For example, see Foot (1985), Thomson (1993), and Taurek (1977).
[24] See Dougherty (2013) p. 528.
[25] It is also advanced in Scheffler (1982).
[26] Some might nonetheless reject this starting point because they believe that a theory with no moral constraints could still be deontological. For example, perhaps they count the 'prerogativism' of Scheffler (1982) as a deontological theory (in opposition to Scheffler's own characterisation of it as a 'hybrid theory') because, although the theory contains no constraints on promoting the good, it does contain an *option* not to promote the good. Anyone with this kind of view can read the argument that follows as showing that the subset of deontological theories that contain constraints are necessarily agent-relative. This is in itself an important conclusion given that several authors have thought they could show that an ethical view containing deontological constraints could be agent-neutral.

absolutely or via a threshold, and thereby accommodates absolutist and non-absolutist varieties of deontology. However, there is a problem with DT. Some theories that are clearly not versions of deontology can be formulated in such a way that they satisfy conditions (i) and (ii) and thus count as deontological according to DT. For example, according to classical utilitarianism each agent is required to promote total net pleasure. If any theory is not a deontological theory then classical utilitarianism is not. However, it is possible to formulate the central rule of classical utilitarianism as a constraint as follows:

> UC: Each agent must not fail to maximize total net pleasure

A theory with such a constraint satisfies condition (i) of DT. Furthermore, in classical utilitarianism the constraint UC is absolute, meaning there are no conditions where an agent is morally permitted to violate it. Thus, condition (ii) is also satisfied because, if there are no circumstances where one is allowed to violate UC, then one is not allowed to violate it in circumstances where violating it results in fewer violations overall. However, if UC satisfies conditions (i) and (ii) then, according to DT, classical utilitarianism is a deontological theory. Yet this is clearly false and thus DT fails as a definition of deontology. Nonetheless, DT appears to be on the right track. It gives the wrong verdict with respect to UC because UC satisfies condition (ii) trivially merely by it being impossible for the circumstance described in (ii) to occur in UC. For these circumstances to occur there would need to be a case where an agent can cause several other agents to maximize total net pleasure but only at the cost of failing to maximize total net pleasure herself. However, if I perform an action because it causes others to act in a way that will maximize net pleasure, then in performing that action I am myself maximizing net pleasure. Similar reasoning shows that this problem arises for other maximizing moral theories such as ethical egoism. Such theories are clearly not deontological and yet satisfy condition (ii) in this trivial way.

To resolve this problem we need to reformulate DT so that it rules out gimmicky constraints like UC that trivially satisfy condition (ii). I suggest the way to do this is to add a condition that requires there to be a possible circumstance where agents can prevent more overall instances of some act-type by performing an instance of that act-type. Thus, we get the following:

> **DT′**: A theory is deontological *iff* (i) it contains a constraint on performing acts of some act-type *T*, (ii) there is some possible circumstance where the only way for an agent to prevent more occurrences of acts of type *T*, is to perform an act of type *T*, and (iii) the constraint prohibits (either in general or below a relevant threshold) an agent performing acts of type *T* in these circumstances

The second condition in DT′ rules out gimmicky constraints like UC and thus, according to DT′, UC is not a deontological rule. On the other hand, all the standard types of constraints that deontologists endorse appear to satisfy conditions (i)-(iii). Thus DT′ succeeds as a criterion for classifying deontological theories via their possession of deontological constraints.[27]

The next step in my argument is to show that any theory satisfying the conditions in DT′ is an agent-relative theory. Recall from §2 that a moral theory is agent-relative just in case it contains at least one agent-relative rule. Thus, my argument must show that any constraint satisfying the conditions in DT′ is an agent-relative rule. If this is proven then it follows via the definition in DT′ that all deontological moral theories are agent-relative.

At first glance it may appear easy to prove that any constraint satisfying the conditions of DT′ is agent-relative. For any constraint on performing an act-type appears to give different agents different aims and thus count as agent-relative. For example, a constraint on performing acts of type T appears to give Alfred the ultimate aim that Alfred does not perform an instance of T and Sita the ultimate aim that Sita does not perform an instance of T.

However this simple proof fails because it neglects a clarification we made in §2 when explaining agent-relativity. The clarification said that a rule is agent-neutral if and only if there is a valid interpretation of it that gives all agents the same aim, and agent-relative otherwise. This clarification was necessary because agent-neutral rules like MK (the minimize killing rule) can be interpreted as giving different agents different aims, even though they are most naturally interpreted as giving all agents the same aim. On the other hand, agent-relative rules can only be interpreted as giving all agents different ultimate aims. It follows from this that to show that any constraint that fits DT′ is agent-relative it is not enough to show that any such constraint can be interpreted as giving different agents different ultimate aims. We also need to show that there is no constraint that meets the conditions of DT′ and yet can be interpreted as giving all agents the same ultimate aim.

---

[27] One might object that DT′ fails as a criterion because it wrongly counts agent-relative consequentialism as a version of deontology. After all, agent-relative consequentialism appears to contain deontological constraints. However, this objection is mistaken. Agent-relative consequentialism may be capable of giving the same deontic verdicts as any deontological theory. However, it still fails to meet all the conditions in DT′. Specifically it fails condition (ii) because the rules it contains are promotion rules and yet circumstances where one fails to promote T in order to bring about more promoting of T by others are impossible. They are impossible because, by bringing it about that there is more promoting of T by others, an agent is herself promoting T. Thus, DT′ correctly counts agent-relative consequentialism as a non-deontological theory.

Another way to put the point is as follows. The proof that deontology is necessarily agent-relative must show that it is impossible for there to be a moral constraint that meets conditions (i)-(iii) below and also meets condition (iv):

(i)   It is a constraint on performing acts of act-type T

(ii)   There is some possible circumstance (C) where the only way for an agent to prevent more occurrences of acts of type T, is to perform an act of type T

(iii)   The constraint prohibits (either in general or below a relevant threshold) performing acts of type T in circumstance C.

(iv)   There is a way of interpreting the aims of the constraint such that it gives all agents the same ultimate aims.

I will now offer a proof that this is indeed impossible. First note that condition (i) requires that there be a constraint with the following form:

(1) Each agent must not perform acts of type T

Now T could be any act-type at all. So the crucial question to ask is: *is there any act-type that, when plugged into (1), allows (1) to be interpreted such that it gives all agents the same ultimate aims?* For logically simple act-types the answer is 'no'. For example, suppose that $T = $ *kill an innocent person*. Thus we end up with the following constraint:

(2) Each agent must not kill an innocent person

The constraint (2) satisfies conditions (i)-(iii). It satisfies (i) because 'kill an innocent person' is an act-type. It satisfies (ii) because it is possible for there to be a circumstance where the only way for an agent to prevent several killings of innocent people is by killing an innocent person herself. Finally, it satisfies condition (iii) because it prohibits killing an innocent person in the circumstances described in (ii). However, (2) does not satisfy (iv). For, there is no way of interpreting (2) such that it gives all agents the same ultimate aim. In fact, (2) is only able to pass condition (iii) because it gives each agent the aim that *she* does not kill an innocent person and thus it gives different agents different aims.

Logically simple act-types like the act-type in (2) will all fail to satisfy (iv). However, there are complex act-types that can be plugged into (1) and result in all agents having the same ultimate aims. For example, suppose that $T = $ *fail to make it the case that no innocent person is killed*. Thus we end up with the following constraint:

(3) Each agent must not fail to make it the case that no innocent person is killed

Or, eliminating the double negative:

(3') Each agent must make it the case that no innocent person is killed

The constraint (3) can be interpreted as giving all agents the same ultimate aim. For it appears to give all agents the aim *that no innocent person is killed*. Thus, (3) satisfies (iv). However, there is a problem with constraints like (3). They fail to satisfy condition (ii). For there are no possible circumstances where an agent fails to make it the case that no innocent person is killed and, in doing so, prevents more occurrences of other agents failing to make this the case. For, insofar as any innocent person is killed every agent has failed to make it the case that no innocent person is killed. Thus, one agent cannot fail at this goal to a greater degree so that other agents fail at it to a lesser degree. All agents succeed or fail at this goal equally.

The problem just highlighted is a problem that occurs for any complex act-type that, when plugged into (1), can be interpreted as giving all agents the same ultimate aims. Such act-types will fail condition (ii) because, by giving all agents the same aim, they will not allow there to be circumstances where several agents can better fulfil their theory-given aims if one agent fails to fulfil her theory-given aims. Thus, the only act-types that are able to satisfy condition (ii) are logically simple act types like the one found in (2). However, such act types necessarily give different agents different ultimate aims.

In summary, we can see that, whatever act-type we take T to be, T will either give all agents the same ultimate aims but fail condition (ii), or satisfy condition (ii) but give different ultimate aims to different agents. It follows from this that no constraint that satisfies conditions (i)-(iii) can be interpreted as giving all agents the same ultimate aims. Thus, according to the definition given in DT′, every deontological theory will contain a constraint that cannot be interpreted as giving all agents the same ultimate aims. Therefore, every deontological theory is agent-relative.

7. Conclusion

This chapter has defended the thesis that deontology is necessarily agent-relative. I showed how several prominent attempts to argue that deontology can be agent-neutral failed. Furthermore, I presented a general argument to prove that any deontological moral theory will be an agent-relative theory. In doing this I argued that deontological moral theories are conceptually tied to deontological constraints and this is best cashed out by DT′. The upshot of these arguments is that we have a clearer understanding of the nature of deontology. We have clarified what features are essential to all deontological moral theories and established that such theories are necessarily agent-relative. Such a result is important and interesting in its own right. However, it is worth reviewing a couple of contemporary debates where this result has further applications. First, as we saw in §1,

several have argued that agent-relative moral theories are implausible because they fail to give all agents common moral aims. By showing that all deontological theories are agent-relative we have shown that the viability of deontology is at stake in this debate. Thus, anyone hoping to defend a deontological moral theory will have to confront this objection. Second, one response to 'consequentializing' morality is 'deontologizing' morality. Agent-neutral deontology appears to provide a strategy for doing so. However, by showing it is impossible we have shown that this strategy fails.

References

Broome, J. (1991). *Weighing Goods*. New York: Blackwell.

Dougherty, T. (2013). Agent-Neutral Deontology. *Philosophical Studies*, 163: 527-537.

Dreier, J. (1993). Structures of Normative Theories. *The Monist*, 76: 22-40.

Foot, P. (1985). Utilitarianism and the Virtues. *Mind*, 94: 107-123.

Frankena, W. (1973). *Ethics*. Eaglewood Cliffs: Prentice-Hall.

Hammerton, M. (2017). Is Agent-Neutral Deontology Possible? *Journal of Ethics and Social Philosophy*, 12: 319-324.

Hurley, P. (2013) Consequentializing and Deontologizing. *Oxford Studies in Normative Ethics*, 3: 123-153.

Kagan, S. (1989). *The Limits of Morality*. Oxford: Clarendon Press.

Kamm, F.M. (1989). Harming Some to Save Others. *Philosophical Studies*, 57: 227-260.

Kamm, F. M. (1992). Review: Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status. *Philosophy and Public Affairs*, 21: 354-389.

Kamm, F. M. (1996). *Morality, Mortality: Volume II: Rights, Duties, and Status*. Oxford University Press.

Kuflik, A. (1986). A Defense of Common-Sense Morality. *Ethics*, 96: 784-803.

Louise, J. (2004). Relativity of Value and the Consequentialist Umbrella. *Philosophical Quarterly*, 54: 518-36.

McNaughton, D. & Rawling, P. (1991). Agent-Relativity and the Doing-Happening Distinction. *Philosophical Studies*, 63: 167-185.

McNaughton, D. & Rawling, P. (1992). Honoring and Promoting Values. *Ethics*, 102: 835-843.

Nagel, T. (1986). *The View from Nowhere*. New York: Oxford University Press.

Nozick, R. (1974) *Anarchy, State, and Utopia*. New York: Basic Books.

Parfit, D. (1984) *Reasons and Persons*. Oxford: Clarendon Press.

Pettit, P. (1989) Consequentialism and Respect for Persons. *Ethics*, 100: 116-126.

Pettit, P. (1991) Consequentialism. In Singer (Ed.), *A Companion to Ethics*. Oxford: Blackwell.

Pettit, P. (1997) The Consequentialist Perspective. In Baron, Pettit, and Slote (Eds.). *Three Methods in Ethics*. Oxford: Blackwell.

Portmore, D. (2001). Can an Act-Consequentialist Theory be Agent Relative? *American Philosophical Quarterly*, 38: 363-377.

Portmore, D. (2014). Agent-Relative vs. Agent-Neutral. In Hugh LaFollette (ed.), *International Encyclopedia of Ethics*. Wiley Blackwell.

Rawls, J. (1971). *A Theory of Justice*. Cambridge: Harvard University Press.

Ridge, M. (2005a). Agent-Neutral vs. Agent-Relative Reason. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (Ed.), URL = http://plato.stanford.edu/archives/win2005/entries/reasons-agent/

Ridge, M. (2005b). Universalizability for Collective Rational Agents: A Critique of Agent-relativism. *Philosophy and Phenomenological Research*, 70: 34-66.

Scheffler, S. (1982). *The Rejection of Consequentialism*. Oxford: Clarendon Press.

Taurek, J.M., (1977). Should the Numbers Count? *Philosophy and Public Affairs*, 6: 293-316.

Thomson, J. J. (1993). Goodness and Utilitarianism. *Proceedings and Addresses of the American Philosophical Association*, 67: 145-159.

Tradros, V. (2011). *The Ends of Harm*. Oxford: Oxford University Press.

Vranas, P. (2007). I Ought, Therefore I Can. *Philosophical Studies*, 136: 167-216.

4

A Trilemma for Deontologists[1]

1. Introduction

Deontological moral theories standardly endorse agent-centered constraints. They hold that in certain circumstances an agent must not perform an act of a certain type even if doing so is the only way to prevent more instances of that act being performed by others. For example, a deontological approach to killing will prohibit an agent from killing even when performing one killing is the only way to prevent more killings being performed by others. That deontology has this feature is widely taken as a point in its favour. For commonsense morality also endorses agent-centered constraints, and thus deontology is able to accommodate a key commitment of moral commonsense. By contrast, agent-neutral consequentialism—the historical rival of deontology—is incompatible with agent-centered constraints and thus faces the objection that it conflicts with commonsense morality.

In this chapter I raise a trilemma for deontologists that stems from their commitment to agent-centered constraints. My trilemma starts from the observation that, for any act-type where a deontologist might endorse an agent-centered constraint, there are in fact several logically distinct rules that could give the required constraint. For example, there are several logically distinct rules related to the act-type of killing that prohibit killing even when it prevents more killing by others. Thus, the deontologist must specify which of these possible rules she endorses as her agent-centered constraint. However, I argue that each possible rule has one of three counterintuitive consequences that appear to count against it. First, some rules involve interpreting deontological constraints as maximizing rules. Yet many deontologists reject this interpretation of deontological constraints because it appears to collapse into agent-relative consequentialism. Second, some options involve

taking morality to give no advice in cases where commonsense morality holds that there is moral advice. Third, some options involve accepting a strong anti-aggregation position that conflicts with commonsense morality. The upshot is that deontology's ability to accommodate the agent-centered constraints of commonsense morality involves serious counterintuitive costs.

2. Deontology and Constraints

It is widely accepted that all deontological moral theories contain agent-centered constraints.[2] Indeed, many hold this to be the distinguishing feature of deontology. However, when it comes to the question of what constraints there are, deontologists have a variety of opinions. Some hold that all constraints can be reduced to a single fundamental constraint such as a constraint on doing harm or intending harm. Others, hold that there are multiple constraints that are logically independent from one another. For example, some endorse a constraint on promise-breaking that applies even in cases where no harm is intended or done by a broken promise. So there are many possible forms that a deontological theory can take.[3] The arguments I offer in this paper focus on general logical features shared by all agent-centered constraints rather than the content of specific constraints. As such they apply to all possible deontological theories regardless of which constraints they endorse. Before moving on to these arguments a couple of further remarks are necessary.

First, we should note that agent-centered constraints can hold either absolutely or only below a relevant threshold. For example, some deontologists hold that killing an innocent person is wrong even in situations where it is the only way to prevent a very large number of other innocent people from being killed, whereas others hold that it can be justified if the number that will be saved is high enough (e.g. killing an innocent person to spare a million other innocent people). The arguments I offer here apply to both absolute and threshold constraints, although to simplify things all the examples I use will involve absolute constraints.

Second, some theorists have argued against the standard view that deontology is necessarily agent-relative, claiming that agent-neutral deontology is possible (Dougherty 2013a). I do not think that these arguments are successful, and argue against them in chapter 3 of this thesis. However, even if they did succeed, they would not undermine my

---

[2] For example, see Nozick (1974), Scheffler (1982), Parfit (1984), Nagel (1986), Darwall (1986), Kagan (1989), McNaughton and Rawling (1991), Dreier (1993), Kamm (1993), and Pettit (1997).
[3] For an overview of some of these options see Kagan (1998), pp.94-153.

claim that all deontological theories contain agent-centered constraints. This is because I am using the terminology 'agent-centered constraint' to pick out constraints that prohibit an agent from performing an act-type even when doing so is the only way to prevent more occurrences of that act-type being performed by others. I am not making any further claim that constraints of this kind have the logical property of agent-relativity and lack the logical property of agent-neutrality. My claim that all deontological theories contain agent-centered constraints in this sense is something that agent-neutral deontologists can, and will, agree with. Indeed, Dougherty (2013a) argues that deontology can be agent-neutral by arguing that it is possible to produce an agent-neutral constraint that prohibits killing to prevent more killing by others.

Finally, although all deontological theories contain agent-centered constraints, it does not follow that all moral theories containing agent-centered constraints are deontological theories. Recently, a new kind of consequentialism—often called 'agent-relative consequentialism'—has been proposed that appears to make room for agent-centered constraints by postulating that the good can be agent-relative.[4] Proponents of this theory claim that it can capture all the commonsense moral verdicts of deontology with the attractive theoretical framework of consequentialism.[5] Deontologists typically respond by either rejecting agent-relative consequentialism as incoherent, or arguing that deontology provides a more attractive theoretical framework.[6] The problems I raise for deontology in this chapter stem partly from the challenge posed by agent-relative consequentialism.

3. Four Possible Constraints

Above we saw that all deontological theories endorse at least one agent-centered constraint. In this section I will argue that, for any action a deontologist might want to constrain, there are in fact four logically distinct moral rules that do equally well at capturing the agent-centered verdict that an agent must not perform that action even to prevent several others from performing it. To make this argument I will first focus on an agent-centered constraint on promise-breaking and then demonstrate how the conclusions reached about this constraint apply to any agent-centered constraint. Consider the following case:

> PROMISE: Sylvia has promised to meet a friend at 2pm to help him complete his tax return. She hails the last available taxi and is about to head off to the meeting

---

[4] For versions of this view see: Sen (1982), Broome (1991), Dreier (1993), Portmore (2001/2011), Smith (2003/2009), and Louise (2004).
[5] See especially Dreier (1993), Smith (2009), and Portmore (2011) on this point.
[6] See Schroeder (2007) for an example of the former approach and Hurley (2013) for an example of the latter.

when she realizes that two nearby strangers also want the taxi. Each stranger also needs to meet a friend at 2pm to keep a (comparable) promise. By chance their meetings are taking place in the same location, whereas Sylvia's meeting is at the other end of town. If she takes the taxi for herself then she will keep her promise to her friend, but the two strangers will break their promises. If she allows the strangers to take the taxi then they will keep their promises but she will break hers.

What should Sylvia do in this case? Commonsense morality appears to require Sylvia to keep her promise even though breaking it would result in two others keeping their promises. In response to this kind of case many deontologists endorse an agent-centered constraint on promise-breaking. It is often assumed that doing so is a simple matter, as it is assumed that there is only one candidate agent-centered constraint that prohibits promise-breaking in this kind of case. However, this is a mistake. There are in fact four candidate agent-centered constraints that achieve this result:

(1) Each agent must not break promises.

(2) Each agent, when she has the ability to do so, must not break promises.

(3) Each agent must minimize her own promise-breaking.

(4) Each agent, at each moment, must minimize her promise-breaking at that moment.[7]

Each of these rules prohibits promise-breaking to prevent more promise-breaking by others. Thus, each rule concurs with the verdict of commonsense morality in cases like PROMISE. However, these rules are each logically distinct (and the choice between them is a morally significant one) because there are possible cases where each rule gives a different deontic verdict. For example, consider:

SYNCHRONIC: Sylvia has independently made comparable promises to Asha, Beth, and César. Each promise requires Sylvia to perform a specific task now. However, it is not possible for Sylvia to perform all three tasks simultaneously. The tasks owed to Asha and Beth can be performed simultaneously. However, the task owed to César cannot be performed in conjunction with the others. Thus, Sylvia has two options: (i) break the promise to César but keep the promises to Asha and Beth, (ii) keep the promise to César but break the promises to Asha and Beth.

What do the four rules say about this case? According to (1), Sylvia must not break promises. In SYNCHRONIC Sylvia has two options and each option involves breaking a promise. Thus, rule (1) gives the verdict that Sylvia does the morally wrong thing whatever option she takes. If she takes the first option she does the wrong thing by breaking a

---

[7] This list could be extended by adding further logically possible constraints. However, the four constraints I have listed capture what I take to be the four main kinds of agent-centered constraints. Any further logically distinct constraints that give the correct verdict in PROMISE will be, I claim, relevantly similar to either (1), (2), (3), or (4).

promise to César. If she takes the second option she does the wrong thing by breaking promises to Asha and Beth. Thus, rule (1) classifies SYNCHRONIC as a strong moral dilemma ('strong' because it cannot be resolved by pointing out that one of Sylvia's conflicting moral duties overrides the other). [8]

According to (2), Sylvia must not break a promise whenever she has the ability not to break it. To derive a verdict from (2) in SYNCHRONIC we must first ask whether Sylvia has the ability to keep her promises in this case. On one possible view, Sylvia does have the ability to keep her promises because doing the tasks required to keep her promise to César is within her powers, and doing the tasks required to keep her promises to Asha and Beth is also within her powers. On this view, (2) gives the same verdict as (1) in SYNCHRONIC— the verdict that Sylvia is in a strong moral dilemma. According to a second possible view, Sylvia does *not* have the ability to keep her promises in SYNCHRONIC. For, although she has the ability to keep each promise considered alone, she lacks the ability to keep all three promises considered together, and the latter is the relevant ability. On this view, (1) and (2) give different verdicts in SYNCHRONIC. Rule (1) gives the verdict that Sylvia acts wrongly. Rule (2) gives no verdict, classing Sylvia's action as neither right nor wrong because the ability clause is not met.

I will not take a stand in this paper on which of these two views is correct. However, it's important to note that on the first view rule (2) gives the same deontic verdicts as (1) in all cases discussed in this paper. Thus, I will ignore this interpretation of (2) as anything I might say about it will already be covered by what I say about (1). Henceforth, whenever I mention (2) I will interpret it according to the second view where it sometimes gives a different deontic verdict from (1).

So far, we have seen that rules (1) and (2) give different verdicts in SYNCHRONIC; (1) gives the verdict that each of the available options is wrong whereas (2) (given the caveat above) classes each option as neither right nor wrong. Let's now consider the deontic verdicts given by rules (3) and (4). According to (3) Sylvia must minimize her overall promise-breaking. In SYNCHRONIC Sylvia does this by taking the first option as this option results in her breaking fewer promises than the second option. According to (4) Sylvia must minimize her present-moment promise-breaking. In SYNCHRONIC Sylvia does this by taking the first option as this results in her breaking fewer promises in the present-moment than in the second option. Therefore, (3) and (4) give the same deontic verdict in SYNCHRONIC. However, the verdict they give is distinct from the verdicts given by (1)

---

[8] See Sinnott-Armstrong 1988, Chapter 1, for a detailed explanation of the difference between 'genuine' or 'strong' moral dilemmas and 'weak' moral dilemmas.

and (2) in this case. Thus, we have seen that (1) and (2) are distinct from each other and distinct from (3) and (4). To show that (3) and (4) are distinct we need to consider another case:

> DIACHRONIC: Sylvia has independently made comparable promises to Asha, Beth, and César. The promise to César is due now, whereas the promises to Asha and Beth are due next week. Unfortunately, an unforeseen complication has made it impossible for Sylvia to keep all three promises. She has two options: (i) break the promise to César now, which will allow her to keep the promises to Asha and Beth next week, (ii) keep the promise to César now, in which case she will break the promises to Asha and Beth next week.

In DIACHRONIC (3) and (4) give different deontic verdicts. According to (3), Sylvia must take option (i) as doing so minimizes her *overall* promise-breaking by causing her to break one promise overall rather than two. However, according to (4) Sylvia must take option (ii) as doing so minimizes her *present-moment* promise-breaking. For (4) requires an agent at each moment to minimize her promise-breaking only at that moment, even if this means she will break more promises at future-moments. Sylvia does this in DIACHRONIC by keeping her present promise to César, even though this will result in her later breaking promises to Asha and Beth.

Finally, let's examine the deontic verdicts of (1) and (2) in DIACHRONIC. Each rule gives the same verdict as it gave in SYNCHRONIC. Rule (1) counts Sylvia as doing the morally wrong thing whatever option she takes, as each option involves breaking a promise. Rule (2) does not class either option as right or wrong because in DIACHRONIC Sylvia does not have the ability to keep all her promises and thus the constraint does not apply to her.

So far we have seen that a deontologist who endorses an agent-centered constraint on promise-breaking must decide whether that constraint is (1), (2), (3), or (4). This result can be generalized; for any agent-centered constraint a deontologist endorses there are four logically distinct rules corresponding to (1), (2), (3), and (4) that could constitute that constraint. This is because for any agent-centered constraint cases like SYNCHRONIC and DIACHRONIC can be constructed. For example, consider an agent-centered constraint on killing that requires each agent to refrain from killing even to prevent more killings by others. Now consider the following case:

> GRENADE: Carlo the anarchist is throwing a grenade into a crowd of people. As he accelerates his throwing arm he realizes the wrongness of what he is doing and loses his desire to kill. However, unfortunately the momentum of his arm and his loose grip of the grenade make it impossible for him to avoiding throwing the grenade into the crowd. All he can do is control the direction of the grenade by flicking his wrist to the left or the right. If he flicks his wrist to the left, the grenade

will kill one person. If he flicks his wrist to the right, it will kill two (different) people. Carlo has no other options other than to flick left or right.

GRENADE is a case just like SYNCHRONIC. Like SYNCHRONIC, we might hold that Carlo is required to minimize his killing by acting so as to kill the smaller number. Alternatively, we might say that Carlo is in a moral dilemma where, no matter what he does, he will do the morally wrong thing. Finally, we might say that, because Carlo lacks the ability to avoid killing anyone, the action he takes is neither right nor wrong. We can also construct a killing case analogous to DIACHRONIC by modifying GRENADE such that the one is killed instantly by the grenade but the two are killed at a significant delay (perhaps the grenade becomes jammed, but will ignite after a few minutes, killing the two). In the sections that follow, I focus on the constraint on promise-breaking, showing how a trilemma arises from the need to choose one of four rules. However, the arguments I present can be applied to any agent-centered constraint a deontologist might endorse, as all such constraints involve the same problematic set of options.

We have seen that a deontologist must clarify which of four possible rules constitute the agent-centered constraints she endorses. I will conclude this section by reviewing some reasons why a deontologist might reject some of these rules and endorse others. First, some deontologists (most notably Ross 1930) reject the possibility of moral dilemmas and so cannot accept rules like (1) that result in such dilemmas. Second, some deontologists reject the *ought-implies-can* principle. Rules like (2) only appear plausible if the ought-implies-can principle is true. Thus, deontologists rejecting the principle will typically reject such rules. Finally, some deontologists attempt to justify agent-centered constraints by appealing to facts about moral agents such as the moral importance of agent integrity and keeping clean hands, whereas others appeal to facts about moral patients such as their inherent dignity and inviolability. Deontologists who adopt the former justification often hold that, in cases like DIACHRONIC, an agent must minimize her violations of a constraint, and thus endorse the verdict of (3). Deontologists who adopt the latter justification often hold that an agent must never subject a patient to the violation of a constraint, even if doing so is the only way for the agent to prevent herself from later violating the same constraint with respect to several other patients. In doing so they endorse the verdict of (1) or (4) in cases like DIACHRONIC.[9] Yet, although deontologists holding these different positions have endorsed the verdicts of either (1), (3), or (4), they

---

[9] The former position can be found in McNaughton and Rawling (1993), and the latter in Brook (1991) and Kamm (1992).

have generally failed to notice that there is a choice to be made between the four rules I have listed here, and so failed to be clear on which of the four they endorse.

Before moving on it is worth noting a small qualification in how we might think of rule (4). Rule (4), unlike the other three rules, is moment-relative. It gives priority to the present moment over future moments by requiring an agent *not* to break a promise in the present even if doing so is the only way for that agent to ensure that she does not break several promises in the future. Those who favour this rule will see its moment-relativity as an advantage because it prohibits certain ways of minimizing one's overall promise-breaking that they find morally objectionable. For example, it prohibits breaking a promise to one promisee now as a means to the end of keeping more promises to others promisees. It does this because, unusual cases aside, a means always temporally precedes the end it brings about. However, the moment-relativity of (4) also leads to a problem. By requiring agents to focus exclusively on their present promise-breaking, (4) fails to require them to take steps to prevent their future promise-breaking. We have already seen that this may be an advantage when it comes to using a present promisee as a *means* to keeping future promises. However, in other cases it is a disadvantage because it clearly conflicts with moral commonsense. For example, consider a case where the only way to ensure that you keep a promise due next week is to perform a task now that is not very demanding and does not violate anyone's rights or sacrifice anything of moral significance. Clearly, if you have made the promise then you must take any reasonable steps necessary to ensure that you keep it. Rules (1), (2), and (3) all require this of you in this kind of case. However, rule (4) does not because it only focuses on avoiding promise-breaking in the present-moment. This problem may appear to be a serious one, but it is easily rectified. To avoid the problem someone who endorses (4) must also endorse rule (3), and stipulate that (4) has lexical priority over (3). It follows that in cases like DIACHRONIC where rules (3) and (4) conflict, an agent must always obey rule (4). However, in cases where there are no present promises being traded-off, and thus no conflict between (3) and (4), rule (3) becomes relevant. In these case rule (3) requires us to take any reasonable steps necessary to prevent future promise-breaking.[10]

4. The First Horn: Losing what is Distinctive about Deontology

---

[10] Interestingly, several agent-relative consequentialists have suggested that agent-relative goods also need to be moment-relative in order to prohibit using someone as a means to prevent your future violations of a constraint (e.g. Smith 2003/2009, and Louise 2004). However, none have noticed the point I make here that there would also need to be a moment-neutral agent-relative good that is lexically dominated by the moment-relative good, but makes a moral difference in cases like the ones I discuss where the moment-relative good is not threatened.

In the previous section we saw that a deontologist who endorses an agent-centered constraint on promise breaking must clarify which of four possible rules constitute that constraint. In this section I will raise a problem for any deontologist who interprets the constraint as (3) or (4). This problem will constitute the first horn of my trilemma.

Consider a deontologist who endorses (3) or (4). We have seen above that these two rules give different deontic verdicts in DIACHRONIC. However, despite this difference, these rules are structurally similar. Both rules are 'maximizing rules', giving agents a goal, or end, to aim at, and instructing them to bring about as much of this goal as they can. By contrast, (1) and (2) are *not* maximizing rules as neither rule requires agents to maximize a quantity or bring about a goal or end.

The fact that (3) and (4) are maximizing rules creates an initial worry for the deontologist who would accept them. Historically many deontologists have claimed that deontology is fundamentally concerned with constraining actions and not with bringing about goals. They have seen it as an essential feature of deontological theories that the (distinctively deontological) rules they endorse have a structure like that of (1) and (2) rather than a structure like (3) and (4). For example, Ross (1939, p.3) distinguishes what he calls 'teleological ethics' from deontology in the following way:

> On the one hand there is a group of opinions involving the closely connected ideas of duty, of right and wrong, of moral law or laws, of imperatives. On the other hand, there are opinions involving the idea of goods or ends to be aimed at. In the one case the idea of human life is envisaged as obedience to laws; in the other as the progressive satisfaction of desire and attainment of ends.

Similar claims that contrast deontology with an approach to morality that is goal-directed can be found in Nozick (1974, p.29), Darwall (1986, 292), Quinn (1989, p.312), Scanlon (1998, p.85), and Otsuka (2011, p.46). Thus, a deontologist who accepts (3) or (4) must reject this common view about the nature of deontology and accept that the constraints endorsed by her deontological theory are similar to the rules one finds in consequentialist or teleological ethics—they are maximizing rules that direct agents towards certain goals or ends.

I suspect that many deontologists will be reluctant to abandon the traditional conception of deontology and thus will interpret agent-centered constraints as (1) and (2), rather than (3) and (4). However, some deontologists might respond by rejecting the traditional conception of deontology and arguing that deontological constraints can in fact be maximizing rules like (3) and (4). Such a position might be bolstered with a diagnosis of where the traditional conception goes wrong. It might be argued that containing an agent-

centered constraint on promoting the good is actually the essential feature of deontology and that Ross and others classified deontology as non-goal directed and non-maximizing because they mistakenly assumed that agent-centered constraints necessarily have these two properties. Yet (3) and (4) demonstrate that there can be agent-centered constraints on promoting the good that are goal-directed, maximizing rules.

Two defenders of the traditional conception of deontology have considered this possibility and rejected it. Nozick (1974, p.29 fn.) considers the possibility of agent-centered constraints formulated along the lines of (3) and dismisses them as 'gimmicky'. Otsuka (2011, pp.41-50), argues that deontologists ought to reject constraints formulated as (3) or (4) because they fail to capture the core deontological idea that persons are separate and must always be treated as such. Their arguments are interesting, however, I will not discuss them further because I think there is an even more fundamental objection to the deontologist interpreting her agent-centered constraints as maximizing rules. The objection is that such a version of deontology collapses into agent-relative consequentialism and thus is not a kind of deontology after all. Recall that agent-relative consequentialists postulate agent-relative values and then argue that, as per consequentialism, agents must maximize these values. This move allows them to give the deontic verdicts of commonsense morality in cases like PROMISE from within a consequentialist framework. For example, an agent-relative consequentialist might claim that, for each agent, the state of affairs of that agent not breaking her promises is good-relative-to that agent. When this value claim is combined with consequentialism we get rule (3). Alternatively, she might claim that, for each agent, the state of affairs of that agent not breaking her promises at some time T, is good-relative-to that agent at T. When this value claim is combined with consequentialism we get rule (4). Thus, an agent-relative consequentialist can endorse rules (3) and (4). The only difference between her theory and the theory of a deontologist endorsing (3) or (4) is that the agent-relative consequentialist holds that what is being maximized in (3) and (4) has moral value, whereas the deontologist must deny this. For if the deontologist accepted this then she would hold that (3) or (4) requires agents to maximize the good—a straightforward consequentialist stance.[11]

---

[11] To be clear, it is ok for a deontologist to say in certain circumstances that an agent must maximize the good. However, to be a deontologist she must hold that there are some possible circumstances where it is impermissible to maximize the good (in other words, some circumstances where there is an agent-centered constraint on maximizing the good). If the agent-centered constraint one endorses turns out to be just another way of maximizing the good, then one fails to be a deontologist. This point is importantly different from the point made by consequentializers. Consequentializers argue that the deontic verdicts of any moral theory can be produced by

We have seen that any deontologist who endorses (3) or (4) must insist that these rules do not require agents to maximize moral value as otherwise her theory collapses into agent-relative consequentialism. However, there is something implausible about this view. To see this let's consider rule (3). According to (3), each agent must maximize her non promise-breaking. For example, Akira must maximize *Akira not breaking promises*. If a deontologist accepts (3) then she must insist that, although Akira must maximize *Akira not breaking promises*, his not breaking promises is not something of moral value. However, this seems intuitively implausible. It seems that if morality requires an agent to bring about as much of something as she can, then that thing must have some kind of moral worth or value. If it did not, then why would morality require her to bring it about? Indeed, it seems that the best explanation of a maximizing rule is a value explanation that appeals to the value of the thing maximized. For maximization appears to be the fitting response to the discovery that something possesses moral value, and thus we expect to find maximizing rules only when there is something of value that it is fitting to maximize. Hence, once a deontologist accepts a maximizing rule it seems she must also accept that such a rule requires the maximization of value. But then when a deontologist accepts (3) or (4) she is really accepting a version of agent-relative consequentialism.

Finally, it is worth considering a further problem facing the deontologist who accepts (3) or (4). Suppose such a deontologist argues that it is actually plausible for morality to require the maximization of something lacking moral value.[12] If such an argument succeeded would she be off the hook? Not necessarily. Recently, Douglas Portmore has argued that agent-relative consequentialism does not require that some things are good-relative-to particular agents. According to Portmore (2011), all that is needed for agent-relative consequentialism is a *teleological theory of reasons* which says that all moral reasons are reasons to bring about certain states of affairs.[13] Such a theory provides a ranking of states of affairs, which is enough, in Portmore's view, to get a consequentialist

---

consequentialism. They then typically argue that, whatever deontic verdicts you are committed to, you should prefer the consequentialist theory that produces those verdicts over the non-consequentialist theory that produces them because the former has certain theoretical virtues that the latter lacks. By contrast, I argue here that a deontologist who interprets her constraints as (3) or (4) is giving up on deontology and adopting a consequentialist theory; not because her theory can be *represented* as a consequentialist theory, but because her theory simply *is* a consequentialist theory. Whereas, if she adopted (1) or (2) she would not be adopting consequentialism (although her theory may be consequentializable).

[12] For example, perhaps she denies that the thing maximized has value because, following Schroeder (2007), she rejects agent-relative value.

[13] Portmore's argument also requires moral rationalism. However, because it is accepted by most deontologists, I will not discuss it as a premise that needs defence.

theory going. In an attempt to defend deontology against this argument, Paul Hurley (2014) says:

> … why accept [the teleological theory of reasons]? Although we routinely invoke reasons that appeal to facts relevant to what is better or worse for each of us (first-personal reasons) and better or worse overall (third-personal reasons), we also routinely seem to appeal to second-personal reasons, reasons to do and not to do certain things, e.g., reasons to keep promises and not to commit murders. Such second-personal reasons seem misinterpreted by the teleological conception, e.g., as reasons to desire that my promise keeping be maximized, or to desire that my commission of murders be minimized.

Hurley concedes that if a deontologist accepted constraints with the structure of (3) and (4) then she would be susceptible to Portmore's argument. However, Hurley rests his defense of deontology on the claim that agent-centered constraints are not best represented this way, presumably because he thinks they are better represented as having the structure of (1) or (2). Thus, a deontologist who accepts rules like (3) and (4) also appears vulnerable to Portmore's argument that moral reasons are teleological reasons, which ultimately leads to agent-relative consequentialism. If a deontologist wants to reject Portmore's teleological theory of moral reasons, then she will need to show that some of our moral reasons do not have a teleological structure. However, if she accepts that all agent-centered constraints are maximizing rules like (3) and (4), then she no longer has a theory that can support non-teleological moral reasons.

From this discussion we can see that it is problematic for a deontologist to accept (3) or (4) as her agent-centered constraint because doing so appears to be adopting agent-relative consequentialism and abandoning deontology. Two arguments lead to this conclusion. First, it is implausible to hold that morality requires maximization while denying that the thing maximized has moral value, which means that (3) and (4) must be interpreted as consequentialist value-maximization rules. Second, a strong case has been made that what consequentialism must grant and deontology must deny is that all moral reasons are reasons to bring about certain states of affairs. However, accepting that all moral rules are maximizing rules like (3) and (4) concedes this point to the consequentialist. These problems constitute the first horn of the trilemma.

5. The Second Horn: Moral Dilemmas and Moral Advice

Given the problem I raise above for deontologists who take agent-centered constraints to have the structure of (3) or (4), it seems the deontologist should interpret agent-centered constraints as having the structure of either (1) or (2). However, adopting (1) or (2) as an

agent centered-constraint leads to a different kind of problem for deontology. To see this problem let's review (1) and (2). Recall that if a deontologist accepts (1) then she counts SYNCHRONIC and DIACHRONIC as moral dilemmas where, whatever option an agent takes, she does the morally wrong thing. On the other hand, if a deontologist accepts (2) then she regards SYNCHRONIC and DIACHRONIC as situations where the agent-centered constraint gives no deontic verdict—classing the options available as neither right nor wrong. Rules (1) and (2) are different moral rules as they give different deontic verdicts in these cases. However, despite being distinct they both share a common feature. Each rule gives no moral advice to agents in SYNCHRONIC and DIACHRONIC. Of the various options available to an agent in these cases, neither rule recommend any option as the morally best thing to do. I hold that the absence of moral advice in these cases makes (1) and (2) implausible as moral rules. This problem with (1) and (2) is the second horn of my trilemma.

In general, it is philosophically controversial whether morality is uniquely action-guiding. Some hold that morality must be uniquely action-guiding, always advising an agent on what to do whenever there are things of moral significance at stake. Others reject this view and accept that it is plausible and even necessary that morality is sometimes unable to give advice. However, the objection I want to raise against (1) and (2) does not involve the controversial claim that the former view in this debate is the correct one. Rather the objection is that, even if morality sometimes rightly fails to give agents moral advice, it is implausible that it would give no moral advice in SYNCHRONIC and DIACHRONIC.

To see why the lack of advice might be problematic in these cases it is helpful to consider examples where morality might plausibly fail to give advice. Two examples widely discussed in the literature are cases of symmetry and cases of incommensurability.[14] In cases of symmetry an agent must choose between several bad options that are morally on a par. For example, in the novel 'Sophie's Choice' a mother with two infant twins is asked by a concentration camp guard to nominate one of her twins for execution (and threatened that both will be killed if she refuses to make a nomination). In so far as the mother's two options are completely symmetrical (i.e., there are no moral reasons to favour one twin over the other, such as one having a better chance of survival) then it seems plausible that morality might fail to give advice in this case. By contrast, cases of incommensurability are cases where the available options contain competing values that are incommensurable and thus cannot be measured against each other in any meaningful way. For example, consider Sartre's student who, during the Second World War, had to choose between caring for his

---

[14] For example, see Sinnott-Armstrong (1988).

elderly widowed mother and joining the French Résistance to fight the Nazis. In so far as the two values at stake here—caring for a parent in need, and defending one's country against an evil aggressor—are incommensurable, it is plausible to think that morality might fail to give advice.

Most of the arguments for strong moral dilemmas focus on cases of symmetry or incommensurability. However, neither SYNCHRONIC nor DIACHRONIC are cases of symmetry or incommensurability. They are not cases of symmetry because the options available have morally significant differences—namely, the difference of comparable promises being broken to *one* person or *two* people. They are not cases of incommensurability because, insofar as all the promises at stake are comparable, they are commensurable and can be weighed against each other. Finally, unlike the classic examples of symmetry and incommensurability where our intuitions support the claim that morality gives us no advice, SYNCHRONIC and DIACHRONIC are cases where our intuitions go the other way and favour moral advice. For example, in SYNCHRONIC most have the intuition that the morally right thing for Sylvia to do is to minimize her promise-breaking, breaking a promise to one rather than two. These intuitions become even stronger when we consider that real world cases like SYNCHRONIC and DIACHRONIC generally have a third option available—the option of breaking all three promises. Whatever one thinks about the choice between the first and the second option, we can at least all agree that the agent must take one of these options over the third option of breaking all her promises. Yet (1) and (2) not only fail to give advice in the two-option version of the case, they also fail to give advice in the three-option version. Rule (1) classes the agent as doing the, all-things-considered, morally wrong thing whichever of the three options she takes, and thus does not advise her to at least take the first or second option over the third option. Rule (2) does not give any verdict in the three-option case and thus classes the agent as doing neither the morally right nor wrong thing, even when she decides to break all her promises. Any theory that falls short of commonsense morality in such a blatant way does not look very plausible.

6. Supplementary Rules

From the arguments of the previous two sections we appear to have a dilemma. A deontologist must interpret her agent-centered constraint on promise-breaking as either (1), (2), (3), or (4). If she interprets it as (3) or (4) then she faces the first horn that her theory collapses into agent-relative consequentialism. If she interprets it as (1) or (2) then she faces the second horn that her theory is not uniquely action-guiding in cases where

we expect it. However, there is a way for the deontologist to escape this dilemma. She can avoid the second horn by supplementing (1) or (2) with another rule that gives advice in SYNCHRONIC and DIACHRONIC. In this section I will examine this possibility, and argue that it either leads back to the first horn, or leads to a third horn. Thus, what initially appears to be a dilemma is actually a trilemma.

Before looking at what supplementary rules might be appropriate, we need to clarify how supplementing might work. In the case of rule (2) supplementing is straightforward. Rule (2) gives no deontic verdict in SYNCHRONIC and DIACHRONIC, thus any rule giving an appropriate verdict in these cases can stand in as a supplementary rule. In the case of rule (1), supplementing is more complicated. Rule (1) gives the deontic verdicts in SYNCHRONIC and DIACHRONIC that each of the actions the agent might perform are, all-things-considered, morally wrong. By contrast, a supplementary rule that advises an agent on what to do in these cases must hold that one of the available actions is the morally right act. Thus, rule (1) gives a deontic verdict that conflicts with the verdict of the supplementary rule in these cases. To avoid this conflict, and produce a set of rules that is uniquely action-guiding in SYNCHRONIC and DIACHRONIC, one of two routes must be taken. First, it could be stipulated that the supplementary rule has lexical priority over (1). Thus, in SYNCHRONIC and DIACHRONIC the verdict of the supplementary rule overrules the verdict of (1), and agents are given advice (via the supplementary rule) on which option to take. Second, the supplementary rule could be what deontic logicians call a *contrary-to-duty obligation*. A contrary-to-duty obligation is a conditional rule telling an agent what to do in circumstances where she will break her all-things-considered obligation. Applied to rule (1) this means that, in SYNCHRONIC and DIACHRONIC, (1) gives the agent an all-things-considered obligation to not perform any of her available options, while the supplementary rule gives the agent an all-things-considered *conditional* obligation to perform one of the available options. To many such a rule will appear paradoxical, and this has led some to wonder whether the idea of a contrary-to-duty obligation is incoherent. Indeed, there is an unresolved debate in deontic logic about whether contrary-to-duty obligations result in a deontic paradox.[15] I will have more to say on this later, but for now we will leave these obligations on the table as one possible way of supplementing.

Given these clarifications on how one might supplement (1) or (2) with an additional rule, we can now look at candidate supplementary rules. To find candidates we should first ask what advice the supplementary rule should give in cases like SYNCHRONIC and

---

[15] For example, see: Forrester (1984), Sinnott-Armstrong (1985), Goble (1991), and Bonevac (1998).

DIACHRONIC. An obvious answer is that morality should advise us to minimize our promise-breaking (either overall, or in the present-moment) in cases like SYNCHRONIC and DIACHRONIC. This answer is attractive because commonsense morality appears to hold that, all else being equal, doing the minimal amount of harm, or maximal amount of good, is the right thing to do. Furthermore, SYNCHRONIC appears to be a case where all else is equal, as the only thing at stake between the two options is whether Sylvia breaks a promise to one person or two. Other factors that might make a moral difference, such as using a person as a mere means, are not at stake in this case. Given these considerations, (3) and (4) are two obvious candidates for the supplementary rule. Each of these rules require an agent to minimize her promise-breaking (either overall, or in the present-moment) in cases like SYNCHRONIC and DIACHRONIC.

Putting all this together, there are six ways of supplementing to consider:

(i)   Adopt (1), supplement with (3), hold that (3) has lexical priority.

(ii)   Adopt (1), supplement with (3), hold that (3) is a contrary-to-duty obligation.

(iii)   Adopt (2), supplement with (3), (priority and conditionality not needed).

(iv)   Adopt (1), supplement with (4), hold that (4) has lexical priority.

(v)   Adopt (1), supplement with (4), hold that (4) is a contrary-to-duty obligation.

(vi)   Adopt (2), supplement with (4), (priority and conditionality not needed).

If a deontologist adopts any of these six combinations, then she avoids the two horns I have considered. Each combination escapes the first horn because it interprets the constraint such that it is not a maximizing rule, and thus avoids the collapse into agent-relative consequentialism. Each combination escapes the second horn because it contains a supplementary rule that gives advice in cases like SYNCHRONIC and DIACHRONIC. However, despite these advantages, each combination faces a version of the same challenge. It adopts either (3) or (4) as a supplementary rule, which leads to the question: *Why not accept the supplementary rule as the constraint in the first place?* For example, the first option involves accepting (1) and supplementing it with (3). However, why should we adopt this combination of (1) and (3) when we can just adopt (3) instead? There is a serious objection to each of the six combinations behind this rhetorical question. To show this, I will consider each combination, comparing it to the alternative of accepting just the supplementary rule.

Let's start with the first combination, comparing it to accepting (3) outright. The crucial thing to notice here is that combining (1) with (the lexically prior) (3) gives the same deontic verdicts across all possible cases as (3) alone gives. This is because (3), being

lexically prior, always overrules (1) when (1) and (3) give different verdicts.[16] It follows that we cannot adjudicate between (3) and the first combination by appeal to deontic verdicts, as each gives the same deontic verdicts across all cases. So how can we decide between these options? One weak reason for preferring (3) to the combination of (1) and (3) is its simplicity. These options are equally plausible when it comes to the deontic verdicts they give, however, the former option is simpler, using one rule where the latter uses two. All else being equal, a simpler theory is preferable to a more complex one. However, there is a much stronger reason for preferring (3) to the combination of (1) and (3). The problem with the combination is not just that it is more complex than (3), but that it takes rule (3) and adds another rule to it without that rule doing any new work in the moral theory. Rule (1) would be redundant when added to (3) because there are no new deontic verdicts or moral explanations it provides that are not already offered by (3). To see why this redundancy is so problematic, let's consider an analogy. Suppose a sufficientarian principle gives the wrong deontic verdict in a certain case where a rival prioritarian principle gives the correct verdict. One might attempt to preserve sufficientarianism by supplementing the sufficientarian principle with the prioritarian principle, giving the latter principle lexical priority in the relevant case. Such a theory has the virtue of giving the correct deontic verdict. However, if it turns out that, in all other cases, the prioritarian principle gives the same verdicts as the sufficientarian principle, then sufficientarianism is in trouble. For then, what the cases really show is that the prioritarianism principle is superior because it alone is able to capture the correct deontic verdicts. Adding the sufficientarian principle to the (already adequate) prioritarian principle is redundant and is not a plausible way to defend sufficientarianism.

Now let's consider the third combination, comparing it to accepting (3) outright. This combination adopts (2) and supplements it with (3). However, just like the previous combination, combining (2) with (3) gives the same deontic verdicts across all cases as (3) alone gives. This is because (2) and (3) give the same verdicts in all cases except for those cases like SYNCHRONIC and DIACHRONIC where an agent lacks the ability to keep all her promises. In these cases, rule (2) gives no verdict (because its ability-clause is not satisfied) whereas (3) classes the action that minimizes total promise-breaking as morally right. Thus, the verdict of (3) takes precedence in these cases. This result gives us the same strong reasons we saw above to adopt (3) as opposed to (2) supplemented with (3).

---

[16] To be precise, a second premise is also needed to prove identical deontic verdicts. There must be no cases where (1) gives a deontic verdict and (3) does not. However, there are indeed no cases of this sort.

Adopting (3) is simpler, and combining (2) with (3) is redundant because (3) alone already gives all the deontic verdicts produced by their combination.

Next, let's consider the second combination, which adopts (1) and supplements it with (3), formulated as a contrary-to-duty obligation. When we compare this combination with adopting (3) alone we see that they both give the same rightness verdicts in all possible cases but different wrongness verdicts in certain cases. For example, suppose in SYNCHRONIC Sylvia keeps her promises to Asha and Beth but breaks her promise to César. According to the combination, Sylvia acts wrongly at the moment she breaks the promise to César because of rule (1). However, given that she cannot keep all her promises, the contrary-to-duty obligation applies and she acts rightly by minimizing her promise-breaking. According to (3), Sylvia acts rightly by minimizing her promise-breaking (and does not act wrongly at the moment she breaks her promise). Thus, although the combination and rule (3) give the same rightness verdict in this circumstance, they give different wrongness verdicts. It follows that this combination can resist the redundancy argument that favoured (3) over the other combinations, because (1) produces a verdict that cannot be produced by (3) alone, and thus is not redundant. Nonetheless, I think there are two good reasons to favour (3) over the combination of (1) supplemented with (3) as a contrary-to-duty-obligation. First, favouring (3) avoids the deontic paradoxes that arise from accepting a contrary-to-duty obligation. Insofar as these paradoxes are troubling and resistant to resolution, avoiding them is an important desideratum. Second, (3) appears to give a better account of the timing of the agent's wrongdoing in SYNCHRONIC. Recall that according to (3), Sylvia acts rightly at the moment her promises are due by minimizing her promise-breaking. By contrast, according to the combination, Sylvia at that moment acts both wrongly and conditionally rightly. Perhaps some will be attracted to the verdict of the combination because they want the agent who breaks a promise to be classed as doing the morally wrong thing, even though she is taking the best option given her circumstances. However, (3) also offers an explanation of the agent's wrongdoing, an explanation that I think is much more plausible. Typically, when an agent finds herself in a situation like SYNCHRONIC there has been some point in the past when she has failed to minimize her overall promise-breaking. For example, perhaps she irresponsibly made a set of incompatible promises. In this case, (3) classes her as doing the wrong thing at the moment she makes the incompatible promises because she fails to minimize her promise-breaking at that moment. Alternatively, perhaps the promises are jointly satisfiable when made, but later when the agent must take steps to ensure she will keep all her promises (such as making a start on the tasks she promised to complete), she instead procrastinates. In this scenario, (3) classes her as doing the wrong thing at the

moment she procrastinates because she fails to minimize her promise-breaking at that moment. In classing the agent as doing the wrong thing at these moments, rather than at the later moment when she takes the best option available and minimizes her promise-breaking, (3) does a better job capturing our commonsense intuitions about the timing of the agent's wrongdoing. Therefore, (3) is a superior constraint to the combination of (1) and (3) as a contrary-to-duty obligation.

If the arguments given so far succeed then the first three combinations must be rejected because accepting (3) outright is more plausible than accepting each combination. The remaining three combinations mirror the first three, with the only difference being that (4) rather than (3) serves as the supplementary rule. Thus, the arguments I have employed against combinations (i), (ii), and (iii), work against combinations (iv), (v), and (vi). In each case, they show that accepting (4) outright is more plausible than accepting the combination. Therefore, we can conclude from this section that supplementing (1) or (2) with either (3) or (4) is not a viable solution the dilemma I posed because attempting this kind of supplementing leads us to the conclusion that we should accept (3) or (4) outright, and this takes us back to the first horn.

7. The Third Horn: The Numbers Don't Count

In the previous section I argued that accepting (3) or (4) as a supplementary rule to (1) or (2) is not viable as it leads back to the first horn. However, this does not prove that the supplementary rule strategy cannot be employed to escape the second horn. For, although (3) and (4) are the most obvious supplementary rules, they are not the only possible ones. If there are other supplementary rules a deontologist might add to (1) or (2) that give advice on what to do in SYNCHRONIC and DIACHRONIC, then she might still escape the first two horns. I know of only two other (minimally plausible) supplementary rules meeting this description. Both of these rules come out of a strong anti-aggregation position some deontologists take known as the 'numbers don't count' view.[17] According to this view, when we need to make a trade-off between different people involving comparable goods, and all is equal other than the numbers, it is wrong to decide what to do by favouring the greater number. For example, suppose you find yourself able to rescue either one person from a flood or ten other people, and all else is equal (e.g. you have no personal connections to any victims, no one is more deserving, etc.) Commonsense morality says that we should favour the ten in this case as, all else being equal, ten saved and one drowned is morally preferable to the alternative. However, 'numbers don't count'

---

[17] See Taurek (1977), for the classic presentation of this view.

deontologists object to this kind of aggregation arguing that by adding up the goods of different people's lives we are treating them as mere receptacles of value, and thereby disrespecting them. Instead of aggregating, these deontologists suggest one of two alternative decision procedures. First, you might give each person an equal chance of having her interests satisfied. In the rescue case you could do this by flipping a fair coin to decide whether to save the ten or the one, giving both the one and each of the ten a one in two chance of being saved. Second, you might give each person a proportional chance of having her interests met. In the rescue case you could do this by holding a weighted lottery, giving the one a one in eleven chance of being saved and the ten a ten in eleven chance of being saved.[18] Each of these decision procedures can be adapted to promise-breaking examples, producing rules that advise us on what to do in cases like SYNCHRONIC and DIACHRONIC:

> (5) Each agent must give all her promisees an *equal chance* of having their promises honoured whenever it is not possible for her to keep all her promises.

> (6) Each agent must give all her promisees a *proportional chance* of having their promises honoured whenever it is not possible for her to keep all her promises.

Either of these rules could be used to supplement (1) or (2). Importantly, a deontologist who adopts (1) or (2) and supplements it with (5) or (6) avoids the two horns discussed above. She avoids the first horn because she ends up with a constraint that is not a maximizing rule. She avoids the second horn because the supplementary rule she endorses (either (5) or (6)) gives moral advice in SYNCHRONIC and DIACHRONIC.[19] However, escaping the first two horns comes at a cost, as there is a problem with accepting (5) or (6). The problem with these rules, and with the 'numbers don't count' view in general, is that they conflict with deeply ingrained intuitions of commonsense morality—intuitions telling us that, when all else is equal, we should aggregate and favour the greater number. Because of these commonsense intuitions many deontologists reject the 'numbers don't

---

[18] See Taurek (1977) for the first decision procedure and Timmermann (2004) for the second.

[19] Notice that (5) and (6), as I have stated them, cannot stand in as general constraints on promise-breaking. In this respect they appear different from (3) and (4), which were plausible as general constraints on promise-breaking and as supplementary rules to (1) and (2). However, this appearance is deceptive as there is a way of stating (5) as a general constraint (although I know of no way of doing the same for (6)):

> (5′) Each agent must give all her promisees the greatest possible chance of having their promises honoured, consistent with every promisee having an equal chance of having her promises honoured.

It follows that we could use (5′) as an outright constraint instead of using (5) as a supplementary rule. However, (5′) as a constraint faces the same problem (the third horn) as (5) and so does not need to be discussed separately.

count' view.[20] Thus, I see this problem as a third horn, and the arguments I offer here as presenting a trilemma to deontologists. The deontologist must clarify which of (1)-(4) she takes as her agent-centered constraint. If she adopts (3) or (4) she faces the first horn of her theory collapsing into agent-relative consequentialism because all the rules it endorses are maximizing rules. If she adopts (1) or (2) she faces the second horn of her theory not giving moral advice when commonsense morality requires it. If she attempts to escape this horn by adopting (3) or (4) as a supplementary rule then several strong arguments lead her back to accepting (3) or (4) outright, and thus back to the first horn. Alternatively, if she attempts to escape the second horn by adopting (5) or (6) as a supplementary rule then she faces the third horn of accepting a strong anti-aggregation position that conflicts with commonsense morality. There are no other (minimally plausible) supplementary rules to (1) and (2) that give moral advice and thus no further options available.

Of course, those who are already committed to 'numbers don't count' deontology, will not see a problem with accepting rules like (5) and (6). They may accept that such rules conflict with commonsense morality, but they will likely argue that commonsense morality is confused and inconsistent, and that adopting the anti-aggregation position is the best way to resolve these inconsistencies. Thus, they will not see a trilemma here at all. However, my argument does have something to offer these deontologists. It gives them a novel argument to bring other deontologists over to their view. This argument is an eliminative disjunction telling deontologists they must either accept the first horn, the second horn, the numbers don't count view, or abandon deontology. The 'numbers don't count' deontologist can then argue that her view is more plausible than the other options, and thus should be endorsed by all deontologists. Whether other deontologists would accept such an argument, or instead see the third option as a horn to be avoided at all costs, remains to be seen.

8. Conclusion

I have presented a trilemma for deontology which shows that one of deontology's perceived strengths, its ability to accommodate agent-centered constraints, is also one of its weaknesses. This trilemma is especially pressing because most deontologists are already inclined to reject all three horns. Most deontologists reject the 'numbers don't count' view and yet hold that deontological constraints are not maximizing rules, and that morality is uniquely action-guiding in cases like SYNCHRONIC and DIACHRONIC. My trilemma

---

[20] For deontologists rejecting the view see: Kavka (1979), Kamm (1993), Scanlon (1998), Kumar (2001), Hirose (2001), Raz (2003), Hsieh, Strudler, and Wasserman (2006), Liao (2008), and Dougherty (2013).

shows that this combination of views is incompatible. Thus, the deontologist must either embrace one of the horns of the trilemma or reject deontology altogether, perhaps adopting agent-relative consequentialism in its place. I have given reasons why each of the horns appears problematic. However, I do not claim to have shown conclusively that each horn must be rejected. Thus, I do not claim to have proven that deontology is false. The main achievement of my argument is that it forces the deontologist to clarify her position in a way that she may not have realized was necessary. Perhaps she can respond to this challenge by finding a way to square deontological ethics with maximizing rules. Or perhaps she can find a justification for embracing the counterintuitive view that the 'number don't count', or that morality does not give advice where we expect it. However, if she cannot mount a successful defence of such views then she must abandon deontology and embrace a new moral outlook.

References

Bonevac, D. (1998). Conditional Obligation. *Nous*, 32: 37-53.

Broome, J. (1991). *Weighing Goods*. New York: Blackwell.

Darwall, S. (1986). Agent-Centered Restrictions from the Inside Out. *Philosophical Studies*, 50: 291-319.

Dreier, J. (1993). Structures of Normative Theories. *The Monist*, 76: 22-40.

Dougherty, T. (2013a). Agent-Neutral Deontology. *Philosophical Studies*, 163: 527-537.

Dougherty, T. (2013b) Rational Numbers. *The Philosophical Quarterly*, 63: 413-427.

Forrester, J. (1984). Gentle Murder, or the Adverbial Samaritan. *Journal of Philosophy*, 81: 193-197.

Goble, L. (1991). The Paradox Deepens. *Philosophical Studies*, 64: 217-227.

Hammerton, M. (2020) Deontic Constraints are Maximizing Rules. *Journal of Value Inquiry*, 54: 571-588.

Hirose, I. (2001). Saving the Greater Number without Combining Claims. *Analysis*, 61: 341-342.

Hsieh, N.-H., Strudler, A., Wasserman, D. (2006). The Numbers Problem. *Philosophy & Public Affairs*, 34: 352-372.

Hurley, P. (2013). Consequentializing and Deontologizing. *Oxford Studies in Normative Ethics*, 3: 123-153.

Hurley, P. (2014). Comments on Douglas Portmore's Commonsense Consequentialism. *Philosophy and Phenomenological Research*, 88: 225-232.

Kamm, F. M. (1992). Review: Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status. *Philosophy and Public Affairs*, 21: 354-38

Kamm, F. M. (1993). *Morality, Mortality: Volume I*. Oxford University Press.

Kagan, S. (1989). *The Limits of Morality*. Oxford: Clarendon Press.

Kagan, S. (1998). *Normative Ethics*. Boulder: Westview.

Kavka, G. (1979). The Numbers Should Count. *Philosophical Studies*, 36: 285-94.

Kumar, R. (2001). Contractualism on Saving the Many. *Analysis*, 61: 165-70.

Louise, J. (2004). Relativity of Value and the Consequentialist Umbrella. *Philosophical Quarterly*, 54: 518-36

McNaughton, D. & Rawling, P. (1991). Agent-Relativity and the Doing-Happening Distinction. *Philosophical Studies*, 63: 167-185.

Nagel, T. (1986). *The View from Nowhere*. New York: Oxford University Press.

Nozick, R. (1974). *Anarchy, State and Utopia*. New York: Basic Books.

Otsuka, M. (2011). Are Deontological Constraints Irrational? In Bader and Meadowcroft (Eds.) C*ambridge Companion to Nozick's Anarchy, State, and Utopia.* Cambridge: Cambridge University Press.

Parfit, D. (1984) *Reasons and Persons*. Oxford: Clarendon Press.

Pettit, P. (1997). The Consequentialist Perspective. In Baron, Pettit, and Slote (Eds.) *Three Methods of Ethics*. Oxford: Blackwell.

Portmore, D. (2001). Can an Act-Consequentialist Theory be Agent-Relative? *American Philosophical Quarterly*, 38: 363-77.

Portmore, D. (2011). *Commonsense Consequentialism*. Oxford University Press.

Quinn, W. (1989). Actions, Intentions, and Consequences. *Philosophical Review*, 98: 287-312.

Raz, J. (2003). Numbers with and without Contractualism. *Ratio*, 16: 346-67.

Ridge, M. (2001). Agent-Neutral Consequentialism from the Inside-out. *Utilitas*, 13: 236-254.

Ross, W.D. (1930). *The Right and the Good*. Oxford: Oxford University Press.

Ross, W.D. (1939). *The Foundations of Ethics*. Oxford: Oxford University Press.

Scanlon, T.M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Scheffler, S. (1982). *The Rejection of Consequentialism*, Oxford: Clarendon Press.

Schroeder, M. (2007) Teleology, Agent-Relative Value, and 'Good'. *Ethics*, 116: 265-295.

Sen, A. (1982). Rights and Agency. *Philosophy and Public Affairs*, 11: 3-39.

Sinnott-Armstrong, W. (1985). A Solution to Forrester's Paradox of Gentle Murder. *Journal of Philosophy*, 82: 162-168.

Sinnott-Armstrong, W. (1988). *Moral Dilemmas*. Oxford: Basil Blackwell.

Smith, M. (2003). Neutral and Relative Value after Moore. *Ethics*, 113: 576-98.

Smith, M. (2009). Two Types of Consequentialism. *Philosophical Issues*, 19: 257-272.

Styron, W. (1980). *Sophie's Choice*. New York: Bantam Books.

Taurek, J. (1977). Should the Numbers Count? *Philosophy & Public Affairs*, 6: 22-40.

Timmermann, J. (2004). The Individualist Lottery. *Analysis*, 64: 106-112.

5

A Defence of Agent-Relative Value[21]

1. Introduction

It is often assumed that all moral value is agent-neutral, meaning that it is impersonal value, equally accessible from all agential perspectives. However, an alternative view holds that at least some moral values are agent-relative, meaning that these values are relativised to particular agential perspectives. The possibility of agent-relative value is particularly interesting because it appears to allow for a novel kind of consequentialism—often called 'agent-relative consequentialism'—that combines attractive features from both traditional deontology and traditional consequentialism together into one moral theory. However, skeptics have argued that there are no good reasons for thinking that moral value can be agent-relative, and hence that agent-relative consequentialism is unjustified.[22] In this chapter I offer a novel defence of agent-relative value. My defence appeals to several commitments of commonsense morality and argues that these commitments ultimately lead to the conclusion that some moral value is agent-relative. In this way, my argument defends agent-relative consequentialism against the skeptical challenge. In addition it offers a general defence of agent-relative values, giving us all a prima facie reason to accept their existence.

2. Agent-Relative Value Explained

What does it mean to say that moral values can be agent-relative? To best answer this question we should start by explaining what agent-neutral moral value is. When we describe a state of affairs as having moral value it is common to suppose that this value is impersonal, meaning that it is present for any agent no matter what perspective she occupies. For example, most will agree that it is a morally good thing that women in Saudi

[22] See especially Schroeder (2006/2007).

Arabia are now permitted to vote. When we endorse such a claim it is natural to suppose that the value we are talking about is agent-neutral value. This value is present at all times and all places (even if agents at those times and places fail to see it) and it is not value that is relativised to some particular perspective or other. We might use Sidgwick's phrase and say that this means it is valuable 'from the point of view of the universe'. Or we might say that it is valuable from no point of view, just valuable simpliciter. When 'good' is applied to states of affairs in this agent-neutral way it is common to suppose that it is a monadic predicate that attaches to a state of affairs and not any other kind of grammatical category such as a relation with multiple relata, or a predicate modifier.

Although it is common to interpret claims about moral value as claims about agent-neutral value there is a rival interpretation available. We might instead interpret such claims as referring to agent-relative values. We might think that when states of affairs are morally good they are not good simpliciter but only good relative to an agent or agential perspective. On this view, when 'good' is applied to a state of affairs it is not a monadic predicate but rather a relation that connects the state of affairs with an agent. Thus, strictly speaking, it would be incorrect to say: "It is a morally good thing that women in Saudi Arabia are now permitted to vote". For this sentence to be complete we must add the agent place, saying something like: "It is a morally good thing, relative to agent *A*, that women in Saudi Arabia are now permitted to vote".

When presented in this way the thesis that moral value is agent-relative may appear absurd. For most of the value claims we accept appear to be agent-neutral; they appear to apply generally and not just relative to some agents. For example, it may be pointed out that it would be strange for someone claiming that *women in Saudi Arabia being given the vote* has objective moral value to then qualify this claim by saying 'but it is only good relative to certain agents'. However, two things are worth keeping in mind here. First, in this chapter we are interested in the thesis that *some* moral value is agent-relative. It may be true that most moral values actually have an agent-neutral structure, but in so far as some moral values are agent-relative the thesis is vindicated. Second, for anyone wanting to claim that *all* moral value is agent-relative there is a good response available to the objection we are considering. Such a theorist can claim that examples which appear to be cases of agent-neutral value are actually just cases where a state of affairs has relative value for all agents. According to this view, some states of affairs, such as Saudi women being granted the right to vote, are valuable relative to all agents and hence it seems natural to describe them as having agent-neutral value. However, strictly speaking, the value being described is still a two-place relation with an agential place and hence agent-relative. It is

just that filling out the agential place becomes uninformative and hence unnecessary because we already know the claim is true regardless of which agent we make it relative to. (According to an alternative way of describing this view there *are* agent-neutral values because the term 'agent-neutral value' is best redefined as just the special case of agent-relative value where the relative value holds for any arbitrary agent.[23]) Of course, these replies do not show that we have any reason to think that there are agent-relative moral values. They merely show that we cannot quickly dismiss as absurd the thesis that there are such values.

At this point it is helpful to clarify what agent-relative moral value is not in order to avoid some common lines of confusion. First, the claim that some states of affairs have agent-relative moral value should not be confused with the claim that some states of affairs are *good for* particular agents.[24] Both claims refer to a two-place relation that takes agents and states of affairs as relata. Furthermore, the relation picked out by each claim makes use of the term 'good'. However, they are nonetheless different claims. Agent-relative values are a kind of moral value. On the other hand, 'good for' claims are claims about prudential value. This is why one can say: 'Margaret Thatcher's Poll tax was good for the rich and yet was not a morally good tax policy' without contradicting oneself. Thus, it does not follow from the claim that a state of affairs *S* is *good-relative-to* a particular agent that *S* is *good for* that agent. A state of affairs can be *morally good-relative-to an agent* without being *prudentially good for that agent*. Second, the claim that some values are agent-relative should not be confused with the claim that some values are subjective, or mind-dependent. For example, suppose that the state of affairs that *Alfred does not kill innocent people* is good relative to Alfred. It does not follow from this that this state of affairs is subjectively valuable, i.e. valuable because Alfred judges that it is valuable. To the contrary, claims about agent-relative value, just like claims about agent-neutral value, are neutral when it comes to debates between moral objectivists and subjectivists. Thus, in the example above the state of affairs can be valuable relative to Alfred even if Alfred does not recognise this value and finds that it conflicts with his own interests.

The discussion so far has explained what agent-relative value is. Later I will defend the claim that agent-relative values exist. However, before I do so it is worth mentioning one special reason why the question of whether there are agent-relative values has philosophical importance. Over the last 25 years several philosophers have argued for a

---

[23] See Smith (2003), pp. 587-590.
[24] This point is made by Regan (1983), p. 96, Broome (1991a) p. 8, and Schroeder (2007), p. 272.

theory known as *agent-relative consequentialism*[25] or *agent-relative teleology*.[26] According to act consequentialism, the right action to perform is the action that will result in the morally best consequences. Historically, act consequentialists (along with everyone else) have taken it for granted that moral value is agent-neutral and thus have assessed which consequences are best in terms of such neutral values. However, when neutral-value is assumed, act consequentialism faces the objection that it permits things forbidden by commonsense morality. For example, consider the following case:

> MAFIA: The Mafia make a credible threat that they will kill three innocent people unless you kill one innocent person.

What should you do in such a case? According to commonsense morality it is wrong to kill an innocent person even if you are in MAFIA and doing so is the only way to prevent several such killings. Commonsense morality gives this answer because it is committed to an agent-centered constraint on killing innocent people. This constraint requires each agent to have a special concern with *her* own killing over the killings that others may perform. However, act consequentialism (when combined with the assumption that all value is agent-neutral) is unable to secure this commonsense moral verdict. This is because there is no quantity that (i) might plausibly be said to have moral value, and (ii) will, whenever it is maximized, prohibit agents from killing in cases like MAFIA.[27] Thus, act consequentialism appears to be incompatible with agent-centered constraints, and so permits acts that commonsense morality prohibits. This problem has led many to reject act consequentialism.

---

[25] There is some controversy over whether this theory can properly be described as 'consequentialism' given that it is agent-relative. However, as this is merely a terminological issue on which nothing substantial hangs, by stipulation I will use the term 'consequentialism' to denote any theory that explains the deontic status of acts (or any other evualand, such as rules, motives, or institutions) solely in terms of the moral value of the consequences it brings about. Anyone who prefers to further restrict this definition by insisting that the moral value used in consequentialism must be *neutral value* can translate talk in this chapter of 'agent-relative consequentialism' as actually referring to 'agent-relative teleology'—a kind of maximizing, teleological theory that is not a version of consequentialism.

[26] For versions of this view see: Sen (1982/1983), Broome (1991b), Dreier (1993), Portmore (2001/2005/2011), Smith (2003/2009), and Louise (2004).

[27] One might attempt to counter this claim by pointing to the act-type of *killing to prevent more killings by others*. Let's call acts of this type 'preventative killing'. Perhaps such killings are morally bad and this badness is agent-neutral. If this is assumed then the act consequentialist can then say that all agents are required to minimize preventative killings. Thus, in MAFIA an agent must not kill the innocent person because this would be failing to minimize prevenative killings that are morally bad. However, although this would give the commonsense moral verdict in MAFIA, it would give the wrong verdict in other MAFIA-like cases. For example, it would require an agent to perform one preventative killing herself if doing so is the only way to prevent two other agents from performing preventative killings. Yet commonsense morality clearly prohibits such an act for the same reasons it prohibits killing in MAFIA.

Agent-relative consequentialism can be seen as a response to this problem. Agent-relative consequentialists claim that by revising the common assumption that moral value is always agent-neutral, and instead postulating certain agent-relative values, they are able to construct a version of act consequentialism that is compatible with agent-centered constraints and commonsense morality. For example, to obtain the commonsense moral verdict in MAFIA, an agent-relative consequentialist would endorse the following claim about value:

> (1) For each agent, the state of affairs of that agent not killing is good-relative-to that agent.

It follows from (1) that the state of affairs of *Alfred not killing* is good relative to Alfred, and the state of affairs of *Marisa not killing* is good relative to Marisa, and so on for each agent. And it follows from act consequentialism that each agent must perform the available action that will bring about the maximal amount of moral good. Thus, if *Alfred not killing* is morally good relative to Alfred (and there are no other goods that outweigh this good in a killing case) then Alfred is required to maximize his own non-killing, as from Alfred's position this would be maximizing the moral good. On the other hand, if *Marisa not killing* is morally good relative to Marisa then Marisa is required to maximize her own non-killing, as from Marisa's position this would be maximizing the moral good. In this way, agent-relative consequentialism is able to prohibit each agent from killing innocent people, even if they are in cases like MAFIA where killing is the only way to prevent more killings by others.

From the above discussion we can see that agent-relative consequentialism is able to produce verdicts of commonsense morality that agent-neutral consequentialism cannot. Many agent-relative consequentialists claim that this makes the theory very attractive. For they claim that consequentialism has certain theoretical attractions that have made it a popular moral theory despite the fact that it appears to conflict with commonsense morality.[28] By constructing a version of consequentialism that combines these theoretical advantages with the commonsense verdicts of non-consequentialist theories, agent-relative consequentialists claim that their theory has the best of both worlds. It combines the most attractive features of non-consequentialism and agent-neutral consequentialism while avoiding each theory's most serious problems.[29] James Dreier goes even further

---

[28] For example, they have claimed that consequentialist theories are better than non-consequentialist theories because they avoid the paradox of deontology (see Dreier 1993, and Portmore 2011), and they can adequately accommodate epistemic uncertainty (see Jackson and Smith 2011).

[29] See: Smith (2003), Louise (2004), and Portmore (2011).

arguing that agent-relative consequentialism is so plausible that the most charitable interpretation of both deontologists and agent-neutral consequentialists is that they are all agent-relative consequentialists.[30] However, these arguments in favour of agent-relative consequentialism only get off the ground if there are agent-relative values. For if all moral values are agent-neutral then a central presupposition of agent-relative consequentialism is false and thus the theory fails. Thus, the possibility of agent-relative consequentialism is one important reason to care about whether moral value can be agent-relative.

3. The Skeptical Challenge

We saw above that if there are agent-relative values then agent-relative consequentialism becomes viable. This is one reason why we might be interested in the question of whether there are agent-relative values. However, ever since agent-relative consequentialism was first proposed in the early 1980's, critics have responded with skepticism to its key claim that there is a *good-relative-to* relation.[31] These skeptics point out that, unlike many of the concepts that moral philosophers discuss, agent-relative value appears to have no basis in ordinary moral thought and talk. For commonsense morality appears to treat the moral good as a monadic predicate and does not relativise it to agents. On this basis the skeptics question whether we have any good grounds for thinking that there are agent-relative values. Some take this further, arguing that our folk concept of the 'moral good' provides no room for a *good-relative-to* relation. According to these skeptics the idea of agent-relative value is not just lacking in justification, but is actually conceptually incoherent. For example, Mark Schroeder says the following:

> I don't understand what "good-relative-to" talk is all about, I don't understand how it could be appealing to think that you shouldn't do something that will be worse-relative-to you. I don't even understand what that means! Until the [agent-relative consequentialists] give me some reason to think that the good-relative-to relation is somehow very much like the good property and the good for relation, I don't see why I should remotely find such an idea deeply compelling.[32]

It is helpful here if we examine Schroeder's version of the criticism in a little more detail as his version is the most developed and most influential in the literature. Schroeder (2006/2007) argues that we have no pre-theoretical grasp of the *good-relative-to* relation. In particular, he argues that attempts to explain this relation in terms of our pre-theoretical grasp of the *good-for* relation and the *good-from the point of view of* relation

---

[30] See: Dreier (1993)

[31] Such skepticism can be found in Regan (1983), pp.96-99, Broome (1991a), p.8, Broome (2004), p. 74, Schroeder (2006/2007), and Hurley (2013), p. 135.

[32] Schroeder (2007), p. 291. Broome (1991a), p.8 expresses a similar view.

fail. Concepts for which we have no pre-theoretical grasp can still be accepted if they can be justified as a theoretical posit. For example, we have no pre-theoretical grasp of the concept of an 'electron'. Yet our acceptance of electrons is justified because they play an essential role in atomic theory and we have good independent reasons for accepting atomic theory. However, Schroeder argues that, unlike electrons, the *good-relative-to* relation cannot be justified as a theoretical posit. To do this he considers several attempts to justify agent-relative consequentialism and argues that each of these justifications fails to offer independent reasons for accepting the theory (i.e. they fail to offer reasons that don't already take for granted the *good-relative-to* relation). Schroeder's argument supports a skeptical conclusion in the pyrrhonic sense of 'skeptical' rather than the error-theoretic sense. For undermining the justifications that have been offered for a theory does not prove that no such justification is possible, it merely establishes doubts about the prospects of such justification. Thus we can draw from Schroeder's arguments (and the arguments of others) a skeptical challenge for anyone inclined to accept agent-relative value. Either show how we have a pre-theoretical grasp of agent-relative value, or justify it as a theoretical posit supported by a theory that has independent justification.

Various responses to this skeptical challenge are available although it is still a controversial question whether any of these responses succeed.[33] In the remainder of this chapter I will present a novel defence of agent-relative value that can function as an answer to the skeptical challenge. My defence will show where the arguments of Schroeder and other skeptics go wrong. In doing so, it will defend agent-relative consequentialism against this particular objection. However, it will do more than this. It will also provide a general defence of agent-relative value that gives anyone curious about the status of such values a prima facie reason to accept their existence.

4. Commonsense Morality and Agent-Centered Constraints

As we saw above in §2 commonsense morality is committed to agent-centered constraints. For example, commonsense morality tells us not to kill an innocent person even if this is the only way to prevent more killings of innocents by others. Thus, commonsense morality is committed to the following rule:

> DK: Each agent must not kill innocent people

---

[33] Different kinds of responses can be found in Sen (1983), Smith (2003), Portmore (2007), Smith (2009), Suikkanen (2009), Dreier (2011), and Cullity (2015).

DK is an *agent-centered constraint*, or an *agent-relative rule*, because it gives each agent a different ultimate aim. For example, it gives Alfred the aim that *Alfred does not kill innocent people*. On the other hand, it gives Sita the aim that *Sita does not kill innocent people*.[34] In doing so it requires each agent to give special priority to avoiding her own killings of innocents over and above the killings that others might perform. In this respect, DK can be contrasted with an agent-neutral rule such as the following:

> MK: Each agent must minimize the killing of innocent people

MK is agent-neutral because it gives each agent the same ultimate aim—the aim that the killing of innocent people is minimized. Because of this feature MK conflicts with commonsense morality in cases like MAFIA where killing an innocent person is the only way to prevent more killings overall. In such cases MK requires an agent to kill whereas DK (and commonsense morality) require agents to refrain from killing.

In addition to agent-relativity, there is another kind of relativity that philosophers have discussed known as moment-relativity.[35] To see this category of relativity lets consider a new case that differs from MAFIA:

> RASKOLNIKOV: Raskolnikov's mind is gradually being overcome by homicidal thoughts. He judges correctly that: (1) In the next few days he will give in to these urges and commit multiple homicides, (2) the only option available to him that will prevent this killing spree is to commit a single homicide now (perhaps this will lead to his immediate imprisonment, or satiate his homicidal urges, eradicating them from his psychology).[36]

In MAFIA the agent has to decide whether to kill one innocent person, and thereby prevent the Mafia from killing three, or whether instead to refrain from killing, in which case the Mafia will kill three. In RASKOLNIKOV the agent has to decide whether to kill one innocent person now, and thereby prevent his future self from killing three, or whether instead to refrain from killing now, in which case his future self will kill three.[37]

---

[34] I am using here the account of Parfit (1984) to explain the *agent-relativity* of moral rules. There are other available accounts such as that of McNaughton and Rawling (1991) that, in some special cases, give different verdicts to Parfit's account. However, my discussion of agent-relative rules in this chapter is only concerned with the common cases that all theories agree on and thus nothing important hangs on the account used. I have opted to use Parfit's account here because I find it more intuitively graspable than alternative accounts.

[35] See Parfit (1984), chapter 9, Broome (1991a), Kamm (1992), Ridge (2001), and Lousie (2004) for discussion of moment-relativity (some of these authors use the terms 'temporal-relativity' and 'synchronic relativity' rather than 'moment-relativity').

[36] Cases similar to RASKOLNIKOV are discussed in Broome (1991a), Kamm (1992/1996), and Louise (2004).

[37] Some might worry whether the case as described is really possible. They might claim that either: (i) Raskolnikov has a third option available in this case of neither killing now nor killing later which is obviously the option he should take, or (ii) he does not have this option because in the scenario

RASKOLNIKOV raises an interesting question in relation to the agent-centered constraint on killing. According to DK, Raskolnikov must not kill innocent people. However, in the case above where he has to choose between killing one now and killing several later what should he do? There appear to be three possible answers available to the theorist who endorses DK. The first answer says that an agent in this kind of case should give equal consideration to all of his potential killings, present or future, and prevent as many of them as possible. This answer requires Raskolnikov to kill the one now to prevent his later killing of several. In doing so it embraces moment-neutrality because it does not give priority to any particular moment in time such as the present. The second answer says that an agent in this kind of case must give special consideration to his present killings and attempt at each moment to ensure that he does not kill at that moment. This answer requires Raskolnikov to refrain from killing one now even though this will result in him killing several later. In doing so it embraces *moment-relativity* because it gives moral priority to the present moment over all other moments. Finally, a third answer says any agent in this kind of case is in a moral dilemma where, no matter what course of action he takes, he will perform an action that is all-things-considered morally wrong. According to this answer Raskolnikov is morally required to not kill one now and morally required to not kill several later, and there is no further advice that morality can give him even though it appears that he will fail to fulfil both these requirements. Such an answer embraces neither moment-neutrality nor moment-relativity but instead counts the case as a moral dilemma. Although these answers are all concerned with DK and the RASKOLNIKOV case, they actually correspond to three general possibilities that apply to all agent-centered constraints. For given any such constraint we can construct a RASKOLNIKOV-like case and ask what the agent ought to do. Whenever we do so we will find that there are three possible answers:

(i)   Give the constraint a moment-neutral interpretation

(ii)   Give the constraint a moment-relative interpretation

---

where he kills several he has lost control of his agency and thus is no longer morally responsible for his actions, in which case rules like DK no longer apply to him. I think this worry is confused. Strictly speaking, Raskolnikov does have the third option available of not killing anyone because this course of action is possible for him qua agent. However, Raskolnikov also knows that he will fail if he attempts to take this option because he knows he is disposed to suffer a weak will in this kind of circumstance. Thus, in practice he must make a choice between killing one now and killing several later. However, any reader who is still skeptical about the possibility of RASKOLNIKOV can still accept the central arguments in this chapter. For, as we will see in §6, the RASKOLNIKOV-like cases that are actually crucial to my argument are cases where it is external factors (and not internal psychological factors) that constrain an agent's choices, and the present worry does not extend to such cases.

(iii)   Count the case as a strong moral dilemma where the agent will violate an all-things-considered obligation no matter what he does.

Which of these three answers should we endorse when faced with these kinds of cases? There is much dispute on this question. Some argue that the moment-neutral answer is generally correct. For example, McNaughton and Rawling (1993), and Portmore (2011, pp.103-108) argue for this answer. Others argue that the moment-relative answer is generally correct. For example, Brook (1991) and Kamm (1992) argue for this answer. Still others may argue for the moral dilemma answer. This debate is particularly controversial because commonsense morality (at least as judged by folk intuitions) does not clearly support one of these answers over the others, as empirical work by Lopez et al. (2009) demonstrates.

In this chapter I will not take a stand on which position is correct in this debate. I will not argue that constraints are generally best interpreted as moment-neutral, or moment-relative, or as leading to moral dilemmas. Instead, I will defend a much more modest thesis. I will argue that there are at least some special cases of agent-centered constraints where the moment-neutral answer is the correct answer. In particular, I will present two cases where I think commonsense morality clearly tells us that an agent-centered constraint is moment-neutral. For my purposes it does not matter if the cases I discuss are outliers. Perhaps the vast majority of agent-centered constraints are best interpreted as either moment-relative constraints, or constraints involving strong moral dilemmas. However, all I need for my argument to go forward is for there to be at least one agent-centered constraint that is best interpreted as moment-neutral. With such an example in place I can then go on to argue that there are agent-relative values.

5. Agent-Relativity with Moment-Neutrality

As I explained in the previous section, I posit that there are at least some cases where an agent-relative constraint is best given a moment-neutral interpretation. This posit will provide a key premise in the argument that follows. To prove this posit I will present two examples where commonsense morality appears to strongly endorse the moment-neutral interpretation of an agent-relative constraint. My first example concerns the special duty parents have to care for their children. The most plausible account of such a duty would not require parents to provide a constant minimal level of care. Instead it would require parents to provide care of a quality that responds to their child's needs. Thus, when their child is especially vulnerable a high quality of care is required and, when their child can fend for herself, little or no care is required. Let's call this need-based care 'needed-care'.

Many will accept that parents have a special duty to provide needed-care for their children. Now consider the following case:

> CARE: Mark has a special duty to provide needed-care for his daughter. However, he is faced with the stark choice of either (i) providing her with needed-care now, which will have the unfortunate consequence that she will miss out on a greater amount of needed-care from him later, or (ii) not providing her needed-care now, which will somehow ensure that he is able to provide her with a greater amount of needed-care later.

In CARE Mark has a special duty to care for his daughter that is agent-relative. The duty is agent-relative because it requires Mark to provide needed-care for his daughter even if, by neglecting his daughter, he could make it the case that three other parents will provide needed-care for their children and thus cause there to be more parental care overall. Furthermore, Mark is in a RASKOLNIKOV-like case as his available options involve a trade-off between a smaller amount of something that is morally problematic in the present moment and a greater amount of the morally problematic thing in the future. What should Mark do in this case? Commonsense morality gives a clear and uncontroversial answer. Mark should forgo providing his daughter with needed-care now if doing so is the only way to ensure that she will receive a greater amount of needed care in the future. Thus, commonsense morality gives a moment-neutral interpretation of the special duty parents owe to their children, as opposed to a moment-relative interpretation, or moral dilemma interpretation. Given this moment-neutrality we might say that in CARE Mark is morally required to maximize his giving of needed-care to his daughter. Generalizing this requirement to all agents we might then endorse the following agent-relative, yet moment-neutral rule:

> (2) Each agent must maximize his giving of needed-care to his child.[38]

Thus, the special duty to provide needed-care for our children results in a maximization rule with respect to each child we are required to care for. The maximization rule appears to follow from the moment-neutrality of these agent-relative special duties.[39]

---

[38] We should remember that one can endorse such a rule without being an absolutist about it. Other duties may override this duty in certain circumstances. For example, above we said that a parent must obey (2) even if breaking it is the only way to ensure that two other parents provide needed care for their children. However, in a case where you have to choose between giving your child lunch today and allowing thousands of people to suffer terribly painful deaths we might think that (2) is overruled. Likewise, in a trade-off between completing an important personal project (such as a PhD thesis) and providing a very small amount of needed-care to one's child, the duty to maximize needed-care may be overridden by the agent-centered permission one to pursue personal projects.

[39] Some economists claim that rationality requires us to temporally discount goods, giving more weight to those that are closer to the present. If they are correct then temporal discounting must

The example above involves a special duty directed at a single moral patient (i.e. a duty owed to a single child). However, we might wonder whether the combination of agent-relativity with moment-neutrality only occurs in this kind of case. That is, we might wonder whether there are any cases where trade-offs between *different* moral patients requires rules that are agent-relative yet moment-neutral. If we think the moment-neutral answer is correct in RASKOLNIKOV, then DK is an example of such a rule. However, in this case the moment-neutral interpretation is controversial. Nonetheless, I believe there is a case involving different moral patients where the moment-neutral interpretation is uncontroversially true. Consider the following case:

> FRAUD: Natasha has set up an elaborate scheme to defraud four pensioners of their life-savings. Just as the scheme is about to defraud the first pensioner she has a moral epiphany and realizes the wrongness of what she is doing. Unfortunately the scheme is set up such that it cannot be easily reversed and it is not possible to prevent everyone from being defrauded. She has two options: (1) prevent the first pensioner from being defrauded now and allow the other three to be defrauded later, (2) Allow the first pensioner to be defrauded now and prevent the other three from being defrauded later.

In FRAUD, Natasha is in a RASKOLNIKOV-like case. For Natasha must choose between (i) preventing one of her victims from being defrauded now, but allowing several other victims to be defrauded later, and (ii) allowing one of her victims to be defrauded now, but preventing several other victims from being defrauded later. What should Natasha do in this case? Again, commonsense morality seems to give a clear and uncontroversial answer. Natasha should attempt to minimize the damage that her attempt at fraud causes to her victims, and thus should prevent as many cases of fraud as she can. Thus she should ignore the plight of the first victim in the present if this is necessary to ensure that the other three victims are not defrauded later. In endorsing this answer, commonsense morality is endorsing a moment-relative rule. Furthermore, this rule appears to be agent-relative because Natasha has a special duty she owes to *her* victims to undo the harm she is going to cause to them. Thus, in a scenario where Natasha can either prevent one of her victims from being defrauded by *her*, or three victims of some other criminal from being defrauded, she is obligated to help her victims first before helping the victims of others.

---

be added to the quantity that is maximized in (2). But then (2) would no longer be a moment-neutral rule. I do not think the economic arguments for temporal discounting are very plausible. However, it is worth noting that even if temporal discounting was required the arguments I make here would still succeed. For, although I have been arguing that rules like (2) are moment-neutral, I actually only need to claim that they are maximizing rules for my argument to proceed. And, even if we temporally discount (2), we still end up with a maximizing rule.

Therefore, FRAUD is a second case where commonsense morality says we are bound by a rule that is agent-relative yet moment-neutral. Again, this rule has a maximizing structure:

> (3) If an agent has wrongly set a causal process into play that will harm several victims, and that process cannot be stopped, then that agent must, whenever possible, minimize the harm caused to the victims by the process.

In this section we have seen that there are at least some cases where commonsense morality is committed to rules that are agent-relative yet moment-neutral. Some of these rules are concerned with duties owed to a single moral patient. Others are concerned with duties covering interpersonal trade-offs. We have also seen that the moment-neutrality of these moral rules means that they are maximization rules that require agents to maximize certain some quantity.

6. From Maximization to Value

So far I have established that commonsense morality is committed to certain agent-relative rules that have a maximizing structure. These rules require agents to maximize certain quantities. Furthermore, they do *not* do this 'accidentally' by requiring agents do an act that, by chance, results in some quantity being maximized. Instead, they do it directly by picking out the relevant quantity and then requiring agents to maximize it. These rules look similar to the rules that consequentialists endorse because they are rules that require agents to maximize certain quantities. However, consequentialist rules do not just require agents to maximize a quantity. There is a further condition that all consequentialist rules meet. The quantity they require agents to maximize must have *intrinsic moral value*. Yet so far, nothing we have said tells us that the quantities picked out by rules like (2) and (3) have intrinsic moral value. Nonetheless, I think we are warranted to infer that these quantities do have moral value.[40] This is because I think there are good reasons to accept the following conditional claim about the relationship between maximization rules and moral value:

> SV: If there is a moral requirement to maximize some quantity S, then S has moral value.[41]

---

[40] I leave it open whether this value is *instrumental* value or *intrinsic* value to simplify the argument that follows. I do this because I take a proof that there are *agent-relative instrumental values* to lead to the conclusion that there are *agent-relative intrinsic values*. For, all instrumental values are derived from intrinsic values, and it seems a basic fact that an agent-relative instrumental value can only be derived from an agent-relative intrinsic value.

[41] Broome (1991b), p. 129 appears to endorse SV but does not offer any arguments for it. As far as I know, no one else has considered SV, or taken a stand on its truth.

Why should we accept SV? One of the strongest reasons for accepting SV is that commonsense moral intuitions appear to endorse it. To see why SV is a natural consequence of commonsense intuitions, consider rule (2) discussed above. According to this rule Mark must maximize his caring for his daughter. Now exactly one of the following must be true: either the quantity that Mark must maximize has some moral value, or it does not possess any moral value. If the second possibility were true then the following would be the case. Morality requires Mark to bring about as much of his caring for his daughter as he can, and yet this thing he is bringing about has no moral value. However, this seems counterintuitive. Why would morality tell Mark to bring about as much of this quantity as he can if it has no moral worth or value? Such a thing seems odd because commonsense morality intuitively connects the maximizing of a quantity with value. Thus, whenever we have a maximizing rule commonsense morality tells us that the quantity being maximized must have some kind of moral value attached to it. If it did not then then it wouldn't be suitable candidates for maximization. In this way commonsense morality supports SV.

From this starting point we may go beyond commonsense morality and look to moral theory to explain why a principle like SV is true. There are several candidate explanations. One candidate is *value explanation*. Perhaps SV is true because maximizing rules are best explained by appeal to the value of the quantities they are directed at. Perhaps, whenever there are maximizing rules there must be corresponding moral values to explain them. Value explanations of maximizing rules are attractive because they are simple and intuitive. For example, we saw above that morality requires Mark to maximize his caring for his daughter. But why might morality require such a thing of Mark? Value explanation tells us that this quantity should be maximized because it is of moral value. This idea is appealing because maximization appears to be a fitting response to the discovery that some quantity has value. At this point it is worth noting that accepting this kind of explanation in these circumstances is not to adopt teleology as a general stance in moral theory. In others words, this explanation does not require the controversial assumption that all deontic concepts are explained by evaluative concepts. Instead, it is committed to something much more modest—the claim that a subclass of the deontic (rules that require maximizing a quantity) are explained by an evaluative concept (moral value). Thus, someone who rejects teleological explanation in general can still endorse it when it comes to maximizing rules. In fact, such a stance is already common with critics of teleology, as such critics often accept that teleology makes sense when it comes to maximization rules,

but argue against it in general because they think it is inappropriate for other kinds of rules such as absolute prohibitions.[42]

Many will find value explanations of maximization rules appealing. However, such explanations remain controversial and have been criticised by some. In particular, Scanlon (1998) rejects such explanations and instead argues for what he calls the 'buck-passing view of value'. Scanlon's 'buck-passing view' makes two claims—a negative claim and a positive claim.[43] According to the negative claim, moral rules (including maximizing rules) are not explained by the value of what is maximized. According to the positive claim, it is *reasons* that explain both moral rules and moral values. For Scanlon reasons are a primitive normative concept that cannot be reduced to any other concept. Importantly, despite holding this view, Scanlon still appears to accept SV. [44] For even if value claims do not explain maximizing rules, it can still be the case that whenever there is a maximizing rule there is a corresponding value. Indeed, Scanlon's positive thesis appears to offer an explanation of the truth of SV that rivals value-explanation (although Scanlon himself does not explicitly make this point). According to this rival explanation, maximizing rules are correlated with corresponding moral values because the reasons that explain maximizing rules also explain the corresponding moral values. Thus, consider a moral rule that tells us to maximize happiness. According to Scanlon's buck-passing view this rule cannot be explained by the fact that happiness has moral value. Instead, the rule is explained by a basic moral reason for maximizing happiness. However, the reason to maximize happiness can also explain why states of affairs containing happiness are valuable. For, according to Scanlon's analysis of value, a state of affairs or a quantity has moral value just when there are moral reasons to value it, and "to value something is to take oneself to have reasons for holding certain positive attitudes toward it and for acting in certain ways in regard to it".[45] It follows from this that a reason to maximize happiness is also a reason to value happiness. Thus, the reasons that explain maximizing rules also explain the moral value of that which is maximized.

---

[42] For example, many critics of teleology endorse a value-explanation of the general duty of beneficence. They agree that the general duty to maximize welfare is best explained by the moral value of human welfare. However, they reject teleology overall because they regard such explanations as inappropriate for other kinds of moral duties such as the duty of nonmaleficence.
[43] Scanlon presents these claims as if they necessarily go together, but Schroeder (2009) makes it clear that they are logically distinct.
[44] For example, Scanlon (1998), p. 94 says: "In one sense this is quite true: if a state of affairs is worth striving for, then it is good. But there remains an important question about the order of explanation."

[45] Scanlon (1998), p. 95.

There is a third candidate explanation of SV in addition to the two already considered. Perhaps, rather than values explaining maximizing rules, or a third factor explaining both, it is maximizing rules that explain value. Perhaps quantities like happiness are only valuable because morality requires us to maximize them. I am not aware of any moral theorists who hold this view, but the view is still worth considering as a possible explanatory thesis.[46] Such a view, by itself does not entail the truth of SV. However, if it is held that, in addition to explaining moral value, maximizing rules also necessitate moral values, then such a view offers a third kind of explanation of the truth of SV.

7. Deriving Agent-Relative Value

In the previous section I argued that commonsense morality supports the principle SV. Prior to that, I argued in §5 that commonsense morality supports two rules—rules (2) and (3)—that are agent-relative and yet have a maximizing structure. It follows from these two claims that the quantities picked-out for maximization in (2) and (3) have moral value. Thus, to take the example of rule (2), it follows that *Mark caring for his daughter* has moral value. Given this we can ask a further question: what kind of moral value do these quantities have? In particular, is the moral value they possess agent-neutral or agent-relative? When answering this question one important point needs to be considered. Rules, (2) and (3) are agent-relative rules. They are not rules that give all agents the same aim—the aim of maximizing the some identical quantity. Instead, they are rules that give different agents different aims. Thus, rule (2) requires Mark to maximize *Mark caring for his daughter*, yet does not require any other agents to maximize this quantity. Given this fact it seems that the value possessed by this quantity must be value that is relative to Mark and not neutral value accessible to all agents. For it would be odd for morality to require that Mark maximizes a particular quantity, and for it to follow from this that that quantity has moral value, and yet for that quantity to be equally valuable for everyone. This would be odd because we expect neutral values to correlate with agent-neutral requirements.

To see this more clearly let's consider the three candidate explanations of SV we examined in the previous section and see what kind of value they require in the case of (2) and (3). According to one candidate explanation, SV is true because maximizing rules are always explained by the fact that the quantity they maximize has moral value. Applying this to (2) we see that Mark's duty to maximize *his caring for his daughter* is explained by the fact that Mark caring for his daughter is morally valuable. Now suppose that this value is agent-

---

[46] Broad (1930), p. 279 discusses this as a possible view but does not ultimately endorse it.

neutral. On this supposition we get a mismatch between the explanans and the explanandum. For Mark's caring is of neutral value, and this value explains Mark's duty to maximize it, then it should be equally able to explain duties possessed by all agents to maximize Mark caring for his daughter. However, there are no such duties; only Mark has such a duty. Hence, to adequately explain why Mark has this duty and no one else, we need value that is relative to Mark.

According to a second explanation, SV is true because there are moral reasons which explain both the maximizing rule and the corresponding value. Applying this to (2) we see that Mark's duty to maximize *his caring for his daughter* is explained by Mark having a reason to maximize *his caring for his daughter*. Furthermore, we see that this reason must be a special reason for Mark (i.e. an agent-relative reason) and not a general reason for everyone, as a general reason would fail to explain why there is a special duty for Mark.[47] Thus, the reason that explains both the maximizing rule and the corresponding value is an agent-relative reason. However, it follows from this that the moral value of *Mark caring for his daughter* must be agent-relative value because an agent-relative reason for Mark to bring about a certain quantity cannot explain why such a quantity has neutral value for everyone; it can only explain how that quantity has value relative to Mark.

Finally, according to the third explanation, SV is true because a duty to maximize some quantity explains and necessitates that quantity being valuable. Applying this to (2) we see that Mark's duty to maximize *his caring for his daughter* explains why Mark caring for his daughter has moral value. However, if we suppose that this value is agent-neutral then, once again, we get a mismatch between the explanans and the explanandum. For, an agent-relative duty does not adequately explain an agent-neutral value. The fact that morality gives Mark a special duty to maximize his caring for his daughter does not offer a good explanation for the claim that his doing so has neutral value. It only seems to offer an explanation of why it might be valuable relative to Mark.

From the preceding discussion we can see that whatever explanation we give of SV, it will turn out that, in agent-relative maximizing rules like (2) and (3), the moral value of the maximized quantity is agent-relative value. At this point it is worth explaining how this defence of agent-relative value has been derived from the commitments of commonsense

---

[47] To be clear, we might think that each agent has not only a strong special reason to care for her own child, but a weak general reason to maximize the state of affairs of parents caring for their own children. Thus, we might think that everyone has a weak reason to maximize Mark caring for his daughter. However, the weak reason is not sufficient to give Mark a strong special duty to maximize his caring for his daughter. To explain his duty we need to appeal to a strong moral reason that only he possesses.

morality. Commonsense morality tells us that there are agent-centered constraints, as cases like MAFIA demonstrate. Commonsense morality also tells us that some of these agent-centered constraints are moment-neutral, as cases like CARE and FRAUD demonstrate. Thus, commonsense morality tells us that some of these agent-centered constraints are actually maximization rules that require an agent to maximize a certain quantity. In addition to this, commonsense morality tells us that if morality requires the maximization of a quantity then that quantity must possess moral value. Thus, commonsense morality tells us that some agent-centered constraints require particular agents to maximize quantities that are morally valuable. But when we consider the fact that only some agents possess these maximization duties it appears that the value they are maximizing must be value relative to them. Thus, by considering several claims that commonsense morality is committed to we reach the conclusion that there are agent-relative values.

Given all this, we may agree with the agent-relative value skeptic that commonsense morality does not directly give us the concept of *good-relative-to*, as it is not an explicit part of our ordinary moral thought and talk. However, we can answer the skeptic's arguments by pointing out that commonsense morality indirectly justifies the good-relative-to relation by its commitment to several claims that jointly entail agent-relative value. Thus, there is an important sense in which the agent-relative value is part of ordinary moral discourse and not some artificial theoretical construct. To frame this in terms of Schroeder's version of the skeptical challenge we might say that we have offered independent grounds for accepting agent-relative value, grounds that do not beg the question by assuming that such values exists. We ought to accept such values because they are required to make sense of certain commitments we find in commonsense morality.

8. Conclusion

In this chapter I have argued that the skeptical challenge fails. Commonsense morality is already committed to agent-relative values and thus skeptics are wrong to dismiss these values as theoretical constructs that are insufficiently grounded in ordinary moral thought and talk. However, the argument offered in this chapter does more than just answer these critics. It provides a general defence of agent-relative value that gives anyone curious about the status of such values a prima facie reason to accept their existence. Anyone might ask: *Why ought I to accept that there are agent-relative values?* This chapter gives one kind of answer—you ought to accept them (at least initially) because commonsense morality is committed to them. (Compare: Someone might ask: *Why ought I to accept that*

there are moral rights? Answer: Because commonsense morality is committed to rights.) These kinds of appeals to commonsense morality are not the final word when it comes to questions of moral theory. All things considered there may be good reasons to disregard some of the concepts given to us by commonsense.[48] Yet such appeals, despite their defeasibility, are an important part of moral theorising.

Importantly, the general defence of agent-relative value offered in this chapter is of a different kind to the general defence offered by agent-relative consequentialists who appeal to such values. Agent-relative consequentialists typically defend agent-relative values with theoretical considerations. For example, one strategy they use is to defend a fittingness account of value. They then argue that because fittingness is relativised to agents, values must also be relativised to agents.[49] Another strategy is to argue that agent-relative consequentialism must be true because it is the only theory that meets all the key desiderata for a moral theory, and then infer by modus ponens that agent-relative values exist.[50] Such defences of agent-relative value are independent of, yet complementary to, the defence of agent-relative value I offer in this chapter. The independence means that anyone who finds these other defences unconvincing can still find good grounds for accepting agent-relative values in the defence I present. The complementation means that the defence I offer may combine with these other defences to make an especially strong case for agent-relative value.

Finally, I will comment on what this chapter says about the overall plausibility of agent-relative consequentialism. The conclusions reached in this chapter do not entail that agent-relative consequentialism is correct. For, showing that there are agent-relative values does not show that they are of the type required by agent-relative consequentialism, or that we are always morally required to maximize those values. However, this chapter's conclusions do makes agent-relative consequentialism more plausible by defending an essential, yet controversial, premise in the overall argument for agent-relative consequentialism—the premise that agent-relative values exist.

References

Broad, C. D. (1930). *Five Types of Ethical Theory*. New York: Harcourt.

Brook, R. (1991). Agency and Morality. *Journal of Philosophy*, 88: 190-212.

---

[48] For example, Cole (2006) argues this about the concept of 'evil'.
[49] See Smith (2009) and Sussiken (2009).
[50] See Portmore (2001) and (2005).

Broome, J. (1991a). *Weighing Goods*. New York: Blackwell.

Broome, J. (1991b). The Structure of Good: Decision Theory and Ethics. In Bacharach and Hurley (Eds.), *Foundations of Decision Theory; Issues and Advances*. Basil Blackwell, pp. 123-146.

Broome, J. (2004). *Weighing Lives*. Oxford: Oxford University Press.

Cole, P. (2006). *The Myth of Evil: Demonizing the Enemy*. Westport, Connecticut: Praeger.

Cullity, G. (2015) Neutral and Relative Value. In Hirose & Olson (Eds.) *The Oxford Handbook of Value Theory*. Oxford: Oxford University Press.

Dreier, J. (1993). Structures of Normative Theories. *The Monist*, 76: 22-40.

Dreier, J. (2011). In Defense of Consequentializing. *Oxford Studies in Normative Ethics*, 1: 97-118.

Hurley, P. (2013). Consequentializing and Deontologizing: Clogging the Consequentialist Vacuum. *Oxford Studies in Normative Ethics*, 3: 123-153.

Jackson, F. and Smith M. (2006). Absolutist Moral Theories and Uncertainty. *The Journal of Philosophy*, 103: 267-283.

Kamm, F. M. (1992). Review: Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status. *Philosophy and Public Affairs*, 21: 354-38.

Kamm, F. M. (1996). *Morality, Mortality: Volume II: Rights, Duties, and Status*. Oxford University Press.

Lopez, T., Zamzow, J., Gill, M., and Nichols, S. (2009). Side Constraints and the Structure of Commonsense Ethics. *Philosophical Perspectives*, 23: 305-319.

Louise, J. (2004). Relativity of Value and the Consequentialist Umbrella. *Philosophical Quarterly*, 54: 518-36

McNaughton, D. & Rawling, P. (1991). Agent-Relativity and the Doing-Happening Distinction. *Philosophical Studies*, 63: 167-185.

McNaughton, D. & Rawling, P. (1993). Deontology and Agency. *The Monist*, 76: 81-100.

Nagel, T. (1970). *The Possibility of Altruism*. Oxford: Clarendon Press.

Parfit, D. (1984) *Reasons and Persons*. Oxford: Clarendon Press.

Portmore, D. (2001). Can an Act-Consequentialist Theory be Agent-Relative? *American Philosophical Quarterly*, 38: 363-77.

Portmore, D. (2007). Consequentializing Moral Theories. *Pacific Philosophical Quarterly* 88: 39-73.

Portmore, D. (2011). *Commonsense Consequentialism: Wherein Morality Meets Rationality*. Oxford University Press.

Regan, D. (1983). Against Evaluator Relativity: A Response to Sen. *Philosophy & Public Affairs*, 12: 93-112.

Ridge, M. (2001). Agent-Neutral Consequentialism From the Inside-Out: Concern for Integrity Without Self-Indulgence. *Utilitas*, 13: 236-254.

Scanlon, T.M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Schroeder, M. (2006) Not so Promising After All: Evaluator-Relative Teleology and Common-Sense Morality. *Pacific Philosophical Quarterly*, 87: 348-356.

Schroeder, M. (2007) Teleology, Agent-Relative Value, and 'Good'. *Ethics*, 116: 265-295.

Schroeder, M. (2009). Buck-Passes' Negative Thesis. *Philosophical Explorations*, 12: 341-347.

Sen, A. (1982). Rights and Agency. *Philosophy & Public Affairs*, 11: 3-39.

Sen, A. (1983) Evaluator Relativity and Consequential Evaluation. *Philosophy & Public Affairs*, 12: 113-132.

Smith, M. (2003). Neutral and Relative Value after Moore. *Ethics* 113: 576-598.

Smith, M. (2009). Two Types of Consequentialism. *Philosophical Issues*, 19: 257-272.

Suikkanen, J. (2009). Consequentialism, Constraints and the Good-Relative-to: A Reply to Mark Schroeder. *Journal of Ethics and Social Philosophy*, 1-9.

Bibliography

Belnap, N. & Perloff, M. (1988). Seeing to it That: A Canonical Form for Agentives. *Theoria*, 54: 175-199.

Bonevac, D. (1998). Conditional Obligation. *Nous*, 32: 37-53.

Broad, C. D. (1930). *Five Types of Ethical Theory*. New York: Harcourt.

Broad, C.D. (1942) Certain Features in Moore's Ethical Doctrines. In P.A. Schilpp (ed.) *The Philosophy of G.E.Moore*, La Salle: Open Court.

Brook, R. (1991). Agency and Morality. *Journal of Philosophy*, 88: 190-212.

Broome, J. (1991a). *Weighing Goods*. New York: Blackwell.

Broome, J. (1991b). The Structure of Good: Decision Theory and Ethics. In Bacharach and Hurley (Eds.), *Foundations of Decision Theory; Issues and Advances*. Basil Blackwell, pp. 123-146.

Broome, J. (1995). Skorupski on Agent-Neutrality. *Utilitas*, 7: 315.

Broome, J. (1999). Normative Requirements. *Ratio*, 12: 398–419.

Broome, J. (2004). *Weighing Lives*. Oxford: Oxford University Press.

Cole, P. (2006). *The Myth of Evil: Demonizing the Enemy*. Westport, Connecticut: Praeger.

Cullity, G. (2015) Neutral and Relative Value. In Hirose & Olson (Eds.) *The Oxford Handbook of Value Theory*. Oxford: Oxford University Press.

Darwall, S. (1986). Agent-Centered Restrictions from the Inside Out. *Philosophical Studies*, 50: 291-319.

Dougherty, T. (2013a). Agent-Neutral Deontology. *Philosophical Studies*, 163: 527-537.

Dougherty, T. (2013b) Rational Numbers. *The Philosophical Quarterly*, 63: 413-427.

Dreier, J. (1993). Structures of Normative Theories. *The Monist*, 76: 22-40.

Dreier, J. (2011). In Defense of Consequentializing. *Oxford Studies in Normative Ethics*, 1: 97-118.

Foot, P. (1985). Utilitarianism and the Virtues. *Mind*, 94: 107-123.

Forrester, J. (1984). Gentle Murder, or the Adverbial Samaritan. *Journal of Philosophy*, 81: 193-197.

Frankena, W. (1973). *Ethics*. Eaglewood Cliffs: Prentice-Hall.

Goble, L. (1991). The Paradox Deepens. *Philosophical Studies*, 64: 217-227.

Hammerton, M. (2016). Patient-Relativity in Morality. *Ethics*, 127: 6-27.

Hammerton, M. (2017). Is Agent-Neutral Deontology Possible? *Journal of Ethics and Social Philosophy*, 12: 319-324.

Hammerton, M. (2019). Distinguishing Agent-Relativity from Agent-Neutrality. *Australasian Journal of Philosophy*, 97: 239-250.

Hammerton, M. (2020) Deontic Constraints are Maximizing Rules. *Journal of Value Inquiry*, 54: 571-588.

Hirose, I. (2001). Saving the Greater Number without Combining Claims. *Analysis*, 61: 341-342.

Howard-Snyder, F. (1996). A New Argument for Consequentialism? A Reply to Sinnott-Armstrong. *Analysis*, 56: 111-115.

Hsieh, N.-H., Strudler, A., Wasserman, D. (2006). The Numbers Problem. *Philosophy & Public Affairs*, 34: 352-372.

Hurka, T. (2003). Moore in the Middle. *Ethics*, 113: 599-628.

Hurley, P. (2013) Consequentializing and Deontologizing. *Oxford Studies in Normative Ethics*, 3: 123-153.

Hurley, P. (2014). Comments on Douglas Portmore's Commonsense Consequentialism. *Philosophy and Phenomenological Research*, 88: 225-232.

Jackson, F. and Smith M. (2006). Absolutist Moral Theories and Uncertainty. *The Journal of Philosophy*, 103: 267-283.

Kagan, S. (1989). *The Limits of Morality*. Oxford: Clarendon Press.

Kagan, S. (1998). *Normative Ethics*. Boulder: Westview.

Kamm, F.M. (1989). Harming Some to Save Others. *Philosophical Studies*, 57: 227-260.

Kamm, F.M. (1992). Review: Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status. *Philosophy and Public Affairs*, 21: 354-389.

Kamm, F.M. (1996). *Morality, Mortality: Volume II: Rights, Duties, and Status*. Oxford University Press.

Kavka, G. (1979). The Numbers Should Count. *Philosophical Studies*, 36: 285-94.

Kelleher, J. P. (2014). Relevance and Non-consequentialist Aggregation. *Utilitas*, 26: 386-408.

Kuflik, A. (1986). A Defense of Common-Sense Morality. *Ethics*, 96: 784-803.

Kumar, R. (2001). Contractualism on Saving the Many. *Analysis*, 61: 165-70.

Lopez, T., Zamzow, J., Gill, M., and Nichols, S. (2009). Side Constraints and the Structure of Commonsense Ethics. *Philosophical Perspectives*, 23: 305-319.

Louise, J. (2004). Relativity of Value and the Consequentialist Umbrella. *Philosophical Quarterly*, 54: 518-36

McNaughton, D. & Rawling, P. (1991). Agent-Relativity and the Doing-Happening Distinction. *Philosophical Studies*, 63: 167-185.

McNaughton, D. & Rawling, P. (1992). Honoring and Promoting Values. *Ethics*, 102: 835-843.

McNaughton, D. & Rawling, P. (1993). Deontology and Agency. *The Monist*, 76: 81-100.

McNaughton, D. & Rawling, P. (1995a). Value and Agent-Relative Reasons. *Utilitas*, 7: 31-47.

McNaughton, D. & Rawling, P. (1995b). Agent-Relativity and Terminological Inexactitudes. *Utilitas*, 7: 319-325.

McNaughton, D. & Rawling, P. (1998). On Defending Deontology. *Ratio*, 11: 37-54.

McNaughton, D. & Rawling, P. (2002). Conditional and Conditioned Reasons. *Utilitas*, 14: 240-248.

Nagel, T. (1970). *The Possibility of Altruism*. Oxford: Clarendon Press.

Nagel, T. (1986). *The View from Nowhere*. New York: Oxford University Press.

Nozick, R. (1974) *Anarchy, State, and Utopia*. New York: Basic Books.

Otsuka, M. (2011). Are Deontological Constraints Irrational? In Bader and Meadowcroft (Eds.) *Cambridge Companion to Nozick's Anarchy, State, and Utopia.* Cambridge: Cambridge University Press.

Parfit, D. (1979). Prudence, Morality and the Prisoner's Dilemma. *Proceedings of the British Academy,* 65: 539-564.

Parfit, D. (1984) *Reasons and Persons.* Oxford: Clarendon Press.

Pettit, P. (1987). Universality without Utilitarianism. *Mind,* 72: 74-82.

Pettit, P. (1989) Consequentialism and Respect for Persons. *Ethics,* 100: 116-126.

Pettit, P. (1991) Consequentialism. In Singer (Ed.), *A Companion to Ethics.* Oxford: Blackwell.

Pettit, P. (1997) The Consequentialist Perspective. In Baron, Pettit, and Slote (Eds.). *Three Methods in Ethics.* Oxford: Blackwell.

Portmore, D. (2001). Can an Act-Consequentialist Theory be Agent Relative? *American Philosophical Quarterly,* 38: 363-377.

Portmore, D. (2007). Consequentializing Moral Theories. *Pacific Philosophical Quarterly* 88: 39-73.

Portmore, D. (2011). *Commonsense Consequentialism: Wherein Morality Meets Rationality.* Oxford University Press.

Portmore, D. (2014). Agent-Relative vs. Agent-Neutral. In Hugh LaFollette (ed.), *International Encyclopedia of Ethics.* Wiley Blackwell.

Postow, B. C. (1997). Agent-Neutral Reasons: Are they for Everyone? *Utilitas,* 9: 249-257.

Quinn, W. (1989). Actions, Intentions, and Consequences. *Philosophical Review,* 98: 287-312.

Rawls, J. (1971). *A Theory of Justice.* Cambridge: Harvard University Press.

Raz, J. (2003). Numbers with and without Contractualism. *Ratio,* 16: 346-67.

Regan, D. (1983). Against Evaluator Relativity: A Response to Sen. *Philosophy & Public Affairs,* 12: 93-112.

Ridge, M. (2001). Agent-Neutral Consequentialism From the Inside-Out: Concern for Integrity Without Self-Indulgence. *Utilitas,* 13: 236-254.

Ridge, M. (2005a). Agent-Neutral vs. Agent-Relative Reason. *The Stanford Encyclopedia of Philosophy.* Edward N. Zalta (Ed.), URL = http://plato.stanford.edu/archives/win2005/entries/reasons-agent/

Ridge, M. (2005b). Universalizability for Collective Rational Agents: A Critique of Agent-relativism. *Philosophy and Phenomenological Research,* 70: 34-66.

Rønnow-Rasmussen, T. (2009). Normative Reasons and the Agent-Neutral/Relative Dichotomy. *Philosophia,* 37: 227-243.

Ross, W.D. (1930). *The Right and the Good.* Oxford: Oxford University Press.

Ross, W.D. (1939). *The Foundations of Ethics.* Oxford: Oxford University Press.

Scanlon, T.M. (1998). *What We Owe to Each Other.* Cambridge, MA: Harvard University Press.

Scheffler, S. (1982). *The Rejection of Consequentialism,* Oxford: Clarendon Press.

Schroeder, M. (2006) Not so Promising After All: Evaluator-Relative Teleology and Common-Sense Morality. *Pacific Philosophical Quarterly,* 87: 348-356.

Schroeder, M. (2007) Teleology, Agent-Relative Value, and 'Good'. *Ethics*, 116: 265-295.

Schroeder, M. (2009). Buck-Passes' Negative Thesis. *Philosophical Explorations*, 12: 341-347.

Schroeder, M. (2011). Ought, Agents, and Actions. *Philosophical Review*, 120: 1-41.

Sen, A. (1982). Rights and Agency. *Philosophy & Public Affairs*, 11: 3-39.

Sen, A. (1983) Evaluator Relativity and Consequential Evaluation. *Philosophy & Public Affairs*, 12: 113-132.

Sinnott-Armstrong, W. (1985). A Solution to Forrester's Paradox of Gentle Murder. *Journal of Philosophy*, 82: 162-168.

Sinnott-Armstrong, W. (1988). *Moral Dilemmas*. Oxford: Basil Blackwell.

Skorupski, J. (1995). Agent-neutrality, Consequentialism, Utilitarianism: a Terminological Note. *Utilitas*, 7: 49-54.

Skorupski, J. (1996). Neutral Versus Relative: A Reply to Broome, and McNaughton and Rawling. *Utilitas*, 8: 235-248.

Smith, M. (2003). Neutral and Relative Value after Moore. *Ethics* 113: 576-598.

Smith, M. (2009). Two Types of Consequentialism. *Philosophical Issues*, 19: 257-272.

Styron, W. (1980). *Sophie's Choice*. New York: Bantam Books.

Suikkanen, J. (2009). Consequentialism, Constraints and the Good-Relative-to: A Reply to Mark Schroeder. *Journal of Ethics and Social Philosophy*, 1-9.

Taurek, J. (1977). Should the Numbers Count? *Philosophy & Public Affairs*, 6: 22-40.

Thomson, J. J. (1993). Goodness and Utilitarianism. *Proceedings and Addresses of the American Philosophical Association*, 67: 145-159.

Timmermann, J. (2004). The Individualist Lottery. *Analysis*, 64: 106-112.

Tradros, V. (2011). *The Ends of Harm*. Oxford: Oxford University Press.

Vranas, P. (2007). I Ought, Therefore I Can. *Philosophical Studies*, 136: 167-216.

Wedgwood, R. (2006). The Meaning of 'Ought'. *Oxford Studies in Metaethics*, 1: 127–160.