

Deontic Logic and Ethics

Shyam Nair *

March 2, 2021

Abstract

Though there have been productive interactions between moral philosophers and deontic logicians, there has also been a tradition of neglecting the insights that the fields can offer one another. The most sustained interactions between moral philosophers and deontic logicians have not been systematic but instead have been scattered across a number of distinct and often unrelated topics. This chapter primarily focuses on three topics. First, we discuss the “actualism/possibilism” debate which, very roughly, concerns the relevance of what one will do at some future time to what one ought to do at present (§2). This topic is also used to introduce various modal deontic logics. Second we discuss the particularism debate which, very roughly, concerns whether there can be any systematic generally theory of what we ought to do (§3). This topic is also used to introduce various non-modal deontic logics. Third, we discuss collective action problems which concern the connection between the obligations of individuals and the behavior and obligations of groups of individuals (§4). This topic is also used to discuss formal systems that allow us to study the relationship between individuals and groups. The chapter also contains a general discussion of the relation between ethical theory and deontic logic (§1) and a brief consideration of other miscellaneous topics (§5).

Introduction

Though there have been productive interactions between moral philosophers and deontic logicians, there has also been a tradition of neglecting the insights that the fields can offer one another. The most sustained interactions between moral philosophers and deontic logicians have not been systematic but instead have been scattered across a number of distinct and often unrelated topics. This article will to some extent follow this approach.

*Thanks to DEON community and especially Lou Goble, Jeff Horty, and Mark Schroeder who taught me almost everything I know about these topics. Of course, I am solely responsible for my remaining ignorance. Thanks also to Michael Gifford and Zach Horne for help with getting through the basic labor needed for a survey article of this sort. And finally thanks to the editors and especially Xavier Parent for their patience with me and to Jane Spurr for help with some of the figures.

After discussing a preliminary issue concerning the interaction between deontic logic and ethics (§1), the article focuses on three topics from ethics that are useful entry points for thinking about the interaction between ethics and deontic logic.¹ First, we discuss the “actualism/possibilism” debate which, very roughly, concerns the relevance of what one will do at some future time to what one ought to do at present (§2). Second we discuss the particularism debate which, very roughly, concerns whether there can be any systematic general theory of what we ought to do (§3). Third, we discuss collective action problems which concern the connection between the obligations of individuals and the behavior and obligations of groups of individuals (§4).

In discussing each of these issues, we begin with a philosophical discussion informed primarily by work in ethical theory. We then introduce a family of formal systems that offers an interesting perspective on these issues and consider how this is related to work in ethics. Each example introduces a different family of formal systems: the first introduces various modal theories; the second various non-modal theories; the third theories that allow us to study the relationship between individuals and groups. We by no means offer a systematic survey of these families of theories; instead, select representative examples are chosen and the reader is invited to explore these families in more detail by consulting other chapters of this handbook. The discussion of these sections closes by considering general features of these theories that are of philosophical interest even outside of our guiding topics. The article closes by discussing a grab bag of topics that are worthy of further consideration (§5)

But before we begin, four preliminary issues must be mentioned. First, my discussion of issues in ethical theory presupposes a certain limited conception of what ethical theory is. What I have in mind are topics of normative ethics, metaethics, and practical reason as they are typically discussed in contemporary analytic philosophy. This limited conception leaves out many other approaches that have legitimate claim to be called “ethical theory”. But since I am most knowledgeable about these areas of philosophy, I focus on them in what follows. This is not intended to denigrate other approaches. And, to some extent, other chapters in this handbook explore some of these other approaches.

Second, I will not be discussing a number of issues in deontic logic and ethics that are of great interest and have received considerable study. This is because this handbook has already dedicated whole chapters to these issues. They include discussion of moral conflicts (Goble 2013); right, duties, and other normative positions (Sergot 2013); contrary-to-duty obligations (??? vol 2); and supererogation (McNamara vol 2). I recommend to the reader each of these chapters and, for the most part, avoid discussing these issues.

Third,, this article follows a traditional, though somewhat unfortunate, practice of treating ‘ought’ and ‘should’ as a term for discussing obligations. This does not, in my view, correspond to the ordinary English meaning of ‘ought’ and ‘should’. It is easiest to discuss obligation in English by using terms like

¹These topics reflect my conception of the state of play rather than any disciplinary consensus on which topics are most important.

‘must’ and ‘have to’. Terms like ‘ought’ and ‘should’ (in their deontic uses) are most naturally used to discuss things that are optimal in some way. There are a number of issues in ethics where this distinction matters (e.g., the discussion of supererogation). But the topics that we pursue in what follows will not require this distinction so will tolerate our somewhat unfortunate way of speaking.

Fourth, this article is designed to be readable in a variety of ways other than straight through. Each section is self-contained (save for a few places and in these places, the reader is directed to the relevant background). In §2–4 after the introductory philosophical discussion, the next subsection on formal theories has subgroupings that may be read separately. Each subsection of §5 can be read on its own. For those uninterested in the core ethical cases, the subsection on the formal theories and the subsection of further discussion offers a general survey of other ethically relevant features of these theories. Those only interested in the core ethical cases may skip the subsection that describes further details of comparison and generalization of the formal theories. I recommend that the reader feel free to sample various subsections to see whether they are important for their purposes.

1 Deontic Logic, Ethical Theory, and Neutrality

We begin with a general question: what is the relationship between deontic logic and ethics? We may make this question marginally but helpfully more precise by considering two questions:

- (i) what kinds of intellectual interactions (borrowing of ideas, citations, collaboration, etc.) have occurred between individuals who studied deontic logic and individuals who have studied ethical theory?
- (ii) what kinds of interesting relations (epistemological, metaphysical, conceptual, logical etc.) are there between the theories developed in these fields?

(i) is a question about intellectual history that is perhaps best studied by a detailed examination of the historical record. This is not the venue for such a study of the historical record. (ii), on the other hand, is a philosophical and formal question that can be studied using the tools that philosophers and logicians are familiar with. I will, therefore, primarily focus on this question.

One popular answer to the question has been that there is no particularly interesting relation between the theories of ethics and the theories of deontic logic. To the extent a theory of either field is promising, it neither tells in favor or against any theory in the other field. The dominant motivation for this view is a kind of neutrality thesis. To first approximation, the idea is that an adequate deontic logic should place no constraints whatsoever on what the correct ethical theory is. Indeed, it is a strike against a deontic logic if it is incompatible with a given ethical theory: the semantics of the logic should be compatible with various competing ethical theories; the set of valid arguments

and theorems should be compatible with competing ethical theories. For example, if a proposed deontic logic somehow ruled out the possibility that it is wrong to lie to the murderer at the door, this logic should be rejected because it is incompatible with certain Kantian moral theories.²

There are different ways of motivating and developing this neutrality thesis. But I wish to begin with an analogy concerning classical predicate logic that will help to give us a flavor of what motivates the neutrality thesis. A familiar albeit disputable picture is that first-order predicate logic is a topic neutral system for understanding the commitments of any scientific or mathematical theory.³ Logic should not decide between competing theories; scientific and mathematical methods should.

This perhaps initially innocuous sounding idea has substantial consequences. A simple example that illustrates this is that $\exists x x = x$ is a theorem of classical first-order predicate logic. But some believe that the neutrality constraint shows that this should not be a theorem of an adequate logic: It is an empirical matter whether there is at least one thing. Scientific methods not logic should settle this empirical matter.⁴

Analogously, the neutrality thesis that we are considering suggests that logic should not decide between competing ethical theories; moral argument should. So-called Standard Deontic Logic, for example, has it that $O(a \wedge b) \rightarrow O(a)$ is a theorem.⁵ But certain ethical theories reject this claim (as we will see in greater detail in §2.1). So, the argument goes, Standard Deontic Logic should be rejected. Logic should not decide on this matter; moral argument should.

Here it is important to see that the neutrality objection to Standard Deontic Logic is different from the so-called “paradoxes” that are used to object to Standard Deontic Logic. According to the neutrality objection, it does not matter whether the ethical theories that reject the theorem of standard deontic logic are true. All that matters is that they are genuine ethical theories (or perhaps that they are ethical theories that are not totally crazy). Returning to our analogy, the objector to standard first-order predicate logic of course concedes that $\exists x x = x$ is true. What they claim is that since it is an empirical matter whether it is true, $\exists x x = x$ is not a theorem. Similarly, even those who think $O(a \wedge b) \rightarrow O(a)$ is true, the objector claims, should think $O(a \wedge b) \rightarrow O(a)$ is not a theorem. Therefore, it does not actually matter whether we accept or reject the ethical theories that are incompatible with the theorem.

This, I hope, gives the flavor of the neutrality based argument for skepticism about the idea that there is an interesting relationship between deontic logic and

²See Sayre-McCord 1986 for an early discussion of this issue.

³But see MacFarlane 2000 for discussion of the difficulty of making good on this conception.

⁴Here I have in mind thoughts that motivate what Nolt 2018 calls “inclusive” or “universal free” logic.

⁵Though here we use italicized lowercase letters early in the alphabet for sentences and italicized lowercase letter late in the alphabet for variables, our conventions about these matters change at times. In particular, when discussing a formal system due to a particular author, I depart from these conventions and adopt the author’s own notational style. Unfortunately, due to the diversity of approaches discussed in this article it is difficult to adopt a single uniform style throughout.

ethical theory. This neutrality argument has been developed in more detail in a variety of ways. Each of these way of developing it involves spelling out a certain conception of what deontic logic is intended to capture. For example, according to some theorists, the truths of logic are a true in virtue of the meaning of the logical connectives alone. And, according to these theories, semantically competent users of these terms at some level know the meaning of these terms such that they are disposed to accept or at least not reject these truths (at least when they are clear-headed, thoughtful, not misspeaking, etc.).⁶ So, the argument goes, since there are semantically competent, clear-headed, thoughtful, intelligent, well-spoken moral philosophers who think that ' $O(a \wedge b) \rightarrow O(a)$ ' is false and indeed even given arguments that it is false, ' $O(a \wedge b) \rightarrow O(a)$ ' must not be a theorem.

Similar ideas have been developed by focusing on conceptual rather than semantic competence, by focusing on a special notion of logical competence, and by focusing on "metaethical" neutrality. Needless to say, these are deep philosophical waters and navigating them would require us to tackle fundamental issues concerning the relationship between logic, semantics, concepts, the a priori, and the necessary. Here is not the place to decide these issues.

Instead, I wish to just sketch two ways of responding to this concern. The first ways is conciliatory, relatively uncontroversial, and perhaps all that is needed for the purposes of continuing to take interest in the interaction between moral philosophers and deontic logicians. The second way is controversial and spells out an alternative conception of the role of deontic logic. While one need not adopt it to take interest in the issues discussed in the rest of the article, I include it because it is roughly the view that I accept and this view may have important consequences that are worth further consideration

The conciliatory view concedes to the objector their preferred conception of the role of logic. She distinguishes this notion of logic from another wider notion such as the one noted by John Burgess

Among the more technically oriented 'logic' no longer means a theory about which forms of argument are valid, but rather means any formalism, regardless of intended application, that resembles a logic in the original sense enough to allow it to be usefully studied by similar methods. (Burgess 2009: viii)

Or perhaps she only reserves the word 'logic' for the objectors intended conception and notes that one might nonetheless develop a formal system using tools that are similar to the ones used to develop a logic.

The conciliatory view, then, is that various proposed "logics" are really best interpreted (or at least may be interpreted) as formal systems that not properly called logics. So understood, they are not subject to the neutrality constraints. But of course, this also means they do not deserve the privileged status that real

⁶See Carr 2015, Cariani 2016b, Charlow 2016 for a recent discussion of arguments like this one as well as several new arguments related to the linguistic semantics of 'ought' and the decision rules that they enforce (cf. 2.3.1)

logic deserves according to the objector. So these formal systems must contend with the arguments moral philosophers give.

In saying that these systems must contend with the arguments given by moral philosophers, we need not concede that the formal properties of the system are not themselves of interest. These properties may well speak to the explanatory power, simplicity, and coherence of the resulting theory and thereby be evidence for the theory that must be considered together with the specifically moral arguments given by ethical theorists. The resulting picture, then, is one on which ethical theory and deontic logic understood as formal system building are relevant to one another though neither enjoys privilege. They are engaged in the same, or a least substantially similar, project even though they typically emphasize different topics and employ different tools.

The idea that theories in deontic logic can be interpreted this way and that so-interpreted there is an interesting relationship between deontic logic and ethical theory should, I think, be uncontroversial. And it is all that is needed to see the interest in the various topics that we will discuss below.

Nonetheless, some may believe the deontic logic can and should also be interpreted in the stricter sense and so-understood it has important contributions to make. What's more, logic in the strict sense has a kind of privileged status. There are special data that is intended to capture and it has special privilege over certain other domains. This is a view on which deontic logic has priority over ethical theory.

I do not accept this “logical priority” view and will not defend it here. But—and this is my second point about this neutrality based objection—I do believe that deontic logic can and often should be interpreted as being logic in whatever strict sense there is of that term and even so understood the neutrality objection fails. The mistake, I believe, is interpreting logic in the strict sense to somehow capture a set of truths no competent clear headed person would reject. I instead reject the idea there is any such set of truths to be had.⁷ Logic is simply a very abstract theory of reality. As such it is responsible to the totality of evidence including arguments from moral philosophy (as well as semantics, the sciences, etc.). But similarly, moral theories also are beholden to the evidence and argument for a given deontic logic. And, as I mentioned earlier, often the formal features of theories allow us to say quite a lot about the explanatory power, simplicity, and coherence of a theory.

This response, then, is a wholesale rejection of the neutrality thesis. Its relevance here is that if it is right (and I don't pretend to have given an argument that it is right), the search for some neutral real logic is simply misguided. This consequence if correct undercuts the motivation for a number of logics that have been developed or at least requires that these motivations be understood differently.

In any case, the approach taken in this article will be one on which neither ethics nor deontic logic has any kind of priority. Both are beholden to the same

⁷The conception of logic that I accept here is a form of “anti-exceptionalism about deontic logic” akin to the view of philosophical methodology given in Williamson 2008 and discussed as it applies to modal logic in Williamson 2015: 423-429

body of evidence and aim to understand a common subject matter even if their approaches are different.

2 Actualism and Possibilism

We now turn to considering three debates in moral philosophy that are especially closely related to issues in deontic logic. The debate that we will look at in this section is the one between so-called “actualism” and “possibilism”.⁸

It is easiest to introduce the issue by considering an example such as the following one due to Michael Zimmerman:

I have been invited to attend a wedding. The bride-to-be is a former girlfriend of mine; it was she who did the dumping. Everyone, including me in my better moments, recognizes that she was quite right to end our relationship; we were not well suited for one another, and the prospects were bleak. Her present situation is very different; she and her fiancé sparkle in one another’s company, spreading joy wherever they go. This irks me to no end, and I tend to behave badly whenever I see them together. I ought not to misbehave, of course, and I know this; I could easily do otherwise, but I do not. The wedding will be an opportunity for me to put this boorishness behind me, to grow up and move on. The best thing for me to do would be to accept the invitation, show up on the day in question, and behave myself. The worst thing would be to show up and misbehave; better would be to decline the invitation and not show up at all. (Zimmerman 2006: 153)

In this case, should I accept the invitation or not? So far the answer may look straightforward. But Zimmerman adds an important wrinkle to the case: suppose that “if I accepted the invitation, I would show up and misbehave (whereas I would not do this if I declined). I need not misbehave (for, as noted, I could easily do otherwise); nonetheless, this is what I would in fact do.” (ibid.: 153).

In this setting, an interesting case can be made for each answer to the question of whether to accept the invitation. Some—the possibilists—claim that I ought to accept the invitation. They point out that it is perfectly possible

⁸The classic discussions of these kinds of cases include Feldman 1986, Goldman 1976, Goldman 1978, Greenspan 1978, F. Jackson 1985, F. Jackson and Pargetter 1986, Sobel 1976. More recent discussions include Bykvist 2002, Carlsson 1999a, Carlsson 1999b, Cohen and Timmerman 2016, Curran 1995, Gustafsson 2014, F. Jackson 2014, Portmore 2011, Portmore 2013, Portmore 2017, Portmore 2019, J. Ross 2012, Timmerman 2015, Timmerman and Cohen 2016, Vessel 2003, Vessel 2009, Vessel 2016, Zimmerman 1996, and Zimmerman 2006. These examples (together with other data) have also been used to motivate some interesting proposal in natural language semantics such as Cariani 2013, Snedegar 2014, and Snedegar 2017. Cariani 2016a is especially worth consulting as it shows there are a variety of ways these proposal in natural language semantics can define validity and each choice has its unique set of costs. See XXX in this volume for more discussion of proposal in natural language semantics.

for me to accept the invitation and behave myself and this would lead to the best outcome. And possibilist think that in deciding whether to do an act we should consider the best outcome one *can* bring about that involves that act. So, according to the possibilist, I ought to accept.

Others—the actualists—claim that it is not the case that you ought to accept. They point out that while I, of course, can behave myself, I, in fact, won't behave myself if I were to accept the invitation. So if I were to accept, I would actually bring about the worst outcome. And actualist think that in deciding whether to do an act, we should consider the outcome that *would* result if one did the act. So, according to the actualists, it is not the case that I ought to accept.

This actualist verdict is a commitment of standard forms of consequentialism

S ought to do x iff the outcome of S 's doing x is better than the outcome of S 's refraining from doing x

where the outcome of S 's doing x is understood as follows:

a possible world w is the outcome of S 's doing x iff if S were to do x , then w would obtain⁹

So understood, it is not the case that I ought to accept. This is because if I were to accept, the resulting possible world would be one in which I go to the wedding and misbehave. This outcome is worse than the outcome of refraining from accepting the invitation. If I were to refrain from accepting the invitation, the resulting possible world would be one in which I do not go to the wedding and so do not misbehave. Thus according to consequentialism it is not the case that I ought to accept.

What is at stake in this debate is not just some claims about what we ought to do in certain examples. These examples are, instead, intended to make vivid different perspectives on whether and how what one will do in the future can affect what one ought to do presently; different perspectives on how facts about one's future agency can affect what one ought to do presently.

Of course, no serious deontic logic “out of the box” has a commitment about whether one ought to accept or reject the invitation. Indeed, all serious deontic logics are compatible with the actualist claim that it is not the case that I ought to accept and the possibilist claim that I ought to accept. But there is another more subtle commitment of actualists and standard consequentialism that is at odds with certain deontic logics.

To see this, consider the conjunctive act of accepting the invitation and behaving myself at the wedding. Now this, by stipulation of the case, is something that I am able to do. Since this would bring about the best possible outcome, the possibilist say that I ought to accept the invitation and behave myself.

But what does the actualist and the standard consequentialist say about this conjunctive act? To see what they say, we need to consider what would result

⁹This definition makes most sense in a setting where the so-called uniqueness assumption holds. We discuss theories that embrace alternatives to this assumption below.

from doing the conjunctive act. We know that what would result is the best outcome. So according to actualism (and standard consequentialism), I ought to accept the invitation and behave myself.

Thus, actualists and standard consequentialists are committed, then, to the claim that I ought to accept the invitation and behave myself ($O(\textit{accept} \wedge \textit{behave})$) and that it is not the case that I ought to accept the invitation ($\neg O(\textit{accept})$). This means actualists are committed to denying the following principle:

DEONTIC INHERITANCE: if p entails q , then Op entails Oq

In what follows, we will explore how this controversy looks from the perspective of deontic logic. We use this as an opportunity to discuss deontic logics that are modal logics as they are most closely connected to the consequentialist way of thinking and to the cases that we are interested (§2.1). We then return to some work in ethics about the possibilism and actualism debate and note some ways this work may contribute to the work in deontic logic (§2.2). We close with a more general discussion of some features of these theories (§2.3).

2.1 Modal Theories

We begin (in §2.1.1) by presenting the so-called Standard Deontic Logic (SDL). We then consider theories in the preference deontic logic tradition (§2.1.2). And finally theories that incorporate special representations of agency as well as preferences (§2.1.3). As we will see, SDL is not an especially helpful model of our cases. But it paves the way for more sophisticated theories that tell us something more interesting.

2.1.1 SDL

Let us remind ourselves of the Kripke semantics for modal logic. A semantic structure is a triple $\langle W, R, V \rangle$ where W is understood to be a non-empty set of possible worlds, R is understood to be a relation on W and V is function from sentence-world pairs to truth values that obeys the usual rules for non-modal vocabulary and the following additional rule for the modal operator, \Box :

$\Box p$ is true at w relative to $\langle W, R, V \rangle$ iff for all w' such that wRw' , p is true at w'

Structures like this where we put no constraint on what properties relation R has provide a sound and complete semantics for the axiom system known as **K**. Axiom system **K** consists of the theorems of propositional logic, axiom K that say $\Box(p \rightarrow q) \rightarrow \Box p \rightarrow \Box q$, the rule of necessitation that says if \top is a theorem, one may infer $\Box \top$, and the rule of modus ponens.

When investigating metaphysical possibility, certain additional axioms are interesting to consider such as this:

$\Box p \rightarrow p$

This axiom holds in every structure where R is a reflexive relation in the sense that for all w, wRw . But this axiom is undesirable in a deontic context where we interpret ‘ \Box ’ as ‘it is obligatory that’: it is an obvious albeit unfortunate fact that certain things that ought to obtain fail to obtain.

On the other hand, in the deontic setting an attractive idea is:

$$\Box p \rightarrow \Diamond q$$

where ‘ \Diamond ’ is defined so that it is equivalent to ‘ $\neg\Box\neg$ ’. In the deontic setting where we interpret ‘ \Box ’ as ‘it is obligatory that’ and ‘ \Diamond ’ as ‘it is permitted that’, this axiom expresses the natural thought that that which is obligatory is also permitted. If we add this axiom to the system **K**, and add the restriction that R is a serial relation in the sense that for all w , there is a w' such that wRw' , the semantics is sound and complete for this system. This system is often called “Standard Deontic Logic” or SDL.

How may the formal structure of SDL be interpreted? For each world, SDL assigns it a non-empty set of worlds¹⁰ and anything that is true in every world in this set is what ought to be. The key question then is how we can understand this set of worlds in a sensible way; what must this set of worlds be like for it to be the right kind of thing to witness the truth of various deontic claims.

One simple idea is that this set of worlds is somehow deontically ideal. Now in moral philosophy one does not typically encounter moral theories that determine what is obligatory by citing deontically ideal ways that the world could be. But moral theories can be understood as determining such a set. For example, a consequentialist theory can be understood as saying that the deontically ideal worlds are the worlds that contain the most value. And perhaps other theories too can be regimented in this way. Perhaps Kantians can understand the set of ideal worlds as the set of worlds where the categorical imperative is obeyed. Perhaps contractualist can understand it as the set of world in which the rules no one can reasonably reject are obeyed. Perhaps virtue theories can understand it as the set of the worlds where things are as they would be if we were fully virtuous.

Each of these interpretations has some initial plausibility and the structure of SDL does not obviously rule one out or rule another in. But there are, in fact, non-trivial constraints imposed by the SDL semantics.

We can see this by considering what a consequentialist interpretation of SDL might be. It is natural to think of the set of ideal worlds relative to some particular world as the best or most valuable available worlds. While this may be a sensible interpretation of SDL that is inspired by taking goodness to be important in much the way the consequentialist takes goodness to be important, the resulting theory is inconsistent with standard act consequentialism.

As we noted earlier, standard act consequentialism rejects DEONTIC INHERITANCE (which, recall, says that if p entails q , then Op entails Oq). SDL, on the other hand, embraces it. To see this, assume that p entails q and that Op holds. According to SDL, Op holds because each of the best worlds is a world in which

¹⁰Each world w is assigned $\{w' \mid wRw'\}$ which is non-empty because R is serial.

p . Since p entails q , then in each of these worlds q is also true. Thus, according to SDL, Oq holds because each of the best worlds is a world in which q . So SDL is incompatible with standard act consequentialism and with actualism.

SDL however is compatible with the possibilist verdict. Should we conclude from this that possibilism is correct or that SDL is incorrect? I do not believe so. SDL offers so little guidance about what to think about this concrete case that it is hard to see it as a definitive argument against the actualist view. It is also hard to take seriously the idea that consequentialism, actualism, and their verdicts about Zimmerman’s wedding case provide a decisive objection to SDL. After all, consequentialism and actualism are themselves extremely controversial views and it is a contested matter what to say about Zimmerman’s case.

I think instead we can draw two morals. First, we need to continue to look for a logic that might tell us something more interesting about the actualism/possibilism debate and in particular we may wish to look for a logic that is compatible with actualism. Second, this example should teach us that to test a moral theories compatibility with a given semantics it is not enough to find a first pass gloss on the semantics that fits with the theory. Each semantics imposes some structural constraints and these must be checked to see if they fit the theory.¹¹

Let’s turn now to richer logics that may give us some more insight into our target cases.

2.1.2 Preference Semantics

Perhaps the logic that is most closely connected with the consequentialist and actualist view is the one developed in and Lou Goble’s “A Logic of *Good*, *Should*, and *Would* Part I” (Goble 1990a) and “A Logic of *Good*, *Should*, and *Would* Part II” (Goble 1990b). Here I rehearse a simplified version of Goble’s framework. Later we will consider Hansson’s logic that puts goodness or preference at center stage as well.¹²

Goble’s theory, like consequentialism, appeals to both facts about counterfactuals and facts about the ordering of outcomes according to their goodness to determine what is obligatory. More precisely, for a given world w_i , we have a comparative similarity ordering on worlds, \leq_{w_i} , and a comparative betterness ordering on worlds \mathbf{Bt}_{w_i} . We read ‘ $w_j \leq_{w_i} w_k$ ’ as saying ‘ w_j is at least as similar to w_i as is w_k ’. We assume that this ordering is transitive (i.e., if $w_j \leq_{w_i} w_k$ and $w_k \leq_{w_i} w_l$, then $w_j \leq_{w_i} w_l$) and connected (i.e., for any w_j, w_k either $w_j \leq_{w_i} w_k$ or $w_k \leq_{w_i} w_j$). We read ‘ $w_j \mathbf{Bt}_{w_i} w_k$ ’ as saying ‘ w_j is, from the perspective of w_i , strictly better than w_k ’. We assume that this ordering is asymmetric (if $w_j \mathbf{Bt}_{w_i} w_k$, then it is not the case that $w_k \mathbf{Bt}_{w_i} w_j$) and transitive (if $w_j \mathbf{Bt}_{w_i} w_k$ and $w_k \mathbf{Bt}_{w_i} w_l$, then $w_j \mathbf{Bt}_{w_i} w_l$).

¹¹§2.3.2 discusses some limitations of non-value-based interpretations even if they pass this test.

¹²There are many other preference-based frameworks for deontic logic, but we focus on these two frameworks because they are perhaps the most directly relevant to our discussion. Two early discussions include Jennings 1974 and Lewis 1974.

Two additional and more controversial assumptions are needed about the comparative similarity ordering. First, we make the “strong centering” assumption that each world is uniquely most similar to itself (more formally: for each world w_i , $w_i \leq_{w_i} w_j$ for every w_j and if there is a w_j such that $w_j \leq_{w_i} w_i$, then $w_j = w_i$). Next we will use \leq_{w_i} to define the notion of the A -alternative to w_i , $[A]_{w_i}$, as follows:

$$[A]_{w_i} = \{w_j \mid A \text{ is true at } w_j \text{ and for all } w_k \text{ such that } A \text{ is true at } w_k, w_j \leq_{w_i} w_k\}$$

In the case where A is true at w_i , $[A]_{w_i} = w_i$ because of the strong centering assumption that we have made. In the case where A is not true at w_i , we make the “limit” assumption about the ordering which requires that $[A]_{w_i}$ is always a non-empty set of worlds.¹³

This allows us to define an obligation operator. Goble defines the operator to ensure that it only applies to contingent claims. I ignore the complexity added to the definition to ensure this but will discuss applications of Goble’s theory involving only contingent claims below. So a simplified version of Goble’s idea is this:

$$O(A) \text{ is true at } w_i \text{ iff for all } w_j \in [A]_{w_i} \text{ and for all } w_k \in [\neg A]_{w_i}, w_j B_{w_i} w_k$$

This is natural formalization of the consequentialist idea. The w_j ’s and the w_k ’s represent the outcomes of A and $\neg A$ respectively. An act is obligatory exactly if its outcomes are better than the outcomes of its negation. The notion of outcome is given to us through counterfactuals because $[A]_{w_i}$ is defined using the comparative similarity ordering that we use to evaluate counterfactuals.

This theory, as we might suspect, does not validate DEONTIC INHERITANCE (which recall is that if P entails Q , then $O(P)$ entails $O(Q)$). More interestingly however, the theory allows us to explore the formal structure of cases that are more complex variants of our wedding case. As the wedding case is told, we know that there are two possibilities where one accepts. In one possibility, one behaves and the other one does not behave. We know the first is the best possibility. We also know there is a possibility where one does not accept. In the standard telling there is one such possibility (that is relevant). But we can consider other more complex cases where there are two possibilities where one does not accept. In one possibility, one behaves; in the other, one does not. Suppose that the rankings of these possibility is that the best possibility is one in which one accepts and behaves, the next is one in which one does not accept and behaves, the third is one in which one accepts and does not behave, and the worst is one in which one does not accept and does not behave. What ought to be done in this more complex case?

According to the actualist, the answer to this question depends on what one would do. We know that for if I were to accept, I would misbehave. So it must be that I do not accept and behave at the actual world. This follows

¹³We discuss further issues related to these assumptions in §2.3.1

from our strong centering assumption which says that if I in fact accept and behave, then if I were to accept, I would behave. This gives us one (perhaps obvious) constraint on the formal analysis of the example. The actual world is not one in which I both accept and behave. This leaves us with three potentially interesting different ways the actual might be:

$$\textcircled{1} : \neg A \wedge \neg B \qquad \textcircled{2} : A \wedge \neg B \qquad \textcircled{3} : \neg A \wedge B$$

where ‘ A ’ stands for ‘I accept the invitation’, where ‘ B ’ stands for ‘I behave’, and where the numerical subscript tells us how valuable each world is (where larger the number, the more valuable). We further know that the world in which I accept and behave is not “nearby” in the relevant sense. Table 1, then, summarizes different ways this case might play out.

The spatial distance from the actual world gives us the similarity ordering on worlds relative to the actual world. The numerical subscript tells you the value of each world so that the world whose subscript is a higher number is a more valuable world. So the most valuable world is the world where I accept and behave, followed by a world where I don’t accept and behave, followed by a world in which I accept and don’t behave, and finishing with a world in which I don’t accept and don’t behave. Notice the world in which I accept and behave is always the furthest from the actual world so the counterfactual “If I were to accept, then I would not behave” comes out true in all examples.

Looking at these example and checking definitions, we can see that in all of them I ought to accept and behave and in all them I ought to behave. But consider whether I ought to accept. In Example 1–4, I ought to accept. But in Example 5–9, it is not the case that I ought to accept. In Example 5–9, consider whether I ought to not accept or where I am merely permitted to not accept. As it turns out, in Example 6–9 I ought to not accept. And only in Example 5 am I merely permitted to not accept.

These results are not, on reflection, especially surprising. But they are worth noticing. We already knew that consequentialism leads to a failure of DEONTIC INHERITANCE because we it that $O(\textit{accept} \wedge \textit{behave})$ but $\neg O(\textit{accept})$ (and Example 5–9 also show this). But looking at Goble’s natural formalization of consequentialism allows to see that the theory also rejects a number of other natural principles:

- NO CONFLICTS: $\neg(O(A) \wedge O(B))$ if $\{A, B\}$ is inconsistent
- NO PERMISSIBLE CONFLICTS: $\neg(O(A) \wedge P(B))$ if $\{A, B\}$ is inconsistent
- AGGLOMERATION: If $O(A)$ and $O(B)$, then $O(A \wedge B)$

To see this notice that in examples 6–9, we have that $O(A \wedge B)$ and $O(\neg A)$ even though $\{A \wedge B, \neg A\}$ is inconsistent so we have a counterexample to NO CONFLICTS. Next recall I mentioned as an aside the Goble’s official definition of obligation blocks tautologies and contradictions from being obligatory. So we know that $\neg O((A \wedge B) \wedge \neg A)$. Thus, these examples are also counterexamples AGGLOMERATION. Finally we can notice that in example 5 we only get a counterexample to NO PERMISSIBLE CONFLICTS.

<p>EXAMPLE 1</p>	<p>EXAMPLE 4</p>	<p>EXAMPLE 7</p>
<p>EXAMPLE 2</p>	<p>EXAMPLE 5</p>	<p>EXAMPLE 8</p>
<p>EXAMPLE 3</p>	<p>EXAMPLE 6</p>	<p>EXAMPLE 9</p>

Table 1: The dependence of what we ought to do on what we actually do in Goble's theory

Nonetheless, Goble's theory is provably consistent and avoid trivializing the logic in various ways. It also validates certain restricted principles such as the following:

NO STRICT CONFLICTS: $\neg(O(A) \wedge O(\neg A))$

And Goble's full theory has the resources to represent evaluative claims about what is good and bad and he proves a number of interesting theorems connecting these evaluative claims with claims about what we ought to do. Goble's logic, then, shows one way to have a coherent actualist and consequentialist theory.

Returning again to Table 1, notice that since the ranking of the outcomes according to how good they are is the same in all the cells, the differences between what is obligatory in each illustrates the dependency of what we ought to do on how similar worlds are to one another. Since according to strong centering, the nearest world is always the actual world, this in turn means what ought to be done partially depends on what is done. So for instance compare Example 2, 5, and 8. Example 2 and 8 feature a world in which we accept and a world in which we don't accept that are equally similar to the actual world. They feature distinct actual worlds. This results in the counterfactual 'If I were to not accept, then I would not behave' being true in Example 2 but false in 8. Similarly, the counterfactual 'if I were to not accept, then I would behave' is true in Example 8, but false in 2. Example 5 is an intermediate case where both counterfactuals are false but 'if I were not to accept, then I might behave' and 'if I were not to accept, then I might not behave' are both true. These are of course plausible claims about which counterfactuals are true in these cases.

But this in turn percolates up to what is obligatory in Goble's theory. It makes it so that in Example 2 one ought to accept, in Example 5 one is merely permitted to accept and merely permitted to not accept, and in Example 8 one ought to not accept. And this dependence of obligation on what is actually done does not quite have the same initial plausibility. One, perhaps tendentious, way to put this concern is that in Example 2 and 8 one takes seriously the possibility in which one accepts and does not behave. But in both examples there is another possibility in which one does not accept (in Example 2, the possibility that one does not accept and behaves and in Example 8, the possibility that one does not accept and does not behave) that is equally close to the actual world that one ignores in evaluating what one ought to do. Instead, one merely focuses on what one actually does. There is something, at least initially, strange about this. We will see later that critics of actualism and defenders of something like DEONTIC INHERITANCE in moral philosophy have developed this suspicion that something is strange about this kind of dependence in detail. We will look at this in the §2.2 when I return to discussing some of the developments in ethics.

We now turn to a theory developed by Sven Ove Hansson (Hansson 2001, Hansson 2013) that shares similarities with Goble's theory but, as we will see, is also interestingly different. We follow the presentation of Hansson's theory given in volume one of this handbook (ibid.). Hansson's theory makes use of

the idea of a preference relation on worlds. This is similar to Goble’s idea of using a **Bt** ordering on worlds. But Hansson’s view differs from Goble’s in two respects. First, it does not make use of the notion of comparative similarity. Second, Hansson’s view can be developed so that the ordering of worlds is a consequence of an ordering over propositions. And indeed, it is this development that is of most interest to us here.

So for Hansson we start with an ordering on propositions that is assumed to be transitive and complete. We read ‘ $p \geq q$ ’ as ‘ p is weakly preferred to q ’ and use $>$ for its strict counterpart (i.e., $p > q$ iff $p \geq q$ and $q \not\geq p$). Using this, Hansson is able to define a family of obligation operators. To do so, one selects a threshold proposition f and says $O(p)$ iff $f \geq \neg p$. So if $\neg p$ is at or below some threshold, then it is obligatory that p . The threshold can be thought of as a “least” forbidden proposition. This is because it is forbidden that $p, F(p)$, is equivalent to $O(\neg p)$. Since $O(\neg p)$ iff $f \geq p$, $F(p)$ iff $f \geq p$. Thus any proposition q such that $q > f$, $\neg F(q)$. So in this sense, f is the least forbidden proposition.

So understood the semantics invalidate both DEONTIC INHERITANCE (which recall is that if p entails q , $O(p)$ entails $O(q)$) and AGGLOMERATION (which recall is if $O(p)$ and $O(q)$, then $O(p \wedge q)$). Indeed it validates no interesting theorems that can be stated solely in deontic vocabulary. But one may add structure of \geq and study the results. One particularly interesting property that Hansson disusses is this:

$$\geq \text{ is interplorative iff } (p \geq (p \vee q) \geq q) \text{ or } (q \geq (p \vee q) \geq p)$$

This in turn is equivalent two other notable conditions:

$$\geq \text{ is interpolative iff } \geq \text{ satisfies the following two conditions: (a) } \\ p \geq (p \vee q) \text{ or } q \geq (p \vee q) \text{ and (b) } (p \vee q) \geq p \text{ or } (p \vee q) \geq q$$

The idea of interplorativity says a disjunction can be as preferred as one of its disjuncts or as dispreferred as one of disjuncts or anything in between. What it prohibits is a disjunction being more preferred than both of its disjuncts or more dispreferred than both of its disjuncts.

If we accept this assumption, the resulting logic has more structure. In particular if accept clause (a), we ensure that AGGLOMERATION holds. If we accept clause (b) we ensure the following principle holds:

$$\text{DISJUNCTIVE DIVISION: if } O(a \wedge b), \text{ then } O(a) \text{ or } O(b)$$

Nonetheless we still do not have DEONTIC INHERITANCE. To see this, consider the following ordering:

$$p > f > (p \vee q) \geq q$$

Obviously, this satisfies interpolativity. Additionally suppose that the ordering is such that logically equivalent formula occur in the same place. Now we have $O(\neg p \wedge \neg q)$ because $\neg(\neg p \wedge \neg q)$ is logically equivalent to $(p \vee q)$ and $f > (p \vee q)$. But we have $\neg O(\neg p)$ because $\neg \neg p$ is equivalent to p and $p > f$. Thus, even though $\neg p \wedge \neg q$ entails $\neg p$ and $O(\neg p \wedge \neg q)$, we do not have $O(\neg p)$.

How shall we interpret this preference relation? The name suggests one obvious interpretation is that the preferences are the preferences of the agent in question. Another natural interpretation that fits will with the consequentialist perspective is that it is a goodness ordering.

Whichever interpretation we might choose, however, we need to ask whether it vindicates interpolativity. There is some initial plausibility to the idea that the better-than relation does. Suppose p is the proposition that it is a sunny day and q is the proposition that it is a rainy day. In this case, it hard to see how the the proposition that it is a sunny day or a rainy day could be strictly worse or strictly better than both of these propositions. That said, the assumption is not trivial and is not satisfied by a variety of interpretation. For example, as Hansson 2013: 490-1 observes, expected utility generates preference orderings that do not satisfy interpolativity.

In any case, Hansson’s theory, like Goble’s theory, is capable of vindicating the actualist claim that you ought to accept and write, but it is not the case that you ought to accept. To see this, consider this ordering:

$$\neg\textit{accept} > f > (\neg\textit{accept} \vee \neg\textit{behave}) \geq \neg\textit{behave}$$

Once again suppose that the ordering is such that logically equivalent formula occur in the same place. Now we have $O(\textit{accept} \wedge \textit{behave})$ because $\neg(\textit{accept} \wedge \textit{behave})$ is logically equivalent to $(\neg\textit{accept} \vee \neg\textit{behave})$ and $f > (\neg\textit{accept} \vee \neg\textit{behave})$. But we have $\neg O(\textit{accept})$ because $\neg\textit{accept} > f$. Further, we have $O(\textit{behave})$ because $f > \neg\textit{behave}$.

Thus, Hansson and Goble both can get the result that $O(\textit{accept} \wedge \textit{behave})$ and $\neg O(\textit{accept})$. Interestingly however, Goble’s theory, but not Hansson’s, allows for there to be cases where $O(\textit{accept} \wedge \textit{behave})$ and $\neg O(\textit{accept})$ and $\neg O(\textit{behave})$. Since Hansson’s theory validates DISJUNCTIVE DIVISION from above (if $O(a \wedge b)$, then $O(a)$ or $O(b)$) it cannot allow for this. One might expect given our discussion earlier that this is because Goble does not require that **Bt** to be interpolative. But, in fact, Goble’s theory predicts failures of DISJUNCTIVE DIVISION even in settings where the ordering satisfies interpolativity. Let’s look at this in detail

To do this, we need to fill in the wedding case differently than we have so far. Figure 1 summarizes the situation.

As we can see, what I actually does is stay at home and behave. And just as in the original telling, if I were to accept, I would go to the wedding and misbehave and that would be the worst thing of all. And once again as the original telling goes, if the agent were to accept and behave, I would go to the wedding and everything would great so that is better than both of these outcomes.

But there is an important difference with this example. It is one where if I were to misbehave, I would in fact do it at home having not accepted the invitation and not gone to the wedding. Finally, for whatever reason this is best outcome of all. This information will allow us to see that Goble’s theory does not validate DISJUNCTIVE DIVISION.

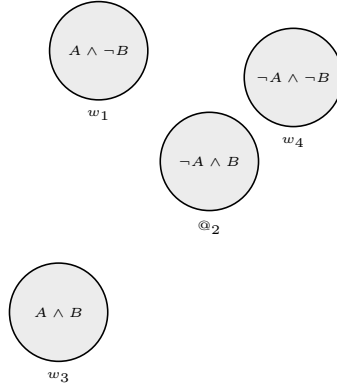


Figure 1: Interpolativity without Disjunctive Division in Goble's Theory

According to Goble, in order to check whether we ought to accept and behave. We need to compare the nearest world or worlds in which you do accept and behave to the nearest world or worlds in which that conjunction is false; if those *accept* \wedge *behave*-worlds are uniformly better than those where the conjunction is false, you ought to accept and behave. We can see that the nearest world and only world in which one accepts and behaves is w_3 and the nearest world in which this conjunction is false is the actual world, $@_2$. Finally, w_3 is better than $@_2$ so we have $O(\text{accept} \wedge \text{behave})$.

Next consider whether you ought to accept. We can see that the nearest *accept*-world is w_1 and the nearest \neg *accept*-world is $@_2$. But $@_2$ is better than w_1 . So $\neg O(\text{accept})$ and indeed, $O(\neg \text{accept})$. For we have seen that the nearest \neg *accept*-world is $@_2$ and the nearest *accept*-world is w_1 and $@_2$ is better than w_1 .

Finally consider, whether you ought to behave. We can see that the nearest *behave*-world is $@_2$ and the nearest \neg *behave*-world is w_4 . But w_4 is better than $@_2$. So $\neg O(\text{behave})$ and indeed, $O(\neg \text{behave})$. Thus, DISJUNCTIVE DIVISION can fail in Goble's theory.

Nonetheless, we can show that $\mathbf{Bt}_{@_2}$ is interpolative in this case. Begin by noticing that Figure 1 is a representation of the following information¹⁴:

$$\begin{aligned} @_2 \leq_{@_2} w_4 \leq_{@_2} w_1 \leq_{@_2} w_3 \\ w_4 \mathbf{Bt}_{@_2} w_3 \mathbf{Bt}_{@_2} @_2 \mathbf{Bt}_{@_2} w_1 \end{aligned}$$

Next replace each world in the ordering with the proposition that is true at exactly that world:

$$\begin{aligned} \neg \text{accept} \wedge \text{behave} \leq_{@_2} \neg \text{accept} \wedge \neg \text{behave} \leq_{@_2} \text{accept} \wedge \neg \text{behave} \\ \leq_{@_2} \text{accept} \wedge \text{behave} \end{aligned}$$

¹⁴Beware in translating between Goble's and Hansson's theories: Hansson's \geq is most similar to Goble's \mathbf{Bt}_w rather than Goble's \leq_w !

$$\neg accept \wedge \neg behave \quad \mathbf{Bt}_{@_2} accept \wedge behave \quad \mathbf{Bt}_{@_2} \neg accept \wedge behave \\ \mathbf{Bt}_{@_2} accept \wedge \neg behave$$

We extract an ordering on *accept* and *behave* and their negations, by placing them in the same spot in the betterness ordering as the closest world in which they hold. We do this because Goble's semantics determines the deontic status of these claims by considering the value of the closest worlds in which they hold. So this gets us the following richer ordering:

$$\neg accept \wedge \neg behave, \neg behave \\ \mathbf{Bt}_{@_2} \\ accept \wedge behave \\ \mathbf{Bt}_{@_2} \\ \neg accept \wedge behave, \neg accept, behave \\ \mathbf{Bt}_{@_2} \\ accept \wedge \neg behave, accept$$

We similarly extract an ordering on disjunction by placing them in the same spot in the ordering as the closest world in which they hold:

$$\neg accept \wedge \neg behave, \neg behave, accept \vee \neg behave \\ \mathbf{Bt}_{@_2} \\ accept \wedge behave \\ \mathbf{Bt}_{@_2} \\ \neg accept \wedge behave, \neg accept, behave, accept \vee behave, \neg accept \vee behave, \neg accept \vee \\ \neg behave \\ \mathbf{Bt}_{@_2} \\ accept \wedge \neg behave, accept$$

Notice that this ordering is in fact interpolative in Hansson's senses. Each disjunction is not strictly preferred to both of its disjuncts and is not strictly dispreferred to each of its disjuncts.

So why the different results about what we ought to do? The crucial difference is not the structure of the ordering. It is, instead, the decision rule one uses to determine what is obligatory given the ordering. On Hansson's theory, there is a fixed spot in the ordering and one checks to see if $\neg p$ is below that spot. If so, p is obligatory. But there is no fixed spot such as this in Goble's theory. Instead for each proposition p , one goes to p 's spot in the ordering and checks to see if $\neg p$ is below that spot. If so, p is obligatory. This difference in the views is what explains why Goble's theory fails to validate DISJUNCTIVE DISTRIBUTION.

Since we are assuming in this example you ought to accept and behave, we know the threshold in the ordering for Hansson will have to be between the row for *accept* \wedge *behave* and its negation $\neg accept \vee \neg behave$. Using this threshold, we can see one also ought to *accept* because $\neg accept$ is below the threshold so DISJUNCTIVE DIVISION holds. Interestingly, in this example, Hansson's theory says that one ought to accept (as we just saw) and that one ought to not accept (accept is at the lowest spot in the ordering so below the threshold).

As we noted earlier, Goble’s theory validates NO STRICT CONFLICTS so does not allow for cases where one ought to accept and one ought to not accept. And this is precisely because the threshold on his theory is relative to each proposition. For *accept* one compares it to $\neg\textit{accept}$ and vice versa, the result is that one ought to not accept and it is not the case that one ought to accept. For $\textit{accept} \wedge \textit{behave}$ one compares it to its negation $\neg\textit{accept} \wedge \neg\textit{behave}$ and similarly for *behave* and $\neg\textit{behave}$. In each of these cases the threshold varies being as low as the fourth row (for *accept*) and as high as the first row (for $\neg\textit{behave}$).

We will discuss issues related to distinct decision rules in greater detail below (§2.3.1). But what should the consequentialist and actualist make of these two different approaches to this variant of Zimmerman’s wedding case? Certainly standard act consequentialism is more similar to Goble’s theory.

Hansson’s theory is however more similar to a consequentialist theories that are sometimes called absolute level satisficing theories. According to these theories, there is a some level of goodness such that if an act produces at least that much goodness, it is permissible.¹⁵ Hansson’s theory is different from this one in two ways. First, one determines *p*’s deontic status on Hansson’s theory not by looking at *p*’s place in the ordering but instead by looking at $\neg p$ ’s place in the ordering. Second, Hansson uses the ordering to determine obligations rather than permissions as the satisficing absolute level consequentialist does. Nonetheless, they are united in thinking there is some threshold. And that threshold determines the forbidden. One interesting result of our discussion is that a further data point on which consequentialist of the standard sort and absolute level satisficing consequentialists may disagree is the status of DISJUNCTIVE DIVISION.

In sum, both Goble and Hansson develop logics that can coherently model actualists and consequentialist reasoning. These logic differ about, among other things, whether they validate DISJUNCTIVE DIVISION and whether they validate NO STRICT CONFLICTS. This in turn is related to a difference in how goodness determines what is obligatory according to each theory, a difference in decision rule. So these logics are models of distinct actualist theories: Goble’s logic fits best with standard consequentialism while Hanssons’s perhaps fits best with absolute level satisficing consequentialist theories.

2.1.3 Semantics with Agency

As we have seen, the debate between actualists and possibilists concerns the relevance of one’s future actions to what one presently ought to do. But so far, we have not considered any logics that explicitly represent the decisions and actions of agents through time. We turn to considering such theories now. We will look at the logic developed by John Horty in his Horty 2001 as it is perhaps the most well-known such system.

Horty’s theory can be thought of as combining a certain logic of agency, so-called stit *logic*, and a certain kind of preference semantics. We begin by

¹⁵See Hurka 1990 and Slote 1984 for developments of satisficing approaches; see Bradley 2006 for criticism

introducing *stit* logic.

stit logic, in turn, can be thought of as beginning with a logic of branching time and adding some way of representing agency within this framework. The logic of branching time is a kind of modal logic where we have set of points, *Tree*, and we call elements of *Tree* *moments*. We also have an ordering on moments $<$ that is irreflexive, transitive, and *tree-like* in the sense that if $m_1 < m_3$ and $m_2 < m_3$, then $m_1 < m_2$ or $m_2 < m_1$ or $m_1 = m_2$. These properties of the ordering on moments ensure that we can represent them in a way that looks like a tree: there is a single trunk and then branches emerging from this trunk. This represents the openness of the future (and the determinateness of the past). A collection of moments is *linearly ordered* just in case for any two of them, m_1 and m_2 , $m_1 < m_2$ or $m_2 < m_1$ or $m_1 = m_2$. A collection of moments is *maximally linearly ordered* just in case it is linearly ordered and not a proper subset of any linearly ordered set. Such maximally linearly ordered collections of moments are called *histories*. Intuitively, a history represent one complete way the world could develop. Using these resources to give us our frames, one can define a model and give conditions for the truth of various claims about the past and the future. It turns out to be best in this setting to define the truth of formulas relative to a pair consisting of a moment and a history through that moment. We skip the details of the semantics for the temporal logic as it will not be of central interest in what is to come.

To this temporal logic, we add resources for representing agency. We add a set *Agent* that is thought of as populated by the individual agents that are of interest to us. We also add a function *Choice* that takes one from an agent and a moment to a partition of the histories through that moment. The cells of the partition are intuitively thought of as the actions that we can perform at that moment. Our acts allow us to select among the histories through a moment which collection of these history will occur. We also say *Choice* applied to an agent, moment, and a history through that moment returns whatever act that history belongs to. This represent the act that an agent does at a moment/history pair.

We now introduce in addition to the normal kinds of sentences one has, sentences concerning what an agent does or claims saying that an agent “sees to it that”. There are a variety of such operators that have been proposed but we focus on a simple one often called *cstit* because it is a “sees to it that”-operator that is closely related to some ideas originally due to Brian Chellas (Chellas 1969). Intuitively, the idea is that an agent sees to it that A at a moment and history through that moment, just in case A is guaranteed to be true by the act that the agent does at at that moment/history pair.

Putting all of this together, we have a *stit frame*, $\langle Tree, <, Agent, Choice \rangle$. A model M based on the frame gives the truth values of formulas at pairs, m/h , of moments and histories through those moments in the usual way for standard formulas with the following truth conditions for the special *cstit* operator:

$$[\alpha \text{ cstit: } A] \text{ is true relative to } M, m/h \text{ iff } Choice^{m,\alpha}(h) \subseteq |A|^{M,m}$$

where $Choice^{m,\alpha}(h)$ represents the act α does at the moment and history

through that moment pair and $|A|^{M,m}$ is the set of histories through m such that A is true at the pair consisting of m and that history. So this just says, as intended, that α sees to it that A at a moment and history through that moment, m/h , just in case the act the agent does at m/h ensures the truth of A . Within a framework like this one, one can study the agency individuals as well as groups (a topic we return to in §4.1.2) including modeling claims about what people or groups are able to do.

Though there is much to say about this, we turn now to how one can add a deontic logic on top of this framework. The simplest way to do this would be to simply include an additional function that maps each moment into a set of histories through that moment which are ideal. While this may be a satisfactory account for what things ought to be, Horty argues at length that it is not a satisfying account of what we ought to do.¹⁶ We do not have space here to consider these arguments in detail. Instead, I will simply state Horty's preferred approach.¹⁷

Horty's approach is to work with a dominance ordering on actions. He constructs this ordering by assuming there is a function *Value* that assigns to each history a number that is understood to be a measure of how good that history is. He then lifts this ordering on histories to be an ordering on arbitrary propositions. We write $P \leq Q$ to mean ' P is weakly preferred to Q '. And $P \leq Q$ iff $Value(h) \leq Value(h')$ for all $h \in P$ and $h' \in Q$. We use $P < Q$ then in the standard way to mean that $P \leq Q$ and $Q \not\leq P$.

Finally, Horty uses this preference ordering on propositions to construct a dominance ordering on act. The idea Horty's weak dominance ordering is intends to capture is the idea of one act being at least as preferred to another in every given state of the world that is independent of the act. These states of the world that are independent of the act are just understood to be given by the choice set of all the other agents (where we may include mother nature as one agent as well). So $State^{m,\alpha}$ is introduced and understood to be $Choice^{m,Agent-\{\alpha\}}$. One then defines the weak dominance ordering \preceq on acts, K_1 and K_2 , by saying $K_1 \preceq K_2$ iff $K_1 \cap S \preceq K_2 \cap S$ for each $S \in State^{m,\alpha}$. Strict dominance is understood in the usual way then so $K_1 \prec K_2$ iff $K_1 \preceq K_2$ and $K_2 \not\preceq K_1$.

This gives us our dominance ordering on acts that in turn allow us to define what it is that we ought to do. We do this via first identifying the optimal act for an agent at a moment. This is understood to be those acts available to the agent at that moment that are not strictly dominated:

$$Optimal^{m,\alpha} = \{K \in Choice^{m,\alpha} \mid \text{there is no } K' \in Choice^{m,\alpha} \text{ such that } K \prec K'\}.$$

Finally, then, we introduce the two place operator $\odot[\dots cstit \dots]$ and we form sentence $\odot[\alpha cstit: A]$ that we read as ' α ought to see to it that A ' and is true just in case every optimal act ensures the truth of A :

¹⁶See Horty 2001: ch. 3 for discussion.

¹⁷But see §5.1 for an overview of the issues Horty and others are responding to.

$\odot[\alpha \text{ cstit: } A]$ is true relative to $M, m/h$ iff $K \subseteq |A|^{M,m}$ for each $K \in \text{Optimal}^{m,\alpha}$

where we assume our models are now based on frames enriched with *Value* and the other material we have defined in terms of these items.¹⁸¹⁹

So what does this theory have to say about our target cases? We now, as is shown in Figure 2, have the resources to model the case as one where the agent first faces choice of whether to accept or not and later if one accepts facing a choice to behave or not. The idea here is at m_1 , one chooses whether to accept

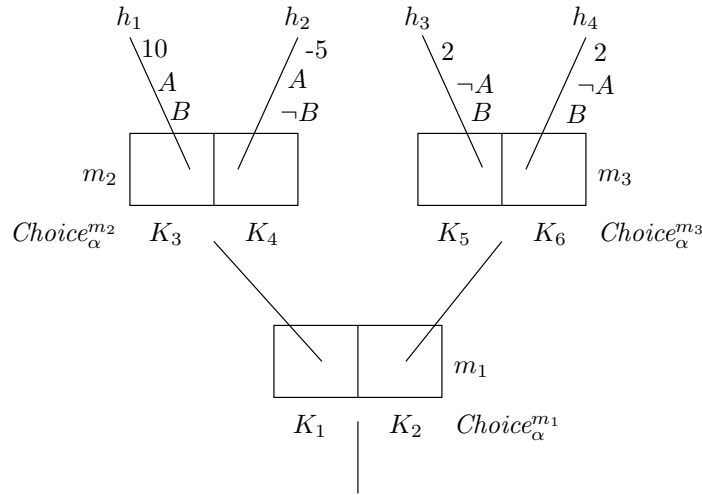


Figure 2: Wedding attendance decision tree

the wedding invitation. K_1 represents the choice of accepting the invitation. Accordingly, A is the proposition that you accept and it is true in every history that you get to if you do K_1 . K_2 on the other hand is not accepting so $\neg A$ is true in all of those histories. If you choose to accept the invitation, you arrive at m_2 and faces a choice of whether to behave. K_3 is the act of behaving and accordingly B which represents the claim that you behave is true in the history where you do K_3 . Since this is a very good outcome, we have given the history the value of 10. On the other hand, one could choose K_4 and not behave. Accordingly, $\neg B$ is true here and it is assigned a low value of -5 . On the other hand, at m_1 you could choose to not accept. This would result in you arriving at m_3 where you can choose between K_5 and K_6 which are perhaps simply different ways of spending sometime like reading a book or watching tv.

¹⁸This definition is actually not Horty's. It is however a theorem of Horty's theory that holds in finite settings. The official definition Horty gives is more complicated and allows him to handle cases that arise in certain infinite settings. Since we ignore those cases here, I opt to use as a definition of the operator what is a theorem in Horty's system.

¹⁹Horty also develops an obligation operator that is closer to standard act consequentialism in Horty 2001: §5.4. We do not have space to explicitly present and discuss this analysis here.

Regardless of what you do, you will have behave. Accordingly, B is true in these histories and each history is of some small middling value 2.

What does Horty’s theory claim that you ought to do in this case? More precisely, at m_1 what should one do. To determine, this we start with the optimal acts. They will come from the set $Choice^{m_1, \alpha}$ so we need to check to see how K_1 and K_2 rank relative to one another. In situations where, the agent is the only one choosing such as this one the set $State^{m_1, \alpha}$ will only contain one element which is the set of all histories through m_1 , $\{h_1, h_2, h_3, h_4\}$. Accordingly, the dominance ordering and the preference ordering will be the same. Since we know $Value(h_1) > Value(h_3)$ and $Value(h_2) < Value(h_3)$, we know neither K_1 nor K_2 are weakly preferred to one another so there is no dominance. Thus, both K_1 and K_2 are non-dominated so both are optimal.

In this setting, $\odot[\alpha \text{ cstit: } A \wedge B]$, $\odot[\alpha \text{ cstit: } A]$, and $\odot[\alpha \text{ cstit: } \neg A]$ are all false. $A \wedge B$ is not guaranteed to be true by K_1 and it is guaranteed to be false by K_2 . A is not guaranteed to be true by K_2 . $\neg A$ is not guaranteed to be true by K_1 . Thus, there is nothing that one ought to do in this case. Instead, one is merely permitted to do any of these things.

This is interesting as it corresponds to neither the actualist nor the possibilist view. It is however like the actualist view in that it does not treat your future decisions differently than simply states of the world that you don’t control. To see this, we can just notice that the very same results follow if we simply have no choices at m_2 and m_3 . It however differs from standard consequentialist actualist views in that it does not assume a single fully specified possibility results from one’s act. We discuss this abstract difference between standard consequentialism, Horty’s theory, and the other theories that we have described in §2.3.1.

But one strange result of Horty’s theory is that even the claim $\odot[\alpha \text{ cstit: } \neg(A \wedge \neg B)]$ is false. While K_2 guarantees $\neg(A \wedge \neg B)$, K_1 does not guarantee the truth of this as it is false at h_2 . Yet it seems hard to deny that one ought to not both accept and misbehave. So it appears that this is not a sensible treatment of our case.

Perhaps unsurprisingly then, Horty does not himself endorse this way of using his theory to analyze this example.²⁰ Instead, Horty expands his theory to deal with cases in which one faces a sequence of choices. The expansion involves introducing the notion of a strategy. It turns out to be a somewhat delicate matter how a strategy must be defined and how a “strategic”-ought-to-do operator, $\odot[\dots \text{scstit} \dots]$ is to be defined. But we will pass over these complications and work with a simplified intuitive picture as this will be enough for our example.

In our example, the available strategies are these:

$$\begin{aligned} s_1 &= \{\langle m_1, K_1 \rangle, \langle m_2, K_3 \rangle\}, s_2 = \{\langle m_1, K_1 \rangle, \langle m_2, K_4 \rangle\}, \\ s_3 &= \{\langle m_1, K_2 \rangle, \langle m_3, K_5 \rangle\}, s_4 = \{\langle m_1, K_2 \rangle, \langle m_3, K_5 \rangle\} \end{aligned}$$

Each strategy, then, is a coherent sequence of choices one could make at a

²⁰Horty’s discussion of the actualism/possibilism debate is found in Horty 2001: §7.4.3.

moment. Roughly, one considers which strategy is optimal by (in a setting where the state is just all of the histories) checking whether every history in one is at least as good as every history in the other. It is easy to see this reduces to comparing the value of h_1 , h_2 , h_3 , and h_4 and obviously h_1 is higher than all of them so s_1 will strictly dominate all of them and so be the only optimal strategy. From here one evaluates $\odot[\alpha \text{ scstit}: A \wedge B]$, $\odot[\alpha \text{ scstit}: A]$, $\odot[\alpha \text{ scstit}: \neg A]$, etc. by considering whether s_1 settles the embedded sentence. And it is easy to see s_1 ensures the truth of $A \wedge B$ and A and ensures the falsity of $\neg A$. Thus, this version of Horty’s view agrees with the possibilist

So on Horty’s picture there is a way of accommodating possibilist reasoning about these cases. But, as Horty himself notes, there is no obvious way of accommodating actualist reasoning in his theory.²¹

In sum then, Horty’s picture gives us two different views about what we ought to do in this case depending on whether focus on the $\odot[\dots \text{cstit} \dots]$ or the $\odot[\dots \text{scstit} \dots]$ operator. And the view we get about the $\odot[\dots \text{cstit} \dots]$ operator corresponds to no major view in ethical theory while the $\odot[\dots \text{scstit} \dots]$ has similarities to possibilist views.

We face, then, a question of what role these different operators play and whether one operator is to be privileged above another operator. We also face a question of whether any plausible view in ethics corresponds to the claims made about our example by the $\odot[\dots \text{cstit} \dots]$ operator.

Thus, while Horty’s view allows us to more explicitly spell out the role of diachronic agency in determining what we are obligated to do and allows to provide a model for coherent possibilist reasoning, a number of questions remain about how to best understand this view. In the next subsection, I present a leading idea in ethics about the actualism and possibilism debate and suggest that it helps to shed new light on some of the issues for Horty’s theory as well as the issues for Goble and Hansson’s theory that I mentioned earlier.

2.2 An Insight From Ethics

We turn now to a prominent line of criticism of actualism and line of defending something akin to DEONTIC INHERITANCE from ethical theory.²² These ideas originate in the work of Holly Smith and have since been developed further by Douglas Portmore and Jacob Ross.²³ In this article, we will only present a simplified version of some of the basics of this approach.

²¹Horty does suggest that the actualist claim may correspond to the claim that one ought to not accept conditional on not behaving. But this claim is uncontroversial and not something the possibilist would disagree with this.

²²Strictly speaking, these approaches defend a restricted version of inheritance. Very roughly, inheritance is restricted to applying to cases where doing one act involves doing another act. The details of how to spell this require a better understanding of the notion of “what is up to me” that is invoked below. This is an area where some formal work could also benefit those in ethics as the notion of an act “involving” another act could do with some formal clarification.

²³See Goldman 1978, Portmore 2011: ch. 6, Portmore 2013, J. Ross 2012. These ideas are also related to so-called “maximalism”, see Brown 2018 and Portmore 2019.

The line of thought involves a focus on the details of the connections between agency and what we ought to do. To get a feel for the view, it helps to begin with a case everyone agrees on. Suppose the best thing for me to do would be to push this button before me. It would, suppose, result in everyone being slightly happier. Suppose further that I easily can push the button: it is right in front of me, I see it, my arms are in good working order. But suppose that I will not push the button because I do not intend to do so currently. Perhaps, this is because I am a cruel person and so have decided to not push the button. In this setting, the fact that I will not push the button is simply not relevant to whether I ought to push it.

Why is that? The idea, I take it, is straightforward enough: since it is up to me whether I push the button the fact that I will not push it is no excuse. Now there is a question of in what way is it “up to me” and different theories can make this notion more precise in different ways. But a simple idea is that it is up to me in that it is under the control of my present intentions where we say:

S 's doing x is under S 's present intentional control at t iff if S were to intend at t to do x , S would do x and if S were to intend at t to do $\neg x$, S would do $\neg x$

The idea is that facts that are under our present intentional control can't get us off the hook for doing things that would bring about what is best.

This idea can be applied to the wedding case. There are, we can suppose, two different interpretations of this case. On one interpretation, if I were to intend now to accept the invitation and behave, I would in fact end up accepting the invitation and behaving at the wedding. Now of course, I do not in fact have this intention even though I could have it. Instead, I have no intention at all with regard to whether I will behave. For this reason, if I were to accept (and indeed even if I were to intend to accept and in fact accepted), I would not behave myself. On this interpretation, the idea is that I ought to accept the invitation even though accepting it would lead to a worse outcome. This is because, the view says, we should not hold fixed the fact that I will misbehave in evaluating whether to accept. We should not hold it fixed because it is under my present intentional control whether I misbehave.

On another interpretation, even if I were to intend now to accept the invitation and behave, I would not end up behaving. In this setting, whether I behave is not under my present intentional control (though, on the natural way of understanding the case, it is at the time of the wedding under my intentional control whether I behave). And the idea here is that I ought not to accept the invitation. Nonetheless, this verdict, plausibly, is compatible with DEONTIC INHERITANCE. This is because it is plausible that if it is not up to me at the present moment whether I accept and behave, then I cannot accept and behave in the sense of ‘can’ that is featured in the ‘ought’ implies ‘can’ principle. For this reason, it is plausible that it is not the case that I ought to accept and behave and so DEONTIC INHERITANCE is not violated.

Now there are many important questions and worries about this position. And there are a number of variants of this position that deal with the interesting

issues about how to make sense of the notion of control or the notion of what is up to me precise and plausible.²⁴ But even without diving into the details of this rich topic, there are a number of lessons that we can learn from this insight from moral philosophy.

First, we noted earlier (§2.1.2) that views like Lou Goble’s consequentialist inspired theory make what is obligatory dependent on what one will actually do in a somewhat curious way. This idea in moral philosophy identifies more precisely what the problematic dependency might be and provides a compelling criticism of it: while the relevant states of the world for determining what we are obligated to do may be partially determined by facts about what will (or would) happen, they do not depend on those facts that are under one’s intentional control.

Second, this insight from moral philosophy puts at center stage connection between features of our agency and our obligations. This suggest that a theory like Horty’s theory that involves representations of features of agency that are relevant to determining what agents ought to do may be closely related to this idea in moral philosophy.

But in fact, at least on first inspection, it does not seem as those Horty’s theory makes use of the notion of intentional control or can help us to get a better sense of it. Horty’s theory, instead, is built with a much sparser and better understood (albeit perhaps more idealized) set of resources than the ideas that I have been rehearsing here. That said, I believe we can use these ideas in moral philosophy to provide a certain helpful interpretation of some features of Horty’s theory.

To see this, recall that we left our discussion of Horty’s theory with two questions. First, what should we make of the fact that the $\odot[\dots cstit \dots]$ approach to the case did not correspond any sensible interpretation of it for it led to results such as $\neg \odot [\alpha cstit: \neg(A \wedge \neg B)]$. Second, while the $\odot[\dots scstit \dots]$ account was a natural development of the possibilist line of thought, it is not clear when this operator is the appropriate one to guide our action as opposed to the simpler $\odot[\dots cstit \dots]$ operator. In other words, it is not clear what different theoretical and practical roles these operators are to play and how they are related.

Now Horty does in fact quite precisely describe a certain kind of relationship between the operators. In the set up to discussing the strategic-cstit operator, Horty points out that often we do not want to consider every single possible future decisions but instead only decisions up to a time. He implements this in his models by adding a parameter for what he calls a “field of concern”. Generally a field of concern is some subset of the total histories and strategic cstit is defined only over strategies constructible in the field of concern. Horty shows that if the field of concern is only this very moment $\odot[\dots cstit \dots]$ and $\odot[\dots scstit \dots]$ are equivalent.

But what field of concern is appropriate for a given problem? This is a question Horty does not answer. Perhaps, this is because there is no fixed

²⁴See especially Portmore 2019.

answer or perhaps because it is a context-sensitive issue. But one suggestion that I would make is that the view from moral philosophy that I have just described gives us grounds for thinking that there is a privileged field of concern for each question about what one ought to do at a given moment.²⁵ It is that space of choices that is currently under your intentional control. If this is right, it suggests that the only relevant operator for modelling the core moral claims that we would like to make is the $\odot[\dots cstit \dots]$ operator evaluated relative to this privileged field of concern. This eliminates the question of which operator to give priority.

Following this line of thought, we can consider each interpretation of the case. The first interpretation recall is one one which one would accept the invitation and behave in one (presently) intended to accept the invitation and behave so the field of concern is not just the present moment but includes moments after this. This explains why $\odot[\dots cstit \dots]$ doesn't correspond to any sensible interpretation of the case: it is inappropriate to have only this moment is your field of concern when your intentional control extends beyond this moment. And we, as we saw earlier, have the possibilist result that you ought to accept and behave and that you ought to accept.

If, on the other hand, we interpret the case so that one does not have present intentional control over whether one will behaves, the field of concern diminishes to the present moment. In this context, we saw that it is not the case that one ought to accept and behave because one cannot do this. Indeed, this follows from theorem of Horty's view that says that $\neg\Diamond[\alpha cstit: A] \rightarrow \neg\odot[\alpha cstit: A]$. What's more on this way of thinking, it make sense why : since one is unable to accept and behave and one is unable to accept and not behave, this cannot be something one is obligated to do.

However, one discrepancy remains between the idea that I have been discussing from ethics and this interpretation of Horty's view. According to the idea in ethics, one ought to reject in the situation where one does not have present intentional control over whether one will behave. This is not so on Horty's view implemented in the way that I have described. According to this view, one is permitted but not required to reject.²⁶

What account for this difference is that the idea from ethics assumes that a single possibilities results from what one does. And in cases where this possibility is not determined by facts that are under one's present intentional control, this possibility is the only relevant outcome of the act. This is not so on Horty's theory. On Horty's theory, any future action available to the agent (even if it is outside the field of concern) makes it so there are multiple potential outcomes of the act. As we will see, this abstract difference between the the theories has

²⁵One thing to notice is that $Choice^{m,\alpha}$ is a partition of the histories through that moment where each "act", K , is cell of the partition. This is a kind of maximalism (of the sort discussed in n. 23) limited acts at a moment. Once we expand to discuss strategies, the resulting choice set also forms a partition that is typically more fine-grained and in this way represents a maximalism that is consonant with the fact that the maximal acts that we perform often extend over time.

²⁶We also continue to have the seemingly strange result that $\neg\odot[\alpha cstit: \neg(A \wedge \neg B)]$ even though $\Diamond[\alpha cstit: \neg(A \wedge \neg B)]$.

a number of broader consequences and it is also a difference between Horty’s theory and the theories that we have discussed before this. We will explore these more abstract differences as well as other issues below (§2.3.1).

For now though, let us take stock. The theories that we have discussed in this previous subsection (§2.1) allowed for both actualist, possibilist, and still other verdicts that are not easy to categorize. While some of the theories, such as SDL, offered us little insight as to why we should expect one set of verdicts or one form of reasoning to be correct, the theories of Goble, Hansson, and Horty offered more insight. This shows some of the fruits of formal theorizing for issues in ethics: it allows us to more fully and systematically explore the range of possible views and it allows to trace the differences between certain verdicts to different views about how rankings on outcomes are related to what we are obligated to do.

We have also seen in this subsection that ideas from ethics can help us to understand certain features of the theories from logic that are harder to grasp in the abstract. In particular, we noticed that we can get a better grip about what is strange about certain dependence of the obligatory on what will be done and we can get a better understanding of which future potential choice opportunities are relevant to our present obligation. And there is every reason to be optimistic that more focused attention than can be provided in a handbook article will yield still more results.

2.3 Further Issues

We close our discussion in this section by considering some further features of modal deontic logic that are relevant to ethical theory.

2.3.1 Decision Rules in Modal Semantics

We start with a certain high level difference between these theories. This difference is relevant to the treatment of actualism/possibilism debate. For example, we touched on it in comparing the theories of Goble and Hansson (§2.1.2). But the difference is so general that it has effects that percolate down to the verdicts that the theories give about many kinds of cases. The heart of the issue is a certain push and pull between two structural features of these theories. The first features concerns which decision rule to adopt for selecting among acts, outcomes, etc. and how fine-grained that decision rule should be. The second structural feature concerns which possibilities are to count as the “relevant outcomes” for determining what is obligatory.

We begin our discussion of this by returning to Hansson’s theory. Above, we only discussed how Hansson’s theory ordered propositions. But Hansson also mentions how relations among proposition may be related to relations among world. These ideas are summarized in Table 2 (from Hansson 2013: 492)) where *max* is a function that takes us from a proposition to the best world in which that proposition is true and *min* is a function that takes us from a proposition to the worst world in which it is true.

<i>Maximin preferences:</i>	
$p \geq_i q$ iff	$\min(p) \geq \min(q)$
<i>Maximax preferences:</i>	
$p \geq_x q$ iff	$\max(p) \geq \max(q)$
<i>Interval maximin preferences:</i>	
$p \geq_{ix} q$ iff	either $\min(p) > \min(q)$ or both $\min(p) \simeq \min(q)$ and $\max(p) \geq \max(q)$
<i>Interval maximin preferences:</i>	
$p \geq_{ix} q$ iff	either $\max(p) > \max(q)$ or both $\max(p) \simeq \max(q)$ and $\min(p) \geq \min(q)$
<i>Doubly maximizing preferences:</i>	
$p \geq_{\ddagger} q$ iff	$\max(p) \geq \max(q)$ and $\min(p) \geq \min(q)$

Table 2: Possible relations between orderings on propositions and orderings on worlds in Hansson’s theory

One way to see what is interesting about these connections is that they enforce a certain kind of decision rule or, more accurately, an elimination or negation selection rule. To see this, recall that $O(p)$ just in case $f \geq \neg p$. So whether $\neg p$ is eliminated and p is adopted is related to the underlying ordering of worlds. To give one example, consider *Maximin preferences*. If the worst $\neg p$ worlds are worse than or equal to the worst f worlds, then $f \geq \neg p$ and thus $F(\neg p)$ and $O(p)$. So $\neg p$ can be eliminated and p can be adopted. So the maximin criteria gives us an elimination/negation selection rule.

Hansson, for his part, prefers *Doubly maximizing preferences* because this insures that the relation among propositions is interpolative. However formally appealing this might be, in many settings the resulting ordering on propositions will be incomplete or tied as the decision rules need not strictly order all incompatible propositions. In cases like this (e.g., any pair of gambles where one has a greater net pay out if you win and also greater net loss if you lose), this will result in both options being permissible even though intuitively they may not both be permissible (e.g., when one option has greater expected value).

Similar points can be made about the other options in Table 2. And indeed, they extend even to the other theories that we have considered. Let’s take a look at this.

To simplify matters, let us consider how these different theories determine whether it is obligatory that p or obligatory that $\neg p$ or neither. SDL does this by simply checking a set of worlds to see if they unanimously say p , or unanimously say $\neg p$, or give no unanimous answer. As a decision rule then, SDL implicitly suggest that if p is obligatory, then it passes a maxi-max test (the best p -worlds are better than the best $\neg p$ -world). The other theories involve more explicit comparisons. Goble adopts a Pareto-like rule so that p is ranked ahead of $\neg p$ just in case every p -world is as good as every $\neg p$ -world and at least one is better. Horty’s theory is harder to succinctly describe as it involves ranking on specific histories, rankings on propositions (sets of histories), and rankings on

acts. But Harty also adopt certain qualitative rules. He adopts a Pareto-like rule for determining the ranking on propositions and then a state-wise dominance rule to determine rankings on acts.

What all of these theories have in common is that they adopt qualitative decision rules that do not appeal to any ideas about the likelihood of outcomes. What's more, many of them also do not appeal to the idea that the goodness of possibilities is numerically measurable. This makes it difficult for these theories to get the result that something is obligatory in any case in which the spread of relevant outcomes for an act is sufficiently wide and diverse in terms of goodness as compared to their competing outcomes.

Strikingly, none of these proposals adopt rules of the sort familiar from expected value decision theory, rules that make use of a weighted average determined by the numerically measurable values weighted by probabilities. Of course, in order for expected value approach to work, we need some sensible way to numerically measure the values and some way to interpret what a probability is. And whatever procedure we adopt for this we need to make sense of how to assign these quantities to worlds, propositions, and acts, and why these quantities can be sensibly multiplied and added. There is a rich philosophical literature on the "additivity" of value and a rich formal literature on the measurability of quantities that can contribute to this.²⁷ In certain cases, the issue is well-understood. For example, if we interpret our ordering as telling us an agent's subjective preferences, decision theorists have representation theorems that tell us what conditions that ordering must satisfy in order for us to make sense of taking a weighted average.²⁸ Other cases are less well-understood, however.

For reasons like this and others, the theories that we have discussed have avoid using these kinds of expected value decisions rules. Here is what Harty and Hanssons have to say:

The particular ordering that results from comparison of expected value relies, however on a kind of probabilistic information concerning outcomes of actions that is often either unavailable or meaningless; and this is true especially in situations in which the outcome resulting from an agent's action may depend, not simply on a roll of the dice, but on the independent choice of another free agent. (Harty 2001: 59-60)

However, this is not suitable explication of preferences for the purpose of deontic logic. Suppose that you are deliberating on whether to keep an extra income for yourself (*s*) or donate it to a charity (*c*). The probability that you will do one or the other [...] should not influence your choice since if it did, then you would not really treat both alternatives as fully open. [...] In addition, there is a

²⁷See Krantz et al. 2007 for a classic introduction to the topic of measurement

²⁸See Ramsey 1931 [1923] and Jeffrey 1990 for classic discussions; see Meachem and Weisberg 2011 for recent philosophical reflection on how these results are to be interpreted.

counter-argument of a more formal nature: Weighted-average preferences are not in general interpolative and they do not even satisfy the highly plausible property that if $p \simeq q$, then $p \simeq p \vee q$, where \simeq denotes indifference (Hansson 2013: 491)

Here is not the place to assess this dispute between expected value decision theories and the theories discussed in this section. But this topic is worth further investigation by moral philosophers and deontic logicians.²⁹

That said, there may in fact be a way of avoiding these difficulties. The difficulties arise because we are connecting a proposition's or act's place in the ranking with the places of various worlds where the proposition is true or the act is performed. The decision rules that we have looked at are either not strong enough to give us rankings that are as discriminating as we'd like them to be or are like the weighted average rule which requires stronger commitments than we might like. The trouble, at root, stems from the fact that there are typically many worlds in which the proposition is true or the act is performed and then having to use this multitude of worlds to determine the place of the proposition or act in the ranking.

One attractive feature of views that appeal to certain counterfactuals to determine the outcomes that are relevant to whether an act is obligatory is that they only require us to consider a (typically proper) subset of the worlds where the proposition is true or the act is performed. According to these views, the world or worlds that are relevant are the ones that would result if you did the act. According to certain standard analyses of counterfactuals, this often means there is a single world that we have to look at. Or, in any case, a small family of worlds that are as close as possible to the actual world. In this setting, it may be that any of the above decision rules will give us orderings that are sufficiently discriminating. Let us compare then how some of our theories determine what the relevant outcomes are.

We already saw Hansson's theory considers the total space of possibilities and this, of course, will also be a feature of SDL. Goble's theory, Horty's theory, and traditional forms of consequentialism, on the other hand, require modal space to have a rich enough structure to accommodate counterfactuals.

We can investigate this by looking at some properties of the closeness-orderings used in the standard semantics for counterfactuals.³⁰ Recall that a standard gloss on how one evaluates the truth of the counterfactual $p \Box \rightarrow q$ at a world w is that one finds the closest p -worlds to w and checks whether q is true at throughout those worlds.³¹ But there are three related proposals presented

²⁹I do not mean to suggest that these issues have never been considered. Indeed, Goble in Goble 1996 develops an expected value consequentialist approach. I do not discuss the details of this paper partially for reasons of space. But more importantly, as I see it, the question of how to integrate decision-theoretic ideas related to values and probabilities with deontic logic is wide-open even though some initial forays have been made.

³⁰See Lewis 2001 [1973] for the canonical treatment.

³¹Or at least this is how we interpret the counterfactual in settings where we assume the limit assumption is satisfied.

in Table 3 about the structure of this similarity ordering that influence what kinds of theorems hold in semantics of this sort.

Properties of the closeness-ordering	Theorems
WEAK CENTERING if p is true at w , then w is one of the closest p -worlds to w	COUNTERFACTUAL MODUS PONENS $(p \Box \rightarrow q) \rightarrow (p \rightarrow q)$
STRONG CENTERING if p is true at w , then w is the unique closest p -world to w	CONJUNCTION CONDITIONALIZATION $(p \wedge q) \rightarrow (p \Box \rightarrow q)$
UNIQUENESS there is a w' that is the unique closest p -world to w	CONDITIONAL EXCLUDED MIDDLE $(p \Box \rightarrow q) \vee (p \Box \rightarrow \neg q)$

Table 3: Correspondence between properties of the closeness-ordering and theorems about counterfactuals

As the interested reader can check for themselves if they work through the details, Horty’s theory accepts only WEAK CENTERING, Goble’s theory in addition accepts STRONG CENTERING, and standard forms of consequentialism all three claims. How is this relevant to what we obligated to do? According to these theories, we determine whether we are obligated to do something in part by the value of various worlds. But which worlds we look at is determined by the similarity ordering. In all of these theories, in cases where I end up performing some act, the value of the actual world is relevant to the acts status. In Horty’s theory, the actual world is one of perhaps several worlds values who is relevant. In Goble’s theory and traditional consequentialism, it is the only world whose value is relevant. In Horty and Goble’s theory, we compare the value an act that you will perform in this world with the value of an incompatible acts, there are a number of worlds whose value is relevant. But in traditional consequentialism, we compare another single world.

As we know, SDL and Hansson’s theory differ from this. But Hansson’s theory “looks” similar to how Horty’s theory works in that a non-singleton set of worlds determine the value of p and it is compared with the value of other acts that are determined by another non-singleton set of worlds. But, in Horty’s theory these sets are typically proper subsets of the set of worlds where p is true while in Hansson’s theory the set of worlds is all the worlds in which p is true.

The upshot of all of this is that our theories give us different “thinning outs” of the space of relevant outcomes. Standard consequentialism is the most extreme. On virtue of this is that there is no longer a wide and diverse set of outcomes that will lead to incompleteness or ties in the rankings of acts. Indeed, given that we are comparing two worlds as the consequentialist does, all of the decision rules in Table 2 give the same verdict about how the worlds rank. As such, consequentialism will be decided about what is obligatory in any case in which the world’s are not exactly as valuable as one another. This makes it so even qualitative decision rules give a rich set of obligations.

It is harder to say in any systematic way what the result of Horty's and Goble's theory are, but we know they represent a thinning out of what the relevant outcomes are and, as such, may and likely will result in more verdicts about what is obligatory than other theories.

It is not my aim to argue that these approaches that use counterfactuals are correct and other approaches are not. All I wish to do is to bring into focus the underlying push and pull between the two structural assumptions that I identified. A theory's decision rule together the modal structure it posits determines how discriminating the ordering on acts is. On one extreme, we have extremely strict uniqueness validating modal structure posited by consequentialism which gives us a very discriminating ordering on acts according to to any decision rule. On the other extreme, we have the unrestricted modal structure together with expected value decision rule which also gives us a very discriminating ordering on acts. In between we have decision rules like maximin, Pareto, etc. and modal structures like weak centering, strong centering, etc. Moral philosophers and logicians alike can fruitfully interact by considering which of these is appropriate for which applications.

In moral philosophy some of these issues have already been explored to some extent. For example, many philosophers believe there are so called "objective" obligations which do not depend on your information but just the facts and so-called "subjective" obligations that do depend on your information.³² For some of them the objective obligations are modelled correctly by any decision rule and uniqueness (traditional consequentialist). But subjective obligation is best modelled by the expected value rule and no modal structure. These different obligations are supposed to correspond to different theoretical roles in our practice. The first is involved in standards of correctness and giving of advice. The second is involved in assessments of rationality and portioning blame. There is also an open question of which of these ought to be taken to be of central interest.³³

There has been considerably less work done about the intermediate cases and what their role might be. This issue about the interaction between these two structural features would benefit from deeper study than I have presented here. There are important formal issues about how to make precise (and whether it can be made precise) some of the comparisons between theories that I have made here. And there are important conceptual issues about how to understand the role of different kinds of structure and whether any of them deserves to be privileged.³⁴

³²See Schroeder 2018: §IIB for a recent discussion of the theoretical role of this distinction. But the distinction itself has played a role in moral philosophy for decades (though sometimes it has gone under other names).

³³See Lord 2017 and Lord 2018b for a helpful presentation of the state of play as well as a defense of a distinctive view about what is of central interest.

³⁴There are still further questions about which comparison class is relevant for determining obligations, the relationship between the values of acts and the values of outcomes, and which outcome in which an act occurs is relevant for determine its deontic status. These questions are preliminary explored in Nair 2020a.

2.3.2 Other Interpretations

The simplest interpretation of the deontic logics that we have looked at are value-based interpretations. We can understand the set of worlds that witness the truth of claims about obligation in SDL as the worlds that are best; we can understand the underlying orderings involved in Goble, Hansson, and Horty's theory as goodness or value-based orderings. While natural, these interpretations are not inevitable as we have seen. And this is good news because in moral philosophy it is, of course, controversial whether anything like a value-based framework is correct. Let us take a look at what some other possible interpretations might be.

Perhaps, the most straight forward alternative interpretation are ones that do not stray far from the value-based approach. So for example, views which take preferences or idealized preferences to determine what is right and wrong could easily provide alternative interpretations of the orderings that our theories make use of. Of course, as with the interpretation in terms of values, one must check that the logical properties of the ordering are ones that are compatible with the interpretation.

This is not trivial. For example, all of the theories that we have looked at assume that the ordering among worlds is connected in the sense that any two worlds or propositions are ranked with respect to one another by the ordering (where we allow this may means some worlds are tied). This amounts to the claim that value comparison can always be made or comparison according to preference can always be made. While plausible, this controversial as some believe that there are incommensurable goods.³⁵ Whatever interpretation we choose, we must take a stand on this issue or provide a way of relaxing the framework to allow for unconnected orderings. Similar issues arise concerning other properties of the ordering such as transitivity.³⁶

But putting these very general difficulties aside, there seems to be no particular reason to be suspicious of interpretations of these logics in terms of preferences. A more interesting question however is whether moral theories that are very different are compatible with the theories that we have discussed here. For example, how can leading deontological theories like Kantianism, contractualism, or Rossian pluralism interpret the logics presented here? And how can theories in the virtue ethics tradition interpret the logics presented here?

The best supported but disappointing answer to this question is that they may be compatible with these theories. For each theory, there is a somewhat trivial way of trying to show that there is license for optimism that it is compatible with the frameworks that we have been discussing. One simply takes what is permitted according to each theory and ranks it higher or ranks the worlds in which it occurs higher than that which is not permitted. If this procedure works (it would requires some care to show it works and we already saw in §2.1.1 some grounds to be cautious about this), then we can rest assured that these moral

³⁵See Chang 1997: Introduction (and the other essays in the volume) for a useful survey of the issues raised by incommensurability.

³⁶See Tempkin 2015 for a book length discussion of problems for the transitivity of goodness.

theories are compatible with the deontic logics that we have been discussing.

But there would be very little interesting about this result. The formal theories would not give us much insight into the underlying interest or structure of the moral theory. And the moral theory would not give us much of an explanation for why the particular formal structures that we are using are sensible. For example, what is interesting about contractualism of the sort Thomas Scanlon has developed is that there is a test for reasonable rejection for rules and the test does not allow for interpersonal aggregation (Scanlon 1998). What a good formal theory would do is give a way of representing rules, representing perspective of those who may have reasons to reject or accept such rules, and then define from these a logic that allows us to study what is obligatory according to this theory. Formally, this would be an interesting task and conceptually it would offer a more precise model of contractualist reasoning that would be helpful in moral philosophy. But none of this is provided by the theories that result from the easy procedure that I described.

I believe this lesson generalizes. While non-value- or preference-based theories may be compatible with the modal theories of the sort that we have been discussing, there is very little of interest in this fact. The distinctive aspect of these moral theories are not represented in any especially illuminating way by the formal structures and the moral theory does little by way of explaining why the formal structure is a sensible one for modeling obligation.

So one important area of research is to develop formal models that are better fit for these deontological theories and theories from virtue ethics. For most of these moral theories, very little work has been done.³⁷ I suspect this is partially because the moral philosophers who have worked on these theories have been less interested in formally developing their theories and more interested in explaining why they are attractive alternatives to consequentialist theories. That said, the theories are discussed in informally precise prose by many leading moral philosophers and are ripe for exploration by the more formally inclined.

That said, a notable outlier are Rossian pluralist theories and the particularistic theories that arose from Ross's insights. The theories have received considerably more attention in deontic logic. The next section introduces these deontic logics.

3 Particularism

Perhaps one of the most fruitful interactions between deontic logicians and moral philosophers concerns particularism. Particularism is, very roughly, the view that there is no codifiable set of principles or norms governing the moral.³⁸ Morality is, as it is sometimes said, "shapeless". We begin by looking at the motivations for this view (§3.1). We then present a formal system that allows

³⁷See however Rechenauer and Roy 2014 for a discussion of contractualist approaches and Braham and Hees 2015 and Lindner and Bentzen 2018 for discussion of Kantian approaches.

³⁸Leading discussions of particularism include Dancy 2004, Hooker and Little 2001, McKeever and Ridge 2006.

us to model many of the phenomena that motivate this view (§3.2). We close by assessing how satisfied we should be with this model (§3.3, §3.4).

3.1 Motivations

For almost any putative principle one can come up with cases where it seems not to be in force. This was famously noted by Henry Sidgwick in his critique of the intuitional method in *The Methods of Ethics* (Sidgwick 1981 [1907]: Book 3). Sidgwick painstakingly considers various principles that, for example, forbid lying or enjoin you to keep your promises and argues that they do not apply correctly in certain cases. And Sidgwick argues that various modifications of these principles are subject to problems of their own as well.

Sidgwick concluded from this that the utilitarian principle is the best explanation of why these principles give us the correct results in some cases and fail in other case.³⁹ But philosophers since Sidgwick have seen utilitarianism as subject to similar problems. G.E. Moore thought the culprit was the simplistic hedonist axiology implicit in Sidgwick's utilitarianism and opted for a perfectionist form of consequentialism instead (Moore 1962 [1903]: especially ch. 3 and 6). But W. D. Ross suggested that even this was not enough. Consequentialism as such cannot accommodate all the cases. Here is one of his examples:

If I have promised to meet a friend at a particular time for some trivial purpose, I should certainly think myself justified in breaking my engagement if by doing so I could prevent a serious accident or bring relief to the victims of one. And the supporters of the view we are examining hold that my thinking so is due to my thinking that I shall bring more good into existence by the one action than by the other. A different account may, however, be given of the matter, an account which will, I believe, show itself to be the true one. It may be said that besides the duty of fulfilling promises I have and recognize a duty of relieving distress, and that when I think it right to do the latter at the cost of not doing the former, it is not because I think I shall produce more good thereby but because I think it the duty which is in the circumstances more of a duty. This account surely corresponds much more closely with what we really think in such a situation. If, so far as I can see, I could bring equal amounts of good into being by fulfilling my promise and by helping someone to whom I had made no promise, I should not hesitate to regard the former as my duty. (W. D. Ross 1930: 18)

As the example shows, Ross's preferred view is that there are a number of *prima facie* duties. In a given case, one or several of them may apply. What we ought to do is determined by which of the duties is stronger in this context.

³⁹Sidgwick also argued that the few intuitional principles that withstand scrutiny, in fact, entail utilitarianism.

But Ross did not think that we could provide any once-and-for-all list that ranks the duties according to strength such that one ought to do what is suggested by one's strongest applicable duty. This is because Ross believed that the relative strength of duties can vary from case-to-case. He writes: "But no act is ever, in virtue of falling under some general description, necessarily actually right; its rightness depends on its whole nature and not on any element in it." (W. D. Ross 1930: 33)

Every act therefore, viewed in some aspects, will be prima facie right, and viewed in others, prima facie wrong, and right acts can be distinguished from wrong acts only as being those which, of all those possible for the agent in the circumstances, have the greatest balance of prima facie rightness, in those respects in which they are prima facie right, over their prima facie wrongness, in those respects in which they are prima facie wrong—prima facie rightness and wrongness being understood in the sense previously explained. For the estimation of the comparative stringency of these prima facie obligations no general rules can, so far as I can see, be laid down. (ibid.: 41)

So for Ross, we can state some moral principles that tell us certain considerations (e.g., promise keeping) have force in any case where they apply (e.g., any case in which an agent has made a promise, keeping the promise is prima facie right). But we cannot give a procedure or rule for going from this list of contributing factor to what we ought to do overall: while it is possible to give a principled account of *what* factors contribute to rightness, it is not possible to give a principled account of *how* they contribute. This then is one grounds for the particularist thought.

But there is an even deeper sense in which a particularist believe the morality cannot be understood in terms of principles. They reject even Ross's thought that we can give a principled account of what factors contribute to rightness. They argue for this by showing that many putative factors that contribute to rightness in one context are *undercut* in other contexts so that they have no force at all or are *intensified* so that they may more strongly contribute in some contexts or are *attenuated* so that they more weakly contribute in other contexts. Here are some examples that have been offered:

we might point out that in some contexts the fact that something is against the law is a reason not to do it, but in others it is a reason to do it (so as to protest, let us say, against the existence of a law governing an aspect of private life with which the law should not interfere) (Dancy 2017)

Not only is it possible to think of cases in which it is false that one ought not to lie, it is also possible to think of cases in which it is false that the fact that some action would involve lying is a reason not to do it. For example, if one is playing the game Bullshit, or the

game Diplomacy—both of which are sometimes said to be designed to involve lying, or at least to not discourage it. (Schroeder 2011a: 331)

This second form of defeat, or something very close to it, is discussed also in the literature on practical reasoning, where it is considered as part of the general topic of “exclusionary” reasons, first introduced by Joseph Raz [...]. Raz provides a number of examples to motivate the concept, but we consider here only the representative case of Colin, who must decide whether to send his son to a private school. We are to imagine that there are various reasons pro and con. On one hand, the school will provide an excellent education for Colin’s son, as well as an opportunity to meet a more varied group of friends; on the other hand, the tuition is high, and Colin is concerned that a decision to send his own son to a private school might serve to undermine support for public education more generally.

However, Raz asks us to imagine also that, in addition to these ordinary reasons pro and con, Colin has promised his wife that, in all decisions regarding the education of his son, he will consider only those reasons that bear directly on his son’s interests. And this promise, Raz believes, cannot properly be viewed as just another one of the ordinary reasons for sending his son to the private school, like the fact that the school provides a good education. It must be viewed, instead, as a reason of an entirely different sort—a “second-order” reason for excluding from consideration all those ordinary, or “first-order” reasons that do not bear on the interests of Colin’s son. (Horty 2007: 14-5)

Beginning with the practical domain, imagine that I have borrowed a book from you. In most situations, the fact that I have borrowed a book from you would give me a reason to return it to you. But suppose I discover that the book I borrowed is one you had previously stolen from the library. In that context, according to Dancy, the fact that I borrowed the book from you no longer functions as a reason to return it to you; in fact, I no longer have any reason to return it to you at all. (ibid.: 20)

The conclusion that we are invited to draw from this is that not only is there no principled account of how various factors contribute to determining what we ought to do, there is not principled account of what the factors are that contribute. What we ought to do is too situationally flexible to be usefully modelled by any once-and-for-all theory.

Instead, the best one can say is that the wise agent sees the situation for what it is and can appreciate the relevant force of the considerations in that context:

In this respect the judgement as to the rightness of a particular act is just like the judgement as to the beauty of a particular natural

object or work of art. A poem is, for instance, in respect of certain qualities beautiful and in respect of certain others not beautiful; and our judgement as to the degree of beauty it possesses on the whole is never reached by logical reasoning from the apprehension of its particular beauties or particular defects. Both in this and in the moral case we have more or less probable opinions which are not logically justified conclusions from the general principles that are recognized as self-evident. (W. D. Ross 1930: 31)

So what we are trying to do is to establish what reasons are present in the case before us. The ability to do this is a sophisticated one, which children develop as they grow up; presumably it is one for which some form of training is virtually essential. If we want to know what it is like to have that ability, we could start by asking what it is that competent judges bring to a new case. [...] The particularist will say here that our skills in reason-discernment are not rule-based, meaning by this that we do not extract rules for the operation of reason-giving features from the cases we have come across and then try to subsume new cases under those rules. [...]

[...] Particularists conceive of the knowledge brought to a new case as much more like knowledge-how than like knowledge-that. That is, it is a skill of discernment, not knowledge of a set of true general propositions discovered by thinking about previous cases and applied somehow to new ones. [...] The competent judge is not the person in command of general truths about the behaviour of reasons, all extracted from experience. She is a person who can tell a difference when she comes across it. (Dancy 2004: 142-3)

So based on reflection on simple examples like this, particularist argue for the conclusion that morality is uncodifiable or shapeless.

But one of the interesting development in recent years is a number of formal systems that can be used to model these examples that motivate particularism. Here I will focus on presenting the ideas of John Horty as they appear in his *Reasons as Defaults* as this theory, in my view, has had the most influence (Horty 2012). As the title of Horty's book suggests, his theory makes use of the notion of reason. Reasons are considerations that count in favor or against some action. Horty's system is designed to allow that the strength of reasons can vary from case-to-case. And it is designed to even allow that what is a reason in one case can fail to be a reason at all in another case. But these changes in strengths of reasons and what is a reason follow from precise principles given by the system. Let's see how this work.

3.2 The Formal Theory

Horty uses of a formal system known as default logic.⁴⁰ Default logics were developed originally to understand inferences like the one where we conclude ‘Tweety flies’ from ‘Birds fly’ and ‘Tweety is a bird’. But Horty proposes to use this formalism to model reasons and what we ought to do.

To begin, we have the notion of a default rule or simply a default which we can write as ‘ $A \rightarrow B$ ’ to mean that once A has been established one may conclude by default B . We will call A the *premise* of the default and B the *conclusion* of the default. Though this formalism looks like a conditional in ordinary logic, it is not to be understood this way. Instead, it will be used as a kind of principle that encodes which reasons we accept. So for example suppose we think that if John were to promise Mary to help her move that would be a reason for John to help Mary move. We can encode this as the default: John promised to help Mary move \rightarrow John helps Mary move. So if we have the information that John did promise to help Mary move, we can conclude by default ‘John helps Mary move’. And in Horty’s development of these ideas, what we ought to do is what we can conclude by default. So the theory tells us that given the default and the information that John promised Mary, John ought to help Mary move.

We develop these ideas more systematically by assuming that we have set of sentences, \mathcal{W} , that is used to represent the information that we accept and have a set of defaults, \mathcal{D} , that represents the “reasons”-principles that we accept. We also assume there is an ordering on this set of defaults, $<$, that tells us about the strength of the reasons. While there are many potential properties one might think this relation has, we only assume that it is transitive (i.e., if $\delta < \delta'$ and $\delta' < \delta''$, then $\delta < \delta''$ for any $\delta, \delta', \delta'' \in \mathcal{D}$) and irreflexive (i.e., $\delta \not< \delta$ for any $\delta \in \mathcal{D}$). Though for certain purposes it may be natural to assume the ordering is connected (i.e., $\delta < \delta'$ or $\delta' < \delta$ for any $\delta, \delta' \in \mathcal{D}$), we do not assume this holds in the general case. So defaults can be tied or can be incomparable. We collect these items together in an ordered 3-tuple $\langle \mathcal{W}, \mathcal{D}, < \rangle$ and call it a *fixed priority default theory*.

The basic idea will be that given our information, \mathcal{W} , and the priority ordering, $<$, among defaults, \mathcal{D} , some (often proper) subset of \mathcal{D} will be the ones that tell us what ought to be done. This is because some defaults may not apply to a case and some defaults will conflict with defaults that are stronger than them. But we only want to pay attention to the applicable defaults that are not in conflict with strong defaults.

We can call a subset of our set of defaults a *scenario*. So our task will be to define which scenarios are the ones that contain the defaults that tell us what we ought to do. Once we identify that set we will, be able to determine what we ought to do.

We build up to this by noting three features had by defaults that tell you what you ought to do. First, defaults that tell you what you ought to do in a

⁴⁰See Reiter 1980 for the seminal presentation. See Makinson 2005: ch. 4 for an approachable (albeit slightly idiosyncratic) contemporary introduction.

case need to be actually applicable or as we will call it *triggered*. The intuitive idea is that a triggered default is a default that actually represents a reason. For example, if John has not promised Mary to help her move, John’s promise is not a reason to help her. But even in cases where John has not actually made the promise, we accept that if he were to make such a promise, it would be a reason. And defaults are just these kinds of “reason”-principles.⁴¹ So in this example, we have a default concerning John’s promises but it is not triggered because John has not made a promise.

We can formalize this in two steps. First, we introduce a pair of functions *Premise* and *Conclusion* that respectively return the premise of a default and the conclusion of a default. So for the default δ^* : John promises to help Mary move \rightarrow John helps Mary move, $Premise(\delta^*) =$ John promises to help Mary move and $Conclusion(\delta^*) =$ John helps Mary move.

We lift this definition to sets of defaults in the natural way:

for any scenario $\mathcal{S} \subseteq \mathcal{D}$,

$$Premise(\mathcal{S}) = \{Premise(\delta) \mid \delta \in \mathcal{S}\}$$

$$Conclusion(\mathcal{S}) = \{Conclusion(\delta) \mid \delta \in \mathcal{S}\}$$

Next we use these to define the set of triggered defaults are for an arbitrary scenario based on on a theory:

$$Triggered_{\langle \mathcal{W}, \mathcal{D}, < \rangle}(\mathcal{S}) = \{\delta \in \mathcal{D} \mid \mathcal{W} \cup Conclusion(\mathcal{S}) \vdash Premise(\delta)\}$$

So for example in situation where it is not part of our hard information that John promised Mary, relevant default would not be triggered but in a situation in which it is part of our hard information, it would be triggered.

More generally, we can say that *A is a reason for B* just in case $A \rightarrow B$ is a triggered default. And interestingly, this conception of triggering also correctly allows for the “chaining of defaults”. So for example, in a setting where $\mathcal{W} = \{A\}$ and $\mathcal{D} = \{A \rightarrow B, B \rightarrow C\}$, we have it that $Triggered_{\langle \mathcal{W}, \mathcal{D}, < \rangle}(\mathcal{D}) = \{A \rightarrow B, B \rightarrow C\}$. Intuitively this is because *A* triggers the first default and the first default then triggers the second.

The next concept to introduce is the concept of a conflicting defaults. This is easy to define:

$$Conflicted_{\langle \mathcal{W}, \mathcal{D}, < \rangle}(\mathcal{S}) = \{\delta \in \mathcal{D} \mid \mathcal{W} \cup Conclusion(\mathcal{S}) \vdash \neg Conclusion(\delta)\}$$

In a set of defaults some might conflict with others, $Conflicted_{\langle \mathcal{W}, \mathcal{D}, < \rangle}$ collects these conflicted default together.

The final concept is the concept of one default defeating another. This is to be understood in terms of two things. First, for one default to defeat another, they must be in competition in the sense that one cannot obey both at the same time. So they must be conflicting defaults. Second, to lose the conflict is for the default to be worse according to our ordering. This suggests that:

⁴¹This shows defaults are not reasons but “reason”-principles. What is a “reason”-principle? We discuss this issue in detail in §3.3

$$\begin{aligned}
\text{Defeated}_{\langle \mathcal{W}, \mathcal{D}, < \rangle}(\mathcal{S}) &= \{ \delta \in \mathcal{D} \mid \text{there is a } \delta' \in \text{Triggered}_{\langle \mathcal{W}, \mathcal{D}, < \rangle}(\mathcal{S}) \\
&\text{such that (i) } \delta < \delta' \text{ and (ii) } \text{Conclusion}(\delta') \vdash \neg \text{Conclusion}(\delta) \}
\end{aligned}$$

As it turns out this definition is not fully adequate for cases in which there are collections of default which can chain together to create defeating relations. There are a number of known proposal for how to deal with this issue, but I set these aside for now because our main ideas can be illustrated with out these complication.⁴²

We make use of these three concepts to define a new operator *Binding* that takes us to the defaults that are triggered, unconflicted, and undefeated:

$$\begin{aligned}
\text{Binding}_{\langle \mathcal{W}, \mathcal{D}, < \rangle}(\mathcal{S}) &= \{ \delta \in \mathcal{D} : \delta \in \text{Triggered}_{\langle \mathcal{W}, \mathcal{D}, < \rangle}(\mathcal{S}) \text{ and } \delta \notin \\
&\text{Conflicted}_{\langle \mathcal{W}, \mathcal{D}, < \rangle}(\mathcal{S}) \text{ and } \delta \notin \text{Defeated}_{\langle \mathcal{W}, \mathcal{D}, < \rangle}(\mathcal{S}) \}
\end{aligned}$$

We will use the label *stable* for when a particular scenario based on a theory is one that we need to pay attention to when determining what we ought to do. We can say:

$$\mathcal{S} \text{ is a stable scenario based on } \langle \mathcal{W}, \mathcal{D}, < \rangle \text{ iff } \mathcal{S} = \text{Binding}_{\langle \mathcal{W}, \mathcal{D}, < \rangle}(\mathcal{S})$$

This is a so-called “fixed-point” definition. If \mathcal{S} is stable, the idea is that it is a fixed point of the *Binding* operator. Intuitively if the scenario you accept is stable, you have no reason to kick any defaults out or add any more defaults in. In practice, one makes use of this definition by working through each possible set of defaults and checking to see if all of the defaults in it are triggered, unconflicted, and undefeated.⁴³

Once we have identified a stable scenario, \mathcal{S} , we say an *extension*, \mathcal{E} is what follows from the conclusion of \mathcal{S} together with our hard information, $\mathcal{E} = \{X : \mathcal{W} \cup \text{Conclusions}(\mathcal{S}) \vdash X\}$. If there is just one stable scenario for a given theory, we say that what we ought to do is anything that is in the extension of that scenario. But as it happens, there can be multiple stable scenarios.

Plausibly enough, this happens when there are ties or incompatibilities among conflicting defaults. In such a case, we have at least two options. We can either say that what we ought to do is anything that is in the extension of *some* stable scenario or we can say that what we ought to do is anything that is in the extensions of *all* of the stable scenarios. It turns out these correspond to deontic logics that allow moral conflicts and ones that do not allow moral conflicts.

This is already on interesting application of Horty’ theory: it allows us to study and compare different approaches to putative moral dilemmas. But we will pass over this interesting application because it is orthogonal to our discussion of particularism and because it is discussed in detail elsewhere in this

⁴²See Horty 2012: ch. 8 and the citations therein for discussion.

⁴³This is not quite Horty’s official view. He instead adopts a slightly more complicated notion of a “proper” scenario, see *ibid.*: §A.1. There are also certain more general complications related to the possibility of there being no stable scenarios, see *ibid.*: §1.3.2 for a preliminary presentation and Delgrande, Schaub, and W. K. Jackson 1994 and Antonelli 1999 for further discussion.

handbook (Goble 2013). We will therefore assume for simplicity the no-conflict version of Horty’s theory that takes what we ought to do in case of multiple extension to be what is in the extension of every stable scenario.

Let us illustrate how Horty’s approach by looking at an example.⁴⁴ Suppose, then, that you have a choice of whether to send your son to School 1 or School 2. When it comes to the cost, School 1 is favored. But when it comes to the quality of education, School 2 is favored. Further, we may suppose the cost provides a stronger reason than the education. So we have the defaults $Cheap(x) \rightarrow Attend(x)$ and $Education(x) \rightarrow Attend(x)$ that represent the idea that a given school being cheap is a reason to attend it and a given school providing high quality education is a reason to attend that school. To simplify things, we instantiate these variables so we have the following defaults:

$$\begin{aligned}\delta_1 &: Cheap(s_1) \rightarrow Attend(s_1), \\ \delta_2 &: Cheap(s_2) \rightarrow Attend(s_2), \\ \delta_3 &: Education(s_1) \rightarrow Attend(s_1), \\ \delta_4 &: Education(s_2) \rightarrow Attend(s_2)\end{aligned}$$

where s_1 and s_2 are constants for School 1 and School 2 respectively. and we collect these defaults together so that $\mathcal{D} = \{\delta_1, \delta_2, \delta_3, \delta_4\}$. We also have the information that cost is more weighty than education so we have:

$$\begin{aligned}Education(s_1) \rightarrow Attend(s_1) &\sim Education(s_2) \rightarrow Attend(s_2) < \\ Cheap(s_1) \rightarrow Attend(s_1) &\sim Cheap(s_2) \rightarrow Attend(s_2)\end{aligned}$$

where we understand $\delta \sim \delta'$ to indicate δ and δ' have equal priority.⁴⁵

Finally, our background information includes $Education(s_2)$ and $Cheap(s_1)$ as well as implicitly the idea that going to one school precludes going to the other. It is easy to see $Education(s_1) \rightarrow Attend(s_1)$ and $Cheap(s_2) \rightarrow Attend(s_2)$ are not triggered. And it is also easy to see that $Education(s_2) \rightarrow Attend(s_2)$ is defeated by $Cheap(s_1) \rightarrow Attend(s_1)$. So the unique proper scenario includes only the default $Cheap(s_1) \rightarrow Attend(s_1)$. This give us an extension that is the closure of $\{Education(s_2), Cheap(s_1), Attend(s_1), \neg Attend(s_2)\}$ and so the result is that you ought send your son to attend School 1 rather than School 2.

This illustrates how the theory can handle conflicting reasons and allow for reasons to be stronger than one another. It can also allow that something can be a reason in one case but not in another simply because it doesn’t obtain in the other case. But we have not yet seen how the strengths of reasons can vary from one case to another. And we have not seen how a consideration can obtain in one case and provide a reason while the same consideration obtains in a distinct case but provides no reason. To accommodate these more complex dynamics, we need to introduce some further ideas from Horty’s theory.

⁴⁴This example and its development is, of course, inspired by the similar case in Raz 2002 [1975].

⁴⁵Technically, Horty’s system as he develops does not have the resources to distinguish equally priority defaults from incomparable defaults. But a simple generalization of Horty’s system which takes a “greater-than-or-equal-to” priority relation as basic and modifies the definitions in the obvious ways would allow for this. I assume this richer framework for simplicity of presentation here.

Conceptually, the key to accommodating these more complex dynamics is a certain picture of what the variability in the strengths of reasons and whether there are reasons amounts to. The picture is that this variability is explained by still further reasons. So for example, there may be a standing reason to not tell a lie, but in a context of playing the game Bullshit it may be that this reason is weaker or non-existent. The idea, then is, is that the fact that you are playing Bullshit provides a reason that *attenuates* the strength of the reason to not lie or simply *undercuts* it.

Formally this is accomplished by removing priorities from the structure of the theory and introducing them into the object language in such a way that we are able to reason about them. So we have a so-called variable priority theory which is simply an ordered pair of $\langle \mathcal{W}, \mathcal{D} \rangle$ of hard information and a set of defaults. But we now assume that we are working with a language that has symbols for an ordering and names for the defaults. We will use \prec as an object language symbol for $<$, \simeq as an object language symbol for \sim , and introduce a ‘ d ’ with a subscript as a name for each default $\delta \in \mathcal{D}$. We add as a further stipulation that every variable priority theory contains as part of its information in \mathcal{W} axioms stating that \prec is a transitive and irreflexive relation. It is easiest to see how these variable priority theories work by returning to our example and considering how it and other variants of it fare.

So to analyze our example, we introduce names into the object for each of our defaults. For each default, δ_i , we introduce the object language name d_i . Next one introduces a new default to the theory:

$$\delta_5: \top \rightarrow d_3 \simeq d_4 \prec d_1 \simeq d_2$$

δ_5 says that by default that cost is more important than education.

To determine which scenarios are stable for a variable priority theory is a bit more complex than for a fixed priority theory. First, to check whether a scenario is proper, one considers what claims about the priority ordering are in the extension of the scenario. So for example the scenario $\mathcal{S}_5 = \{\top \rightarrow d_3 \simeq d_4 \prec d_1 \simeq d_2\}$ has as its extension \mathcal{E}_5 which is the closure of $\{d_3 \simeq d_4 \prec d_1 \simeq d_2\}$ together with the hard information in \mathcal{W} . One then considers a fixed priority theory with \mathcal{W} and \mathcal{D} as before but with the ordering $<_5$ that matches (the object language claims about) the ordering given by \mathcal{E}_5 . We then check to see whether \mathcal{S}_5 is a proper scenario in the old sense of the fixed priority theory, $\langle \mathcal{W}, \mathcal{D}, <_5 \rangle$. That is, we considering whether $Binding_{\langle \mathcal{W}, \mathcal{D}, <_5 \rangle}(\mathcal{S}_5) = \mathcal{S}_5$

Let us work through this for our example. Since δ_2 (i.e., $Cheap(s_2) \rightarrow Attend(s_2)$) and δ_3 (i.e., $Education(s_1) \rightarrow Attend(s_1)$) are not triggered, they cannot be included in any proper scenario. On the other hand, δ_1 (i.e., $Cheap(s_1) \rightarrow Attend(s_1)$), δ_4 (i.e., $Education(s_2) \rightarrow Attend(s_2)$), and δ_5 (i.e., $\top \rightarrow d_3 \simeq d_4 \prec d_1 \simeq d_2$) are all triggered. Nonetheless, $\{\delta_1, \delta_4, \delta_5\}$ is not a proper scenario because both δ_1 and δ_4 are conflicted given the hard information that $Attend(s_1) \supset \neg Attend(s_2)$. $\mathcal{S}_{4,5} = \{Education(s_2) \rightarrow Attend(s_2), \top \rightarrow d_3 \simeq d_4 \prec d_1 \simeq d_2\}$ is also interestingly not a proper scenario. To see why, notice the derived priority of this scenario is the following

$$\begin{aligned} & Education(s_1) \rightarrow Attend(s_1) \sim_{4,5} Education(s_2) \rightarrow Attend(s_2) <_{4,5} \\ & Cheap(s_1) \rightarrow Attend(s_1) \sim_{4,5} Cheap(s_2) \rightarrow Attend(s_2) \end{aligned}$$

and the fixed priority theory based on it is $\langle \mathcal{W}, \mathcal{D}, <_{1,5} \rangle$. In this setting, $\mathcal{S}_{4,5}$ is not a proper scenario because δ_4 is defeated (by δ_1) in this scenario based on the fixed priority theory.

$\mathcal{S}_{1,5} = \{Cheap(s_1) \rightarrow Attend(s_1), \top \rightarrow d_3 \simeq d_4 \prec d_1 \simeq d_2\}$ however is a proper scenario because it is proper scenario the resulting fixed priority theory $\langle \mathcal{W}, \mathcal{D}, <_{1,5} \rangle$. Both defaults are triggered, unconflicted, and undefeated. The scenario does not include the triggered default δ_4 but this is acceptable because δ_4 is defeated. Finally, $\mathcal{S}_{1,5}$ is the unique proper scenario as any of the singleton sets would leave out a triggered, undefeated, and unconflicted default.

The resulting extension then tells us what we ought to do is the same as before. Obviously, in this case, all of the added complexity can seem pointless. But in cases where we want to reason about priorities, this extra complexity is worthwhile.

To illustrate this, let us add to the case that we are discussing. Suppose for example that you have also now promised your partner that in matters involving your child you will give more priority to education than cost. Now we might use P to represent that you made this promise to your partner and add the default:

$$\delta_6: P \rightarrow d_1 \simeq d_2 \prec d_3 \simeq d_4$$

. We introduce d_6 as the object language name for this default.

If we assume that P is part of our hard information, we now have to consider that all of δ_1 , δ_4 , δ_5 , and δ_6 , are triggered. This set of defaults is not a proper scenario because it is now conflicted in two ways. As before, δ_1 and δ_4 are conflicted, but we now also have conflict between δ_5 and δ_6 . So we know that our proper scenario will have a most one of each of these. And indeed, as is intuitive, there are exactly two proper scenarios: one consisting of δ_1 and δ_5 as before and another consisting of δ_4 and δ_6 . The other pairs are, as we might expect, ruled out. Consider for example $\mathcal{S}_{1,6} = \{\delta_1, \delta_6\}$. This scenario is not proper because δ_1 is defeated (by δ_4) in the fixed priority theory based on the derived ordering of this scenario.

We can elaborate this example still further if, for instance, we assume that we take our more specific promise to take priority over our initial views about the relative strength of cost and education. To do this, we need only add the following:

$$\delta_7: \top \rightarrow d_5 \prec d_6$$

If we add this, we now get as may be expected a unique proper scenario, $\mathcal{S}_{4,6,7} = \{\delta_4, \delta_6, \delta_7\}$. This scenario recommends sending one's child to the school that provides the best education, School 2.

Though we have been brief and quite informal, this, I hope, illustrates how we can model the dynamics of how the strengths of reasons can shift. The idea is that reasons themselves explains why in one context a reason can have one strength and in other context in can have a different strength. And Horty works

through a variety of other examples in his book (in greater detail than we can do here) and shows just how flexible this framework is at accommodating the variety of dynamics of reasons that have so impressed particularists.

This does not yet give us a way of modelling cases of undercutting. But there are two promising approaches to this. One approach is to take it that there is some place in the ordering that is a threshold in the sense that anything below that spot in the ordering is not a reason. In this setting, one says a default is triggered when it meets the old condition and is above the threshold. This allows us to preserve the idea that reasons are premises of triggered default.

An alternative approach does not make use of a threshold. Instead, the approach says that reasons are defaults that have a special property. We can introduce a predicate *Out* into our language and we can say that if a default's premise obtains and is *Out*, then it is not a reason but if it lacks this property of being *Out* and the premise obtains it is a reason. We then modify the definition of triggering so that a default is triggered when it meets the old definition and is not *Out*.

It should be easy to see how both approaches allow that a consideration can be a reason in one case but not in another. According to the first approach, this is because we can reason about whether a certain default is above or below the threshold. According to the second approach, this is due to our reasoning about whether a certain default has the property of being *Out*. As it turns out, there are interesting differences between these approaches. We will take a look at this in detail later on (§3.4.2).

But for now we pause to take stock of our basic results are. We have seen that a certain picture about the shapelessness of morality is motivated by the fact that there are a variety of relevant normative considerations that compete to determine what we ought to do, that the strength of these considerations vary from context to context, and that in certain contexts these considerations can even have no strength at all. These phenomena were used by particularist to motivate the idea that there can be no systematic account of the moral. But the formal theory that we have just described can accommodate each of these features with at least as much precision as typically non-particularistic moral theories. This shows one important contribution formal work in deontic logic makes to ethical theory: it undermines a certain argument for particularism. In the remainder of this section, I will further explore how satisfactory this response to the particularist is. And then I will turn to further issues that are raised by the present framework.

3.3 Limitations of the Response to Particularism

Horty's theory gives a systematic account of the shiftiness of reasons and what we ought to do by using a fixed background set of defaults where this set of defaults encodes information not just about which acts to do but also information about the properties of the defaults themselves. Shifts in what reasons there are and how strong reasons are are explained by shifts in hard information and how

these pieces of hard information interact with the fixed background of defaults.

But what are these defaults that form a fixed background against which changes are explained? To start, notice that defaults are not reasons. In Horty's theory, reasons are the premises of triggered defaults. This means (the premise of) a default can fail to be a reason by failing to be triggered and this occurs primarily by the premise failing to hold or by the default itself being undercut.

Horty himself describes defaults as generalizations or defeasible principles (cf. especially Horty 2012: 16-7 and 42-3). But what are these? One interpretation is the following:

ACTUAL NORMATIVE RELATION: defaults model an actual normative relation between two propositions

So this interpretation claims that if $P \rightarrow A$ is a default, then (whether P is true or A is true) there is a normative relation between P and A . And when certain facts about the world obtain (e.g., P), this together with the default explains why P is a reason for A .

This is to be contrasted with a second interpretation:

MODAL-NORMATIVE MIXED RELATION: defaults model under what condition it would be the case that the reason-relation holds between two proposition

This interpretation does not entail that any actual normative relation holds between the two propositions.

To see why accepting MODAL-NORMATIVE MIXED RELATION does not (at least without some further assumptions) require one to accept ACTUAL NORMATIVE RELATION, it perhaps helps to consider two examples by way of analogy. First, consider a theory that tells us under what conditions two people would become married. This theory can be informative and interesting for many reasons, but it need not be a theory according to which there is any marital or otherwise interesting actual relationship between people who would be married under certain non-actual conditions. The theory may simply be describing modal-marital mixed relation that is determined by embedding claims about the conditions sufficient for marriage under modal operators.

Or to take an example closer to the normative case, consider a theory that tells us under what conditions P would be believed by John to be a reason for A . No matter how good this theory is at predicting the facts about what normative relations hold in John's belief worlds under various conditions, it gives us no immediate grounds to think any normative relation actually holds between P and A .

So these two interpretations are distinct and the model is, at least initially, neutral about which interpretation is correct. As such, it is neutral on certain further metaphysical questions that the particularist may be interested in. For the particularism central concern may actually be whether there is any actual normative relation that can be systematically theorized about and that determines what we ought to do.

Why might this be so? If ACTUAL NORMATIVE RELATION is incorrect, then it is natural to think facts about what reasons there would be are not explained by (or wholly grounded in or reducible to) genuine normative relations. Now particularist tend to be non-reductivists who believe that normative relations cannot be explained by (or grounded in or reduced to) anything that we can use descriptive language to talk about. So such a particularist would see claims about what reasons there might be as unexplained by any principles, moral or descriptive. Instead, defaults are just free floating generalizations about what reasons there would be. This, it seems, corresponds to one sense in which the particularist may be interested in a “principle free” approach to ethics.

If this is right, then while Horty’s theory is enough to undermine a certain argument for particularism, it is not itself incompatible with particularism.

So the particularist may agree that their argument fails because it is possible to explain the dynamics of what we ought to do in the manner Horty does. But she may say that this explanation is in tension with particularist explanation only if we accept ACTUAL NORMATIVE RELATION. And this interpretation, she may argue, is less plausible than MODAL-NORMATIVE MIXED RELATION.

To start to show why MODAL-NORMATIVE MIXED RELATION is more plausible than ACTUAL NORMATIVE RELATION, the particularist may point out that the natural way to talk about the connections between the propositions paired by a non-triggered defaults (in English at least) is simply to embed a reason-claim in a counterfactual or other kind of alethic modal expression (e.g, if I were to promise Mary that I will give her \$100, then there would be a reason for me to give Mary \$100). Indeed, it is hard to think of a pretheoretically available term for the kind of normative relation that the first interpretation claims there to be. So we do not seem to think or talk about a relation of the sort envisioned by ACTUAL NORMATIVE RELATION. Furthermore, MODAL-NORMATIVE MIXED RELATION is simpler in the sense that ACTUAL NORMATIVE RELATION is committed to all the claims made by MODAL-NORMATIVE MIXED RELATION together with the claim that some actual normative relation makes these claims true.

According to this particularist reply, what Horty’s theory teaches us is that the particularist’s argument was not strictly speaking sound. But to avoid the argument’s conclusion, one must posit a normative relation that we seemingly heretofore did not think or talk about that determines what reasons we have and what we ought to do.

That said, others have taken a much less skeptical attitude to the kind of normative relation posited by ACTUAL NORMATIVE RELATION. Thomas Scanlon in recent work, for example, goes in for precisely this view and calls the relation R .⁴⁶ He writes:

the essentially normative content of a statement that $R(p, x, c, a)$ [that R holds between a proposition p , an agent x , a context c , and an act a] is independent of whether p holds. This normative content

⁴⁶An idea like this is found as early as Chisholm 1964. Horty 2012: 42-3 also appears to endorse it.

lies in the claim that, whether p obtains or not, should p hold then it is a reason for someone in c to do a . (Scanlon 2014: 40-1)

It is however hard to find any *argument* in Scanlon's book that such a normative relation exists rather than there being a mixed modal-normative relation determined by the embedding of a normative claim under modal expressions.⁴⁷ This may be because he simply assumes that there is such a normative relation and is defending this idea from objections.⁴⁸ While this is a worthwhile project, I do not think it adds any more support for ACTUAL NORMATIVE RELATION. And it provides no response to the concern that we have no evidence that there is such a relation that we have thought or talked about. Indeed, Scanlon's own gloss on what R is simply modally embeds a claim about reasons without doing anything to show us that R is itself a normative relation. So while positing such a relation may ultimately be worthwhile, we still must face up to this consequence.

Of course, there may be still other interpretations of what defaults are that are possible and problematic for the particularist. For example those who are reductivists may have analyses according to which defaults represent certain relations that are in the reductive base and these relations together with certain further facts determine what reasons we have.

If this right, then Horty's theory helps us make dialectical progress by having us focus on providing an interpretation of what defaults are. Since defaults serve a precisely defined role in the theory, this provides some constraints on what it takes for an interpretation to be admissible. Nonetheless, the theory still may allow for a variety of different interpretations corresponding to various particularist and anti-particularist views. It therefore does not settle the debate; rather, it refocuses where the debate should occur.

3.4 Other Problems and Competing Implementations

Here we catalogue some remaining issues and topics that are not directly related to the debate about a particularism but are important for spelling out the correct model of how reasons explain what we ought to do.

3.4.1 Derivative and Non-Derivative Reasons

Our thought and talk about reasons is rich and we often recognize a variety of reasons that are interestingly related to one another. For example, if I promised some friends that I will help them move, I not only take myself to have a reason to help my friends move but I also take myself to have a reason to get up at 6am if that is what it takes for me to help my friends move. There is an interesting relation among these two reasons.⁴⁹ The reason to get up at 6am is, in some

⁴⁷Cf. Schroeder 2015a: 196

⁴⁸The best explanation of Scanlon's remarks in my view is that he assumes that there are certain normative relations that hold of necessity so that their holding in one possible world suffices to show they hold in all the worlds.

⁴⁹This is true even if, due to the existence of weighty competing reason, I ought to refrain from helping my friends move and I ought to refrain from get up at 6am.

way, derivative of the reason to help my friends move; the reason to get up at 6am depends on the fact that getting up at 6am is means to helping my friends move. An interesting question, then, is whether the reasons one is modelling using tools like Horty's are reasons of both the derivative and non-derivative sort or just reasons of the non-derivative sort or whether this makes any difference.

The first thing to notice is that it does make a difference. For example, suppose I promised John that I would both take him to the Chinese embassy (to pick up his visa) and take him to LAX. In this case (I hereby stipulate), this is a single promise that I make to John to do two things. Suppose further still that John would not be interested in me merely committing to do one of these acts as he needs me to do both acts for either act to be worthwhile to him (there is no point in going to get the visa if he can't make it to his flight; there is no point in showing up for the flight without a visa). Next suppose that I have promised Mary that I will take her to Burbank airport. Finally, suppose that while I can take John to the embassy and LAX, can take John to the embassy and Mary to Burbank airport, and can take John to LAX and Mary to Burbank airport, I cannot do all three things a once. There isn't time for all that driving.

In this example, we get different results in Horty's system about what ought to be done depending on which reasons we include in our default theory. So far, we have mentioned two reasons. The reason to take John to the embassy and LAX and the reason to take Mary to Burbank. If these are the only two reasons we include, Horty's theory (on its non-conflict allowing interpretation) delivers the results that I ought to either take John to the embassy and LAX or take Mary to Burbank.

But it is very natural to take there to not merely be these two reasons. Instead, there are also some derivative reasons in this case. In particular, given that the promise provides a reason to take John to the embassy and LAX, it also provides a reason to take John to the embassy and a reason to take John to LAX. These are after all (constitutive) means to those ends.⁵⁰ If we include these reasons as additional defaults in Horty's theory, we get different results about what we ought to do. In particular, Horty's theory (on its non-conflict allowing interpretation) does not deliver the result that I ought to either take John to the embassy and LAX or take Mary to Burbank. Instead, we only get the result that I ought to take John to the embassy and LAX or Mary to Burbank and John to the embassy or Mary to Burbank and John to LAX. What this means is that if I were to only take Mary to Burbank, the first way of modelling this cases says that I will have done what I ought to have done while the second way of modelling the case says that I will have failed to do what I ought to do.

Which modelling choice is correct? I myself believe the first modelling choice is correct. What matters fundamentally is that I keep my promises. In this case, I cannot keep both. But there is no grounds for thinking I fail to do what I ought to do when I keeps one of the promises but do not partially fulfill the

⁵⁰If one thinks only causal means are supported by derivative reasons, it does no harm to change the example so that we discuss such means throughout

other. On the assumption that I haven't separately promised to do that act which is the partial fulfillment and on the assumption that there are no other reasons to do that act, there is, in my view, nothing amiss with what I do when I only take Mary to Burbank.⁵¹

If this is right, this tells us that one must only include the non-derivative reasons when modeling a case using Horty's theory. And this teaches us that there will need to be a substantial role for moral theory in making use of Horty's theory for we need to know what the non-derivative reasons are. This topic is a contested one in moral theory and as such, correctly implementing Horty's theory will be controversial as well.

That said, there may very well be local contexts in which the theory can be used without taking a stand on these issues. And it may be that Horty's theory is a useful tool for adjudicating certain debates about which reasons are derivative and non-derivative because it allows us to see exactly what these different views in moral theory predict about what we ought to do.

To take one example that illustrates this second point, there is a debate about the reasons that are provided by making promises. One view is that one has a standing reason with regard to promise keeping and that by making a promise to someone in particular, one thereby derivatively acquires a reason to keep the promise to that person. Another view is that fundamentally one has no reason with regard to promise keeping prior to making promises.⁵² Rather making a promise brings a (non-derivative) reason into existence. These views differ about whether there is any reason prior to promise making and differ about whether the reason one has because one makes a promise is derivative or non-derivative. As such, these views will end up making different predictions in different cases and Horty's theory provides a precise set of constraints that will allow us to investigate these differences.⁵³

3.4.2 Undercutting Defeat and Downward Closure

Let's return now to discuss two different ways of modelling undercutting defeat. According to one way of doing things, a reason is undercut when it is below a threshold in the ordering. According to the other, a reason is undercut when it has a certain property of being *Out*. So according to the first proposal whether a default is triggered (and in particular whether it is not undercut) depends on its place in the ordering. But according to the second proposal whether a default

⁵¹A slightly more complicated version of this objection is needed for the conflict allowing version of Horty's theory: Suppose one promises to do $A \wedge B$ and promised to do $C \wedge D$ where A and C are not compossible but the remaining acts are compossible. Here a default theory that only includes defaults corresponding to non-derivative reasons tells us that $\neg O(B \wedge D)$. On the other hand, a theory that includes defaults for both derivative and non-derivative reasons gets the result that $O(B \wedge D)$. I believe the results of the theory that only includes defaults corresponding to non-derivative reasons is correct.

⁵²Extremes of these different approaches are typical consequentialist accounts that fall in the first camp and the so-called normative powers approach that fall in the second. But this choice point in the theory of promise keeping also applies to other theories.

⁵³Cf. Schroeder 2007: ch. 3's discussion of the so-called standard model of normative explanations.

is triggered (and in particular whether it is not undercut) does not depend on its place in the ordering. Much like a default that is not triggered because its premise doesn't hold, a default that is not triggered because it is undercut can occur anywhere in the ordering on the second proposal.

This difference is interesting because the ordering has certain structural properties (transitivity and irreflexivity) that make it so anything that is lower in the ordering than an undercut reason is itself undercut if the first proposal is true. Horty calls this feature the *downward closure* of undercutting. If the second proposal is true, undercutting need not be downwardly closed. So a key question then is "Is undercutting downwardly closed?". According to the first proposal, the answer is 'yes'. According to the second proposal, the answer is 'no'.

The first answer is supported by what Mark Schroeder calls *The Undercutting Hypothesis* which says "complete undercutters are simply a limiting case of such partial undercutters (attenuators)" (Schroeder 2011a: 335). The idea here is that partial undercutters or attenuators are things which lower the place of a reason in an ordering. And complete undercutting is just a case where the reason has been lowered to a spot that is sufficiently low. Now one version of this view takes this to be the very bottom of the ordering. According to this view, it would be trivially true that anything lower in the ordering is undercut. But as it happens, there is good reason to think that the threshold in the ordering is not fixed but is instead context-sensitive.⁵⁴ In fact, this is a crucial part of Schroeder's view and indeed part of what motivates The Undercutting Hypothesis for him.

He gives the following example in support of this:

In the basic case, you are standing outside the library, when you see Tom Grabit exit, pull a book from under his shirt, cackle gleefully, and scurry off. This gives you pretty good reason to believe that Tom just stole a book from the library. Case 2 is just the same as the first case, except that Tom has an identical twin, Tim, from whom you can't visually distinguish him. In this case, it has seemed to the judgment of many philosophers that your visual evidence is not a reason to believe that Tom stole a book. Cases like these have been used in order to introduce the notion of undercutting defeat [...]

A simple argument, however, strongly suggests that things are more complicated in the Tom Grabit case. Consider a third version of the case, exactly like the other two except that in the third case, in addition to Tim, Tom has a third identical sibling, Tam, from whom you can't visually distinguish him. This third case underwrites a compelling argument against the intuitive judgment that in the second case, your visual evidence was no reason to believe that Tom stole the book. For if you go on to conclude, in the third case, that

⁵⁴Schroeder 2007: 5.3 presents his pragmatic approach. See Snedegar 2013 for criticism and a contrastive alternative.

Tom stole the book, then you are doing worse than if you had gone on to conclude this in the second case. Your reason to believe that Tom stole the book therefore doesn't seem to have gone away in the second case; it merely seems to have gotten substantially weaker. It seems to have been, in Dancy's [...] phrase, attenuated. [...]

And partial undercutting clearly comes in degrees. If the case in which Tom has two identical siblings shows that in the case in which he only has one, you still have a reason to believe that Tom stole a book, then a fourth case, in which Tom has three identical siblings, will show by analogous reasoning that in the two-sibling case, you still have a reason to believe that he stole a book. And if that is right, then we can construct an indefinite chain of increasingly powerful attenuators, each of which will leave you with a reason to believe that Tom stole a book – simply by arbitrarily increasing the size of his sibling cohort. But once we see that a reason can be arbitrarily attenuated, it is natural to contemplate [...] the *undercutting hypothesis* (Schroeder 2011a: 334-335)

Thus, while complete undercutting really is being at the bottom of the order, on Schroeder's view, the undercutting that we typically detect and talk about will be context-sensitive and depended on the importance of our reasons and what other reasons we think are relevant to deliberation. This supports the claim that undercutting should be a spot in the order because attenuating, it agreed all around, involves changing places in the order.

Against this conception, Horty has argued that there are cases where a reason is undercut but a reason lower than it in the ordering is not. Here is Horty's example:

Consider a normative interpretation in which a soldier, Corporal O'Reilly, is subject to the commands of three officers. We now take A as a command by the Captain that O'Reilly is to perform some action, where P stands for the proposition that O'Reilly performs that action, so that δ_1 represents the fact that the Captain's command favors P ; we take B as a command by the Major that the O'Reilly is not to perform that action, so that δ_2 represents the fact that the Major's command favors $\neg P$; and we take C as a command by the Colonel that O'Reilly is to disregard the Major's command—perhaps the Colonel knows the that Major is drunk—so that δ_3 represents the fact that the Colonel's command favors the exclusion of δ_2 . The priority ordering among defaults now corresponds to the rank, and so the authority, of the various officers, with the Major outranking the Captain and the Colonel outranking the Major.

Under this interpretation, it seems clear that the downward closure outcome is incorrect. Again, δ_3 provides a reason for excluding δ_2 —the Colonel has ordered O'Reilly to disregard the Major's command; this command cannot, therefore, be taken as a reason for $\neg P$. But

it is hard to see why δ_1 should be excluded, or why the Captain's command should be ignored. Imagine O'Reilly trying to explain to the Captain why he has ignored the Captain's command. O'Reilly might say: "The Colonel commanded me to ignore the Major." The Captain could reply: "But I am not the Major." O'Reilly might persist: "The Major outranks you. If I am not supposed to obey even a higher-ranking officer like the Major, why should I obey a lower-ranking officer like you?" But the Captain could again reply: "You were not commanded to ignore orders from the Major and also from all officers of lower rank. That would have been a different command from the one you were actually given, which was simply to ignore orders from the Major." At this point I think the Captain has won the dispute. (Horty 2012: 133)

What should we make of this counterexample? For my part, I believe that Horty is right about what one ought to do in this case. And that the Captain's reply is convincing. *Prima facie*, then, we have a counterexample to the first view. The idea is that the Major's reason is higher in the ordering than the Captain's reason but it is undercut and the Captain's isn't.

I am, however, less than certain about this last claim. The captain is correct when she says "You were not commanded to ignore orders from the Major and also from all officers of lower rank". Indeed, one may be tempted to say that the Colonel said not to pay attention to Major and so thereby treat the Major's commands less seriously than the commands of other officers. If that is true, we could claim that actually the Colonel's command tells us to put the Major lower in the ordering than officers of lower rank and indeed lower than the contextually relevant threshold as well. That is, while we originally start with an ordering according to rank, the ordering changes and then the Major is moved below the threshold. If this is implicit in the case, then the case is no threat to the Undercutting Hypothesis or the downward closure of undercutting.

Against this, one may simply wonder why the case must be interpreted in this way. To this, I have no firm reply other than it is difficult to clearly provide a case where only one of these interpretations is admissible. So it is not clear that the Undercutting Hypothesis is false or that undercutting fails to be downwardly closed.

That said, Horty for his part has a reply to Schroeder's case and accordingly the argument for the undercutting hypothesis. His view is that there may be local contexts in which one does have downward closure. In those contexts, we should explicitly encode this in the default information. His idea is that in cases like Schroeder's, our initial ordering comes from reliability considerations which support downward closure but in domains like the domain of military commands such downward closure is not supported.

For my part, I do not think it is clear which view is correct. It is worthwhile for moral theorists to consider what stance they would take on these issues and consider the different predictions that result. If there turn out to be promising theories that require one treatment rather than the other, this would perhaps

would be the best evidence for one treatment over the other.

3.4.3 The Accrual of Reasons

We have seen that default theories allow us to answer as well as pose many interesting questions in moral philosophy. As such, they are among the most promising formal tools for exploring ethical issues. Unfortunately, these theories fair poorly with accommodating one rather simple and central class of cases: Sometimes one can have two reasons to do an act and one reason against it. And it can turn out that one ought to do the first act even though each reason to do it is individually worse than the reason against it. The two reasons together, what is often called the *accrual* of these reasons, provides sufficient support to make it so you ought to do the act.

This kind of case, which I think you will agree is mundane, cannot be easily accommodated by default theories.⁵⁵ To illustrate the phenomenon, consider the following example that I have given elsewhere:

Suppose for example that there is a movie theater and a restaurant across town. And suppose that in order to get to that side of town I must cross a bridge that has a \$25 toll. The toll is a reason not to cross the bridge. The movie is a reason to cross the bridge and the restaurant is also a reason to cross the bridge. It may be that if there were just the movie to see, it wouldn't be worth it to pay the toll and if there were just the restaurant, it wouldn't be worth it to pay the toll. But given that there is both the movie and the restaurant, it is worth it to pay the toll. (Nair 2016: 56)

But in default theories each individual reason to cross the bridge will be defeated by the reason not to cross the bridge. So these theories get the result that you ought no to pay the toll.

Now as it happens not all cases of multiple reasons work like this. Here is an example adapted from an early case due to Henry Prakken (Prakken 2005: §3.1 and cf. Horty 2012: 61):

Suppose I am deliberating about an afternoon run, and that both heat and rain, taken individually, function as reasons to not run; still, the combination of heat and rain together might function as a weaker reason to not run (say, because the heat is less onerous when there is rain) (Nair 2016: 59)

And indeed one can easily modify my original example to illustrate this point as well. Simply suppose that there is one seating for dinner and one showing of the movie and they are at the same time so one cannot attend both. In this variant of the bridge case, the two reasons combined provide no additional support for crossing the bridge.

⁵⁵There is a small literature on this topic in default logic and argumentation theory: Delgrande and Schaub 2004, Gómez Lucero, Chesñevar, and Simari 2009, Gómez Lucero, Chesñevar, and Simari 2013, Modgil and Bench-Capon 2010, Prakken 2005, and Verheij 1995.

Perhaps, default theories can model these cases if they assume there is always some extra default in cases where having multiple reasons matters. So for example according to this view in the first bridge case one has an extra reason provided by the movie and the restaurant together. But in the second bridge case one does not. This treatment however is inadequate.

While it may be true that two reasons aren't always better than one, it is not true that in cases where two reasons are better than one, this is because there is some further reason that floats free of the original ones. Instead, whether there is a further reason and how strong it is in a given case appears to have a clear explanation. The explanation in the bridge cases has something to do with whether one can do attend both dinner and the movie. The explanation in the running cases has something to do with how onerous it is to run in various weather conditions.

Now these are not full explanations and the second one does not seem to have much prospect of being generalized to other cases. But they are nonetheless enough for us to be confident that there is some kind of explanation of what is happening.

The hard theoretical project is, however, providing a plausible and fully general explanation. To date, there are no especially good proposals that fit with default theories. Indeed, there are some grounds for thinking that to model this phenomena one may need to introduce the kinds of quantitative tools familiar from decision theory such as utility function and a probability function. This is because, in my view, the most satisfactory accounts to date make use of such resources.⁵⁶ That said, the topic is one that is wide open as of now. It is one that neither work in ethics nor in deontic logic satisfactorily treat. There is substantial room for collaboration.

3.4.4 Other Important Theories

I have focused on Horty's theory here. This is because it is perhaps presented in the most accessible single work. But there are other theories that have many of the advantages of Horty's theory. These theories differ however from Horty's in a variety of ways. Though the brevity with which I will introduce these theories does not do them justice, it is worth at least mentioning them so that the interested reader may look at them in greater detail.

Horty's theory comes from the default logic tradition. As it turns out, this is a rich tradition with a variety of alternatives to Horty's approach. Perhaps, the approach that admits of the easiest comparison is Jörg Hansen's approach which generalizes the approach of Gerhard Brewka and others.⁵⁷ Though the

⁵⁶See Nozick 1968 for a prescient discussion of some of these issues. See Sher 2019 for an approach that makes use of both probabilities and utilities. Nair 2020b develops an approach that only makes use of probabilities.

⁵⁷See especially Hansen 2008. Brewka's work is developed in the context of modeling agents reasoning about what to believe, see Brewka 1994 and Brewka and Eiter 2000. Other important alternative approaches in default logic include Delgrande, Schaub, and W. K. Jackson 1994, Baader and Hollunder 1995, Prakken 2010.

details of the formal theories are too complex to describe here, we can look at an important class of cases in which these approaches disagree.

Here is Horty's presentation of a particular version of this class of cases (sometimes called the Order Puzzle) together with his favored verdict about it:

Once again, we suppose that the agent is the hapless Corporal O'Reilly, and that he is subject to the commands of three superior officers: a Captain, a Major, and a Colonel. The Captain, who does not like to be cold, issues a standing order that, during the winter, the heat should be turned on. The Major, who is concerned about energy conservation, issues an order that, during the winter, the window should not be opened. And the Colonel, who does not like to be too warm and does not care about energy conservation, issues an order that, whenever the heat is on, the window should be opened. [...]

[...] O'Reilly's job is to obey his orders exactly as they have been issued. If he fails to obey an order issued by an officer without an excuse, he will be court-martialed. And, let us suppose, there is only one excuse for failing to obey such an order: that, under the circumstances, he is prevented from obeying the order issued by this officer by having chosen to obey another order or set of orders issued by officers of equal or higher rank. [...]

Given the set of commands that O'Reilly has been issued in the Order Puzzle, can he, then, avoid court martial? Yes, he can, by [...] obeying the orders issued by the Captain and the Colonel [...] In this scenario, O'Reilly fails to obey the Major's order [...] but he has an excuse: he was prevented from doing so by obeying an order issued by the Colonel, an officer of higher rank. (5) (Horty 2012: 204-)

Hansen, and other however disagree with this treatment of the case. Here is Hansen's presentation of a seemingly structurally identical case that suggests Horty's approach is incorrect:

Suppose that if I am attacked by a man, I must fight him (to defend my life, my family etc.). Furthermore, suppose I have pacifist ideals which include that I must not fight the man. Now you tell me to provoke him, which in the given situation means that he will attack me. Let self-defense rank higher than my ideals, which in turn rank higher than your request. Should I do as you request? By the reasoning advocated by Horty, there is nothing wrong with it: I satisfy your request, defend myself as I must, and though I violate my ideals, I can point out to myself that the requirement to fight back took priority. But I think if I really do follow your advice, I would feel bad. I think this would not just be some irrational regret for having to violate, as I must, my ideals, but true guilt for having

been tempted into doing something I should not have done, namely provoking the man: it caused the situation that made me violate my ideals. (Hansen 2008: 26)

Both cases are compelling and appear to be a structurally identical. It is an interesting question, then, which approach to choose and why. These very kinds of structures can predictably arise in moral theories that allow of a plurality of reasons that can conflict in complex ways. As such, moral philosophers should also be interested in understanding these kinds of structures and contributing to resolving the question of which way (if any) is best for handling them.

In addition to work in the default logic tradition, there is work in the tradition of argumentation theory that is importantly related to the framework discussed here. Within the formal tradition, Dung 1995 presents the classic approach. Within the philosophical tradition, a framework with some similarities to this is developed in the work of John Pollock (Pollock 1995).⁵⁸ The argumentation approach need not conflict with the default approach but it invites slightly different interpretations and can be developed in a conflicting way. For example, Pollock believed that an undercutting defeater must be stronger than the reason it undercuts in order to defeat it (ibid.: 103-4). This is not true in Horty's system as undercutting defeaters and the reasons they undercut do not conflict in his system and so their strengths are not especially relevant.⁵⁹ But the family of theories itself is very general and can be used to study a variety of perspectives on the issues that we have been discussing.

Finally, there is work in the input/output tradition which has recently been shown to be useful for modeling similar phenomena to the one that Horty's system models. These theories are interesting in their own right as they are formally quite different from other theories and invite different interpretation (see §5.2). Recently, those who make use of this formalism have shown that it can be useful to model contrary-to-duty obligation in a way Horty's system cannot (Parent 2011). And they have shown how to offer a distinctive take on priorities and exclusion among reasons that is relevant both to the Order Puzzle and the dispute about whether undercutting defeaters must be stronger than the reasons they undercut (Tucker 2018).

Each of these perspectives is worth further exploration and engagement from moral philosophers. They often give different verdicts on concrete examples of significance. And they often invite different interpretations that may fit better with certain ethical theories. To date, there has been little engagement with these alternatives to Horty's theory by moral philosophers.

4 Individual and Group Obligations

We turn now to our final main topic. In moral philosophy as well as in deontic logic it is typical to focus on what a given individual is obligated to

⁵⁸See Dung 1995: 4.2 and Prakken and Horty 2011 for comparisons between the argumentation approach and Pollock's theory.

⁵⁹See Horty 2012:§5.3.2 for discussion

do. But there are groups of individuals (or, if you prefer, some individuals) who together act and, seemingly, can be obligated to do various things. What inferential relations, if any, are there between claims about group obligation and claims about individual obligations?

The question is not an idle curiosity. Issues of great moment often tacitly can be characterized as involving group behavior and perhaps group obligation and we often take this to tell us something about what individuals are obligated to do.

Felix Pinkert in a recent article provides a helpful case that illustrates some of the issues that are at stake:

Ann and Ben are owners of two factories which are located opposite each other on a river. Both agents opt for a production process which releases waste chemicals into the river and thereby kill all the fish in the river and destroy the livelihood of a fishing community downstream. The waste from one factory alone would suffice to kill all the fish, and adding the waste from the other factory does no additional damage whatsoever [...]. If Ann or Ben were to unilaterally produce cleanly, this would make their production uncompetitive compared to the other factory, put them out of business, and destroy the livelihood of their employees. However, if they both were to produce cleanly, then this problem would not arise, and both factories would remain in business and the fishing community would flourish. Ann and Ben each employ 100 workers, the fishing community counts 100 people, and all that matters morally in this case are the livelihoods of the workers and fishermen. Further, the only available actions are either to pollute or to produce cleanly. In particular, Ann and Ben cannot come together and suggest and discuss a common strategy.

[...] The Two Factories becomes a challenge for Act Consequentialism only once we assume that Ann and Ben are both “uncooperative”, that is, each would pollute even if the other produced cleanly. [...] In The Two Factories, it is only if both agents are uncooperative that neither could have improved matters by acting differently and that Act Consequentialism judges that both act rightly. Lastly, Ann and Ben are fully aware of this situation. (Pinkert 2015: 973-4)

We can ask a number of questions about this example. Are Ann and Ben together obligated to not pollute? If so, how does this affect what Ann is obligated to and what Ben is obligated to do? Moreover how do facts about what Ann will do effect facts about what Ben is obligated to do and vice-versa?

As Pinkert suggests, standard act consequentialism appears to give certain answers to these questions. Standard consequentialism suggest that what we together *can* do is of no special interest to what I am obligated to do. All that matters is what others *would* do given what I do. I must then consider for each act, what others would do if I were to do that act and how good that situation

would be. In such a setting, Ann ought to pollute and Ben ought to pollute given that the other in fact will pollute. When it comes to the question of what we together should do, standard consequentialism if it applies to collections of people suggests that that we together ought to not pollute because if we together were to not pollute this would lead to the best outcome.

There are many other versions of cases like this. Some of which do not rely on the idea that others are uncooperative, but instead rely on the idea that no single act makes a difference. Though there are important differences between these kinds of examples, we will not dwell on this here.⁶⁰

Moral philosophers have also suggested certain high level theoretical principles concerning the relationship between what individuals ought to do and what they accomplish by collectively doing what they ought. These principles have come to be discussed under the banner of *The Principle of Moral Harmony* due to a famous paper by Fred Feldman introducing these ideas and exploring their importance:

With a few exceptions, moral philosophers seem to be agreed that, at the level of the individual, morality doesn't necessarily pay. Hardly anyone who thinks about it seriously would maintain that doing what he morally ought to do invariably benefits the agent more than would some worse alternative. However, when we rise from the level of the individual to the level of the social group, we find that the reverse is true. Quite a few moral philosophers seem to believe that when all the members of a social group do what they morally ought to do, the group as a whole does benefit more than it would have from the performance of any worse alternative set of actions. I shall say that any such view is a version of the Principle of Moral Harmony. (Feldman 1980: 166-167)

Different moral theories give different answers to the question of whether there is some correct version of the *Principle of Moral Harmony*.⁶¹

Standard consequentialism may claim that the *Principle of Moral Harmony* is false. In Pinkert's case, act consequentialism suggests each of agent ought to pollute. So if each does what she is obligated to do, they end up polluting. But this is worse than if each failed to do what each was obligated to do; namely, pollute. That said, some care is required in the formulation and evaluation of the *Principle of Moral Harmony*. For as Donald Regan (in Regan 1980) points out, if both agents in fact do acts that result in the best outcome (i.e., both do not pollute), consequentialism says they both acted rightly. So it is not obvious whether consequentialism is incompatible with whatever the precisely formulated and true version of the *Principle of Moral Harmony* is.

⁶⁰Regan 1980 is a seminal discussion of these issues in the context of consequentialist theories. Important recent discussions include Woodard 2008, Kagan 2011, Nefsky 2012, and Dietz 2016.

⁶¹An important early discussion of this principle is Regan 1980 especially p. 181ff. Important recent discussions of moral harmony include Estlund 2017, Portmore 2018, Portmore 2019: §5.3, and Kierland 2006.

We now turn to introducing some theories from deontic logic that bear on these matters.

4.1 Quantified Deontic Logic and Group Agency

There are at least two traditions for thinking about claims about what an individual agent ought to do. According to the simpler of the two, the claim that John ought to do x is to be understood as the claim that it ought to be that John does x . In this setting, we can represent the idea that the collection of John and Bill are obligated to do an act as the claim that it ought to be that John and Bill do this act.

4.1.1 Quantified Deontic Logic

We begin with this perhaps overly simple view of group obligation and see how it relates to our problem. We pay particular attention to the representation of these situations using quantification (e.g., claims like everyone or all of us ought to pollute). Later we turn to a more sophisticated approaches to representing agency and obligation of the sort introduced in §2.1.3 and consider how this approaches handles group obligation.

In the factory case, one perspective claims that it is true of each individual (Ann, Ben) that that individual ought to pollute, but it is not true that everyone (Ann and Ben) ought to pollute. According to our scheme for translating these claims in to claims about what ought to be, this means that in this case everyone is such that it ought to be that she pollutes. But it is not the case that it ought to be that everyone pollutes. This suggests that the following claim is false were ‘ $\mathbf{P}x$ ’ is interpreted as ‘ x pollutes’.

$$\forall xO(\mathbf{P}x) \rightarrow O(\forall x\mathbf{P}x)$$

In a setting a setting in which O is understood to be a modal necessity operator, this claim is an instance of the so-called Barcan formula

This same perspective also suggest that everyone (Ann and Ben) ought to not pollute but it is not true of each individual that she ought to pollute. According to our translation scheme, this means that it ought to be that everyone doesn’t pollute. But it is not the case that everyone is such that it ought to be that she doesn’t pollute. This suggests then that the following claim is false:

$$O(\forall x\neg\mathbf{P}x) \rightarrow \forall xO(\neg\mathbf{P}x)$$

This claim is an instance of the so-called converse Barcan formula.

Of course, this way of translating things may be incorrect. And we will consider other formalizations and interpretations of the example later. But for now let us focus on this initial first pass.

Compared to the propositional fragment of deontic logic, quantified deontic logic has received much less attention. So much of the ground that we will cover now is speculative and draws on analogies from other fields.

A standard way of thinking of debates about the Barcan formulae given the usual semantics for alethic modality is to think of it as concerning whether there could've been larger or smaller number of things. Very roughly, in the context of alethic modality, failures of Barcan formula (if there are any) are taken to illustrate that there could be more things than there (actually) are. And failures of the converse barcan formula (if there are any) are taken to illustrate that there could be fewer thing than there (actually) are. Let us examine this and consider what it might teach us about failures of the Barcan formula in the deontic context.⁶²

It is easiest to appreciate putative counterexamples to the Barcan formula and its converse if we work with their equivalent formulations involving existential quantifiers and possibility modals. In this setting the Barcan formula looks like this:

$$\diamond\exists x(\alpha) \rightarrow \exists x\diamond(\alpha)$$

where α is an arbitrary (possibly open) formula. To see why some reject, this claim, consider that it is possible for Wittgenstein to have had a daughter (though actually he did not). Where we let ' $\mathbf{D}xy$ ' stand for x is y 's (biological) daughter and we let w name Wittgenstein, we have it that $\diamond\exists x\mathbf{D}xw$. On the other hand, Wittgenstein doesn't have a daughter. Is there, nonetheless, someone or something that is possibly his daughter? It seems 'no'. Who or what would it be? Certainly not any of Wittgenstein's actual children or anyone else's children. Plausibly, then $\neg\exists x\diamond\mathbf{D}xw$.

How does this example fare in the deontic setting? It certainly seems permissible for Wittgenstein to have a (biological) daughter. And it may seem plausible then that there is among the ideal worlds one in which Wittgenstein has a daughter. But is there someone who exists here and now who is permitted to be Wittgenstein (biological) daughter? This question is harder to answer, but plausibly the answer is 'no' if we accept the counterexample to the alethic version of the Barcan formula. After all, in that case we think $\neg\exists x\diamond\mathbf{D}xw$. But this means there is nothing that is such that it could be Wittgenstein's daughter. And plausible a generalization of the idea that 'ought' implies 'can' suggests that no one is permitted to be a way they cannot be.

Let us consider how this line of thought fares in our pollution example. There it is permissible that someone refrain from polluting (because it is permissible that Ann and Ben together refrain from polluting). But it is not true of anyone that it is permissible for that person to pollute (because Ann ought to pollute and Ben ought to pollute). But the model from the alethic case does not help us to understand why these claims hold. It is not as though there is an deontically

⁶²The discussion below ignores the considerable resources of available to necessitists who accept the Barcan formulae and believe that there is a necessary framework of objects. I myself am sympathetic to this approach and believe many of these issues may be helpfully explored within a necessitist friendly deontic logic. But I do not explore this here for the sake of introducing our topic in a way that is connected to failures of simpler, more familiar formulae. See Williamson 2015 for a systematic defense and development of the necessitist picture. The examples below are inspired by Williamson's discussion.

ideal world where one of Ann or Ben exists even though one of them does not exist in the actual world. The existence or non-existence of certain agents is simply not a relevant feature of these kinds of collective action problems.

That said, while existence and non-existence may be of little relevance to the cases of collective action that we have in mind, there are interesting issues in ethics concerning the existence and nonexistence of agents. And it may be that the present model is better suited to exploring those issues. We will briefly consider this in §4.3. But for now, let us continue to explore how failures of the converse Barcan formula in the alethic setting may be relevant to our topic.

So consider a standard putative counterexample to the converse Barcan formula. Again, it helps to consider the version of it that makes use of existential quantifiers and possibility modals:

$$\exists x \diamond(\alpha) \rightarrow \diamond \exists x(\alpha)$$

Consider now that someone is such that it is possible that she does not exist. That seems true of me, for example. We might formalize this as follows $\exists x \diamond(\forall y(x \neq y))$. On the other hand, it is impossible for there to be something that is not identical to anything. So $\neg \diamond \exists x(\forall y(x \neq y))$.

Let us consider a similar case in a deontic context. Plausibly, there is someone who is permitted not to exist. Perhaps, this is someone whose life has been filled with nothing but pain, is incapable of forming interesting relationships, etc. Whatever the exact details are, it is highly plausible there are creatures that do not have lives worth living and plausibly it is true of them that they are permitted to not exist. So $\exists x P(\forall y(x \neq y))$ where ‘ P ’ is here interpreted as the operator ‘it is permissible that’. And if we accept the counterexample to the alethic Barcan formula, it also suggests we have a counterexample to the deontic one. This is because we generally have as a theorem $P(\alpha) \rightarrow \diamond(\alpha)$. So since $\neg \diamond \exists x(\forall y(x \neq y))$, $\neg P \exists x(\forall y(x \neq y))$.

Yet once again, it is not clear how this kind of understanding of failures of the Barcan formula help us to understand the target case of interest. There we have it that someone (each of Ann and Ben) is such that they are permitted to pollute. But we want to reject the claim that it is permissible for a person to pollute (because it is required that Ann and Ben (together) not pollute). What we just saw is that this can happen when one of the people we are talking about in the actual world fails to exist in an ideal world. But this, as before, is not relevant to the pollution case. Ann and Ben, we may assume, exist in the actual world as well as all ideal worlds.

One reaction to this is that this shows what is wrong with the view that the group ought to not pollute but each individual ought to pollute. But this reaction is mistaken for at least two reasons. First, while the model does not allow for the verdicts the consequentialists gives about the case in any sensible way, the model also does not represent much of the interesting underlying structure of the case. For example, we have no representation of the causal relations or counterfactual relations between the individual and group actions. Second, it may be that this representation of what agents ought to do in term of what ought to be the case is incorrect.

One way of trying to improve on these shortcomings is to consider what richer theories such as Lou Goble and Sven Ove Hansson’s theories (described in §2.1.2) say about this case. But we will not dwell on how these theories treat these cases because their application to these cases is straightforward and can be easily checked by the reader by consulting the more detailed description of the theories provided in §2.1.2.

Instead, I will simply state of the some relevant facts about how these theories treat the cases that we are discussing: First, both theories allow that it may be that some individuals ought to do something while the group ought to do an incompatible act (roughly, this is because of the way in which these theories fail to validate DEONTIC INHERITANCE). Second, since Goble’s theory has a counterfactual structure, it does allow us to model some aspects of the interaction of groups and individuals. Third, Hansson’s theory (on the simplest way of modelling these cases) suggests that if a group ought to do something then at least one individual in the group ought to do it as well (roughly, this is because Hansson’s theory validates DISJUNCTIVE DIVISION). Fourth, neither theory tells us much about the *Principle of Moral Harmony*. In principle both allow failures of it.

With these results in mind, we turn to a more complex theory that allows for a richer representation of the structure of our cases.

4.1.2 Group Agency

We can also explore these ideas in the context of the agency based deontic logic discussed in §2.1.3. Though we will briefly restate the crucial features of this framework, the reader who has not looked at §2.1.3 will need to consult it for a better understanding of these features and what motivates them. Our primary focus is how group obligation and the representation of group agency can be introduced in that framework.

Let us recall, then, the basics of Horty’s framework. We have frames with the following structure: $\langle Tree, <, Agent, Choice, Value \rangle$. *Tree* is a set containing elements that we call moments and use ‘*m*’, perhaps together with a subscript, to refer to a moment. $<$ is an ordering on *Tree* that ensures the moments form a “tree-like” branching time structure. A set of moment that form a complete unbroken path through the tree (more precisely, a maximal linearly ordered set of moments) is called a *history* and we use ‘*h*’, perhaps together with a subscript, to refer to a history. *Agent* is a set of agents and *Choice* is a function that maps an agent and a moment to a partition of histories through that moments. The cells of the partition are *acts* and we use ‘*K*’, perhaps together with a subscript, to refer to an act. Finally, *Value* assigns numbers to histories where these numbers are understood to be a measure of the value of each history. We use *M* to designate a model based on such a frame where we assume the usual semantics for our underlying propositional language and present the remaining interest semantic clause below.

Using these resources, we saw that we were able to analyze obligations and actions of individuals. We begin with the following useful definitions:

- $H_m = \{h \mid m \in h\}$
- $|A|^{M,m} = \{h \in H_m \mid A \text{ is true relative to } M, m/h\}$
- $Choice^{m,\alpha}(h) = \{K \in Choice^{m,\alpha} \mid h \in K\}$
- $State^{m,\alpha} = Choice^{m,Agent-\{\alpha\}}$
- For sets of histories, P and Q , $P \leq Q$ iff $Value(h) \leq Value(h')$ for all $h \in P$ and $h' \in Q$
- For acts, K_1 and K_2 , $K_1 \preceq K_2$ iff $K_1 \cap S \leq K_2 \cap S$ for each $S \in State^{m,\alpha}$
- $K_1 \prec K_2$ iff $K_1 \preceq K_2$ and $K_2 \not\preceq K_1$
- $Optimal^{m,\alpha} = \{K \in Choice^{m,\alpha} \mid \text{there is no } K' \in Choice^{m,\alpha} \text{ such that } K \prec K'\}$

This allows us to define what it takes for α to see to it that A , $[\alpha \text{ cstit: } A]$, and for α to be obligated to do A , $\odot[\alpha \text{ cstit: } A]$:

$[\alpha \text{ cstit: } A]$ is true relative to $M, m/h$ iff $Choice^{m,\alpha}(h) \subseteq |A|^{M,m}$

$\odot[\alpha \text{ cstit: } A]$ is true relative to $M, m/h$ iff $K \subseteq |A|^{M,m}$ for each $K \in Optimal^{m,\alpha}$

This picture can easily be generalized to apply not just to individual agents, but collections of agents. To do this, we need to generalize some of our definitions slightly so that they apply not to elements of *Agent* (i.e., individual agents) but subsets of *Agent* (i.e., groups of agents). The basic idea is just to take the acts available to the group to be the conjunction of acts available to the individuals. In order for this idea to be sensible, it is assumed that at a given moment the action of one agent does not affect the acts available to another agent at that moment. We call this the independence of agents property and require that models satisfy it. These ideas are made precise as follows:

$Select_m$ is a set of functions s that for each $\alpha \in Agent$, $s(\alpha) \in Choice^{m,\alpha}$

Independence of agents is satisfied iff $\bigcap_{\alpha \in Agent} s(\alpha) \neq \emptyset$ for each moment m and $s \in Select_m$

$Choice^{m,\Gamma} = \{\bigcap_{\alpha \in \Gamma} s(\alpha) \mid s \in Select_m\}$

Once we have the *Choice* function generalized to groups, all of the previous definitions and clauses can be directly applied to the choice sets of groups.

Let us consider than what this theory says about our guiding examples and the *Principle of Moral Harmony*. Figure 3 represent the simple two agent case that is discussed by Pinkert in the present framework. In this situation, α faces of choice to do K_1 or K_2 and β faces a choice of K_3 or K_4 . The result of α performing K_1 is α polluting which we represent as A_p ; analogously β

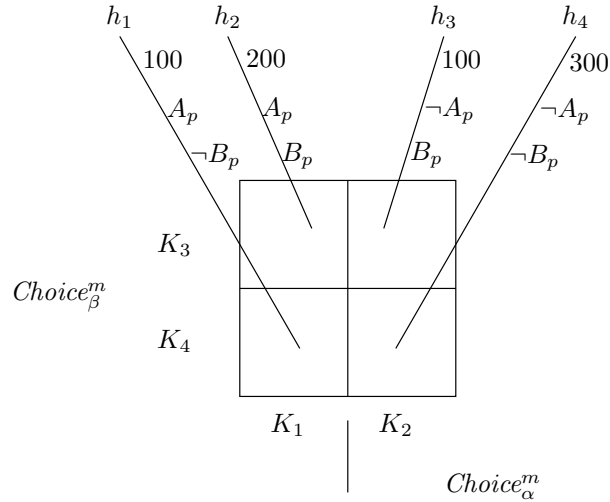


Figure 3: Two Agent Pollution Decision Tree

performing K_3 results in β polluting which we represent as B_p . The other acts result in the agent not polluting and instead producing cleanly. According to Pinkert's telling, when $A_p \wedge B_p$, only the 100 workers in both factories do well. This is what occurs in h_2 so its has a value of 200. When $A_p \wedge \neg B_p$, the 100 workers in α 's factories do well, but the 100 workers in β 's factories do not because they are at a competitive disadvantage because they do not pollute. This is what occurs in h_1 so it has a value of 100. Analogously, when $\neg A_p \wedge B_p$, only the 100 workers in β 's factories do well. This is what occurs in h_3 so it a value of 100. Finally, when $\neg A_p \wedge \neg B_p$, the 100 workers in both factories and the 100 people in the fishing village do well because the waters are not polluted. This is what occurs in h_4 so it has a value of 300.

If we consider now what α and β together ought to do, we must begin with checking what the choice set of this group is. According to our definitions $Choice^{m_1, \{\alpha, \beta\}} = \{K_1 \cap K_3, K_1 \cap K_4, K_2 \cap K_3, K_2 \cap K_4\}$. In this case each of these choice determines a unique history: $K_1 \cap K_3$ results in h_2 ; $K_1 \cap K_4$ results in h_1 ; $K_2 \cap K_3$ results in h_3 ; $K_2 \cap K_4$ results in h_4 . So determining the optimal act is just a matter of comparing the values of these histories and it is clear, therefore, that $K_2 \cap K_4$ is the unique optimal act. Thus what α and β together ought to do is not pollute. In other words, $\odot[\{\alpha, \beta\} \text{ cstit: } \neg A_p \wedge \neg B_p]$.

But what should each of these agents do individually? Begin with α who faces a choice between K_1 and K_2 . To determine, which act is optimal, we need to consider whether either act is state-wise dominant. In this case, the states are just given by the other agents acts. So our question reduces to whether some act results in the better outcome no matter what the other agent does. And it is easy to see both K_1 and K_2 are non-dominated in this sense. In the state K_3 , K_1 produces a better outcome (h_2); in K_4 , K_2 produces the better

outcome (h_4). Thus both K_1 and K_2 are optimal. So we have it that $\neg \odot [\alpha \text{ cstit: } \neg A_p]$ and $\neg \odot [\alpha \text{ cstit: } A_p]$. It is easy to check analogously that $\neg \odot [\alpha \text{ cstit: } \neg B_p]$ and $\neg \odot [\alpha \text{ cstit: } B_p]$.

We noted that in certain setting consequentialism not only suggested that α and β together ought to not pollute, but also that α ought to pollute and β ought to pollute. But notice Horty's theory does not have this result in this example. α is merely permitted to pollute and same for β .

The difference between the results that Horty's theory provides and the results that the consequentialist provide is that Horty treats what other agents will do as a genuinely open matter. The theory does not take it as given for the purpose of evaluating what one agent ought to do what the other agent will do. This is reflected in the fact that one must consider the value of one's act in states where the other agent does the action that the case suggest they will not do. Consequentialism however takes as settled what others will do (at least in cases where one cannot affect what other will do).⁶³

What of the *Principle of Moral Harmony*? It is hard to say whether or not Horty's system validates this principle without trying to state the principle more precisely. But we can notice one feature of Horty's theory that suggests it may validate something like the *Principle of Moral Harmony*. Suppose a group of individuals can realize the uniquely best history through their actions. In this setting, each individual doing what they ought to do can never ensure that this history fails to obtain.

We can informally see this as follows: Suppose h is the history that is the (uniquely) most valuable. According to Horty's theory, it can never be that if each of a set of individuals does what they ought to do, they will have seen to it that h fails to obtain. This is because any act/state pair compatible with h is guaranteed to be non-dominated. And as such, there will be no unique optimal act that ensures that h fails to hold for each individual. Could it be that though each of a pair of individuals, α , β , has no optimal act that ensures h does not obtain, they each have exactly one optimal act, K_1 and K_2 respectively, which are individually compatible with h but when done in tandem ($K_1 \cap K_2$) ensures h does not obtain? No, for if K_1 is uniquely optimal for α , this means the act/state pair $K_1 \cap K_2$ is strictly preferred to any alternative act available to α together with K_2 . But now notice that K_2 must not be the uniquely optimal act for β . This is because we know that there is a state S such that $K_1 \cap S$ is compatible with h (since K_1 is compatible with h). Given what we have said before, this state includes an act, K_3 , that is available to β and distinct from K_2 (because $K_1 \cap K_2$ is incompatible with h). But this means that K_2 cannot be the unique state-wise non-dominated act as $K_1 \cap K_3$ is not dominated by $K_1 \cap K_2$. So it must not be that each individual doing what they ought precludes the best history.

This informal demonstration rides roughshod over some important subtleties and complications. But hopefully, it is enough to give a sense of why the result

⁶³Horty notices this, shows how to define a consequentialist notion of obligation in his system, and shows that his formalization get the same results as consequentialism. See Horty 2001: §5.4 for the treatment of individual obligation and §6.2-6.3 for group obligation.

holds in Horty's theory. An important topic for further consideration is the formulation and evaluation of more precise versions of the *Principle of Moral Harmony*. It would be especially worthwhile to consider how an argument due to Donald Regan (1980) that claims to prove that certain natural formulations of the *Principle of Moral Harmony* cannot be true fares once regimented in a system like Horty's. That argument and related discussion since suggest a principled inadequacy with any approach that only represents and evaluates actions (rather than also considering the attitudes of agents).⁶⁴

So Horty's frameworks allow us to study the interaction of individual obligation and group obligation. And it also gives us some insight into whether the *Principle of Moral Harmony* holds. As it turns out, there are other logics in this tradition as well. While we do not have the space to consider them here, studying these logics in greater detail is a natural next step to take in trying to gain greater philosophical and formal traction on the issues raised by collective action problems.

4.2 A Speculative Alternative Framework

We close our discussion of collective action problems by proposing (but not developing) an alternative framework for understanding these problems that is suggested by some remarks from Derek Parfit. Parfit in his unpublished but much cited paper "What We Together Do" writes:

My suggested version of AC [act consequentialism] may seem incoherent. Suppose that, in Regan's Case, we both do *A*. On my suggestion, though each of us acts rightly, we act wrongly. This may seem impossible. How can truths about each not be true of us?

With some truths, this is not possible, Thus, if each is old, we cannot be young, Youth is a property of individuals: we together cannot be young. But other properties are different. Even though each is weak, we together may be strong.

My suggestion is of this second kind. (Parfit 1988: 8-9)

The case Parfit is discussing is much like our factory case. He is considering the view that it is right for each individual to pollute but it is wrong for us to pollute. Parfit here appears to be drawing our attention to the fact that some predicates like the predicates 'weak' and 'strong' are what linguists call *non-distributive* predicates. They are predicates that can apply to groups of people without applying to their members ('strong' in Parfit's example) or to individuals who make up a collection of people without applying to the collection ('weak' in Parfit's example). Parfit is suggesting 'right' and 'wrong' may exhibit such behavior.

We can take a leaf then from the study of these non-distributive predicates and see how it may apply to studying deontic logic. And it turns out there are

⁶⁴See Portmore 2019: §5.3 for discussion.

two main approaches to this issue. According to one approach, we enrich the kind of things there are to include not just ordinary individuals but also some things for when we are talking about several individuals. For example, perhaps there are also sets of individuals. Other leading proposals have been mereological sums, or events with mereological or set-theoretic structure. Whichever approach one favors, the idea is developed by claiming a predicate may apply to this entity (i.e., this set, mereological sum, etc.) without applying to the ordinary individuals. This approach is often favored by linguists.⁶⁵ And, in effect, this is the approach that we have already pursued in the previous subsection when discussing Horty's treatment of group obligation. There we assigned obligations to sets of individuals as distinct from the individuals who make up the set.

But there is another approach to these cases favored in certain circles of philosophical logic.⁶⁶ It is the approach that does not posit anything in addition to individuals, but rather involves the idea that we may attribute plurally some features to individuals. When some marbles are scattered, being scattered is a property of these marbles. One does not posit that there is some further entity, the set of marbles or the marble fusion, that has this property. These ideas are developed formally in theories that allow for plural predication and quantification.

My suggestion, inspired by Parfit's comments, is that it would be worthwhile to develop deontic logics with the resources of plural predication and quantification in order to model our reasoning about group obligation in collective action problems of the sort that we have been discussing. As far as I know, there has been no discussion in the literature of deontic logic about plural predication and quantification.

Here is not the place to make the first steps toward developing such a logic. But in recent years there has been some useful work in analogous fields. In particular, there has been work on understanding the interaction of plural predication and quantification and alethic modal logic. This work can serve as a tentative guide for us. Some of the highlights of this literature include discussion comparing the relation of an individual being one of some individuals and the relation of identity in modal context and include discussion of Barcan formulae involving plural quantifiers.⁶⁷

While quite speculative, my hunch is that this approach will be fruitful for exploring the relationship between individual obligation and the obligations of groups and for exploring the status of the *Principle of Moral Harmony*.

This concludes our discussion of collective action problems. As can be seen, the formal literature on this topic is not as rich as the literature we have discussed about our previous topics. This is especially true of approaches that are

⁶⁵Important representatives of this approach include Link 2002 and Schein 1993. Schein 2006 provides a survey.

⁶⁶Important representatives of this approach include Boolos 1984, Boolos 1985, Yi 1999, Oliver and Smiley 2001, Rayo 2002. Rayo 2007 provides a useful survey and McKay 2006 is a rich book length treatment.

⁶⁷See Bricker 1989, Linnebo 2016, Uzquiano 2004, Uzquiano 2011, and Williamson 2010.

related to quantification in deontic logic. The field is wide open.

4.3 The Ethics of Existence

We close our discussion by turning away from issues related to collective action problems and instead returning to consider whether there are any other examples from ethics that not only lead to failures of the deontic Barcan formulae but also do so in a way that makes sense given the traditional interpretation of what these formulae say. Since these formulae on their standard interpretation tell us something about what there is or what exists, a good place to look is at issues concerning existence.

And in moral philosophy, there are a number important issues related to existence and non-existence. Perhaps most famous among them is the so-called non-identity problem discussed originally in Derek Parfit's *Reasons and Persons* (Parfit 1986: ch. 16). Though we do not have the space here to explore this topic in any detail, we will briefly present the problem and connect it to our discussion of deontic Barcan formulae.⁶⁸

Molly Gardner in a recent article provides a nice explanation of the non-identity problem:

Consider the following two cases:

Case 1. During her pregnancy, Alice takes a drug that she knows will cause her child, whom she names Alex, to develop poor health. Despite his poor health, Alex has a life worth living. He would have had a higher level of well-being if Alice had not taken the drug.

Case 2. Barbara uses *in vitro* fertilization and screens the embryos for a particular gene that causes poor health. When she finds an embryo with that gene, she implants it and discards the rest. The selected embryo becomes a child named Billy, who develops poor health. Having poor health causes Billy to experience exactly the same hardships, pain, and suffering that having poor health causes Alex to experience. However, like Alex, Billy has a life worth living.

Intuitively, both Alice's action and Barbara's action are objectionable. The objection to Alice's action is that she has clearly harmed her child. Since Barbara's action is similar—it affects Billy in almost the same way that Alice's action affects Alex—we might be tempted to think that the objection to Barbara's action is also grounded in harm.

Nevertheless, there is a difference between Case 1 and Case 2. The difference is that, although Alex would still have existed had his

⁶⁸A recent survey is Roberts 2015. But see also Gardner 2015 and the citations therein.

poor health *not* been induced, Billy is *non-identical* to anyone who would have existed, had his poor health not been selected for. [...]

Many philosophers argue that this metaphysical difference makes a moral difference. According to the *counterfactually worse-off condition* on harming, an action harms someone only if it makes her worse off in at least some respect than she would have been, had the action not been performed. Alice's action satisfies this condition [...] However, Barbara's action does not satisfy this condition. Billy's life is worth living, and plausibly, having a life worth living is not worse for Billy in any respect than not existing; therefore, Billy is no worse off in any respect than he would have been, had Barbara not selected for poor health. [...]

But if Barbara's action does not harm Billy, then we seem to be at a loss to justify the intuition that, in much the same way that Alice's action is objectionable, Barbara's action is also objectionable. [...] The problem of either accounting for this appearance that the individual was wronged or explaining it away is the *non-identity problem*. (Gardner 2015: 428-429)

Crucial to the case is the interaction between a certain metaphysical claim—the claim that in w , $\neg\exists x x = \text{Billy}$ where w is the world that would result if Barbara had chosen a different embryo—and a moral principle that depends on it, the *counterfactually worse off condition* on harming. As Gardner says, the combination of these claims leads to the suggestion that Barbara's action is not wrong or at least seemingly that the grounds for it being wrong is much different than the grounds for Alice's action being wrong.⁶⁹

In the actual world, there is Billy and Billy has poor health. Evidently many believe that it ought to be that Barbara choose a different embryo much like how it ought to be that Alice not take the drug. So it ought to be that there is someone who is both Barbara's child and not Billy and (given the set up of the case) there is no one who is Billy. That is, $O(\exists x x \text{ is Barbara's child} \wedge x \neq \text{Billy} \wedge \neg\exists y y = \text{Billy})$. This is equivalent to the claim that $O(\exists x\forall y (x \text{ is Barbara's child} \wedge x \neq \text{Billy} \wedge y \neq \text{Billy}))$ and the claim that $O(\forall y\exists x (x \text{ is Barbara's child} \wedge x \neq \text{Billy} \wedge y \neq \text{Billy}))$.

From $O(\exists x\forall y(x \text{ is Barbara's child} \wedge x \neq \text{Billy} \wedge y \neq \text{Bill}))$, the fact that obligations entail permissions, and the Barcan formula, we have $\exists xP(\forall y(x \text{ is Barbara's child} \wedge x \neq \text{Billy} \wedge y \neq \text{Bill}))$. If we suppose Barbara's only actual child is Billy, then there is no one who is possibly Barbaras's child and not Billy.⁷⁰ So it seems $\neg\exists x\Diamond(\forall y (x \text{ is Barbara's child} \wedge x \neq \text{Billy} \wedge y \neq \text{Billy}))$ and it is therefore plausible that $\neg\exists xP(\forall y (x \text{ is Barbara's child} \wedge x \neq \text{Billy} \wedge$

⁶⁹Of course, consequentialist will say the reason why both acts are wrong is they do not bring about the best outcome and so they will deny that what makes Alice's act wrong is that it causes harm. Gardner, on the other hand, rejects the *counterfactually worse-off condition* on harming. We do not discuss the details of these approaches here.

⁷⁰This is true at least in the contingentist setting we adopted for the purposes of exposition in this article.

$y \neq \text{Billy}$)). Thus, reasoning about this example appears to suggest that the Barcan formula fails. And it fails because of the fact that in some deontically ideal world there is something that exists that does not exist in the actual world.

Similar results hold for the converse Barcan formula. From $O(\forall y \exists x(x \neq \text{Billy} \wedge y \neq \text{Billy}))$ and the converse Barcan formula, we have it that $\forall y O(\exists x(x \neq \text{Billy} \wedge y \neq \text{Billy}))$. But consider Billy as value for y . Given that Billy cannot fail to be self-identical, the embedded formula cannot be true in any world (ideal or otherwise). Thus, reasoning about this cases appears to suggest the converse Barcan formula fails. And it fails because there is something the exists in the actual world that fails to exist in all deontically ideal worlds.

Unlike the examples involving collective action problems, this example of a failure of the deontic Barcan formulae does seem to be sensibly related to the standard interpretation of the underlying formalism. What is ethically relevant is the existence of something in ideal worlds that doesn't exist in the actual world (Barcan formula) and the non-existence of something in ideal worlds that does exist in the actual world (converse Barcan formula)

This shows that there are examples of significance in ethics that provide motivation for exploring deontic logics where Barcan formulae fail.⁷¹ But can deontic logic provide any insight to help resolve or evaluate important arguments in moral philosophy concerning the ethics of existence? Here it is harder to say because, as far as I know, the topic has not been systematically explored to date.

What's more, there are a number of other important problems in moral philosophy that concern the existence and nonexistence of individuals.⁷² For example, there is the paradox of mere addition and the related repugnant conclusion that concerns issues about whether to bring more people into existence who have lives worth living.⁷³ All of these topics have, to my knowledge, received little to no systematic study by deontic logicians

5 Further Topics

I have chosen to look in detail at three particular issues in moral philosophy and consider their connection to various frameworks in deontic logic. But there are many other topics where there have been fruitful interaction or their could be fruitful interactions between deontic logicians and moral philosophers. We briefly consider a small selection of disparate topics.

⁷¹Or the exploration of necessitist deontic logics that allow for an alternative approach to putative failures of the Barcan formulae. We have, to repeat, focused on raising these issues in a contingentist manner only to simplify exposition.

⁷²These issue primarily originate from Parfit 1986: Part 4's discussion of population ethics but Naverson 1967 is an earlier work that explores some issues related to existence. Greaves 2017 is a recent survey of views from the perspective of formal axiology.

⁷³The paradox originates in Parfit 1986: ch. 17-19. Arrhenius, Ryberg, and Tännsjö 2017 provide a recent survey.

5.1 The Logical Form of Obligation

In ethics and deontic logic, there has been considerable study of the logical form of claims about obligation. One issue that has been of interest in both fields for some time is how best to represent what an agent ought to do as opposed to what ought to be the case. Another issue that has received attention in ethics in the last decade or two is about the proper scope of the obligation operator with respect to conditionals in statements of what rationality requires of us. Let us briefly look at each of these topics.

We speak of things that ought to be or occur (e.g. it ought to be that there is world peace) and we also speak of what agents ought to do (e.g., we ought to keep our promises). How are these related? On one hand, it seems as though when we discuss what an agent ought to do we are interested in a certain relation between an agent and an action. But since there appears to be no relation at all between an agent an action involved in claims about what ought to be, claims about what ought to be and what we ought to do, it seems, have substantially different logical forms. On the other hand, we might try to reduce what we ought to do to what ought to be. And a popular and tempting analysis (which we have used in various places in this article) claims that what we ought to do is just a special case of what ought to be; we ought to do something exactly when it ought to be that we do it. And there are many possible views in between these two extremes.

A variety of evidence from logic, semantic, and ethics have been brought to bear on choosing among these options.⁷⁴ And while the dominant tradition in logic and semantics has been to adopt the tempting analysis, a variety of formal approaches have been developed that depart from it including the framework discussed in §2.1.3 of this article.

Finally, a number of important issues in ethics may turn on which view about these issues is correct. In a recent article Mark Schroeder (Schroeder 2011b) mentions at least four: the viability of certain metaethical analyses, the viability deontology, the adequacy of the agent-neutral/agent-relative reasons distinction, and the prospects of wide-scope accounts of rationality.

This last item on Schroeder's list—the prospects of wide-scope accounts of rationality—is a topic that has received much discussion in its own right.⁷⁵ The issue here concerns what is rationally required of us. To focus on just one kind of example, we know that when someone believes that they ought to intend to do something and fails to intend to do it, something has gone wrong with them. They are, in some way, irrational. The following claim, if true, would explain why this agent is irrational:

⁷⁴Some important discussions from the linguistic, logical, and philosophical traditions include Broome 1999, Broome 2013, Castañeda 1981, Chisholm 1964, Chrisman 2012, Finlay 2014, Finlay and Snedegar 2014, Geach 1982, Horty 2001, J. Ross 2010, Schroeder 2011b, Wright 1951, Wedgwood 2006, and Williams 1981 [1980]

⁷⁵A small sampling of important papers on this topic include Broome 1999, Kolodny 2005, Kolodny 2007, Schroeder 2009. Kiesewetter 2017 and Lord 2018b are recent book length exploration of these and related topics.

If S believes that S ought to intend to do x , then S is rationally required to intend to do x

But many have objected to this claim on the grounds that it involves a kind of illegitimate bootstrapping. Suppose S is not rationally required to intend to do x and indeed S ought not to intend to do x . But nonetheless, suppose S now comes to believe that S ought to intend to do x . The above principle then says that S now in fact is rationally required to intend to do x . But S should not, now, intend to do x . Instead, S ought to stop falsely believing that S ought to intend to do x .

In light of this, some have proposed the following alternative account of our rational requirements:

It is rationally required that if S believes that S ought to intend to do x , then S intend to do x

Here, the thought goes, one can respond to this requirement by dropping the belief (as S should in some cases like the one described in the previous paragraph) or by forming the intention (which will be appropriate in many other cases).

The debate concerning this requirement as well as some related requirements turns not just on the correct verdict about certain examples but also how conditionals and requirements interact in the face of further factual and normative information. Arguments for and against the above view often make some assumptions about under what conditions one can conclude that one is rationally required to intend to do x given a proposed requirement and given certain pieces of factual and normative information.

This of course is an issue that has been studied in detail by deontic logicians. And work in ethics has benefited from this work in deontic logic even if there are still some insights from logic that have yet to be noticed.⁷⁶

5.2 Input/Output Logic

In a series of important papers in the early 2000s, David Makinson and Leedert van der Torre initiated the study of what they called “Input/Output Logics” (Makinson and van der Torre 2000, Makinson and van der Torre 2001). Since then, these logics have been used to study permissions, contrary-to-duty obligation, and a variety of other topic of importance in deontic logic and ethics. This handbook provides a detailed discussion of these logics (Parent and van der Torre 2013).

Here we consider whether there are some additional applications for this formal theory. In particular, from the perspective of the moral philosopher what are some interpretation of the formalism of input/output logic that could allow us to use this powerful formal theory to understand issues in ethics.

⁷⁶Some examples of work that appeals to certain logical principle include Broome 2007, Broome 2013, Schroeder 2009, Schroeder 2015b, Lord 2018b, and Lord 2018a. There are many others.

In their original paper, van der Torre and Makinson introduce the study of input/output logic as follows:

Imagine a black box into which we may feed propositions as input, and that also produces propositions as output. Of course, classical consequence may itself be seen in this way, but it is a very special case, with additional features—inputs are also themselves outputs, since any proposition classically implies itself, and the operation is in a certain sense reversible, since contraposition is valid. However, there are many examples without those features. Roughly speaking, they are of two main kinds.

The box may stop some inputs, while letting others through, perhaps in modified form. Inputs may record reports of agents, of the kind ‘according to source i , x is true’, while the box may give as output either x itself, a qualified version of x , or nothing at all, according to the identity of i . Or it might give output x only when at least two distinct sources vouch for it, and so on. [...] In these examples, the outputs express some kind of belief or expectation.

Again, inputs may be conditions, with outputs expressing what is deemed desirable in those conditions. The desiderata may be obligations of a normative system, ideals, goals, intentions or preferences. In general, a fact entertained as a condition may itself be far from desirable, so that inputs are not always outputs; and as is widely recognised, contraposition is inappropriate for conditional goals.

Our purpose is to develop a general theory of propositional input/output operations, covering both kinds of example. (Makinson and van der Torre 2000: 383-4)

Input/output logic aims to be a general theory of how one can transform some input to get some output where, it appears, that the transformations that we aim to model are those involved in inference.

In the article in this handbook dedicated to input/output logic and its relevance to deontic logic, Xavier Parent and Leendert van der Torre tell us “the first objective states that detachment is viewed as the core mechanism of the semantics of normative reasoning” where by detachment they mean the way in which one reaches conclusions about what one is obligated to do from conditional claims about what one is obligated to do together with other information (Parent and van der Torre 2013: 502). They also tell us:

The view of logic underpinning the I/O framework is very different. Its role is not to create or determine a distinguished set of norms, but rather to prepare information before it goes in as input to such a normative code, to unpack output as it emerges and, if needed, coordinate the two in certain ways. A set of conditional norms is, thus, seen as a transformation device, and the task of logic is to act as its “secretarial assistant”. (ibid.: 506)

The idea then is that for an arbitrary set of conditional norms, input/output logic helps us prepare inputs to the set of norms and retrieve outputs about what we ought to do given these inputs.

From the perspective of a moral philosopher, it is not immediately clear how best to understand these claims in a way that allows us to see what light input/output logic sheds on issue in moral philosophy. But two interpretations stand out corresponding to the two comments by van der Torre and Parent.

The first comment from Parent and van der Torre and the comment from Makinson and van der Torre highlight certain aspects of agents' reasoning about what they are obligated to do. As such, we can take them to be modelling good forms of reasoning about obligation. This in itself is not a model of what we in fact ought to do.

The second comment from Parent and van der Torre, on the other hand, more naturally suggests that the theory models what we in fact ought to do given that a certain set of conditional norms is in force (and certain other factual or normative information holds). This in itself is not a model of good forms of reasoning about obligation.

These different interpretations should be kept conceptually separate as one is a model of an epistemological issues while the other is a model of a metaphysical issue. And it may be that different desiderata bear on the adequacy of the model depending on whether it is a model of good forms of reasoning or a model of what is obligatory.

That said, we should not make too much of the difference either. One may take model of good forms of reasoning about obligation to give us insight into what we in fact ought to do if we are willing to accept certain theories of obligation. For example, according to some, facts about what we ought to do are explained by facts about *correct reasoning* about what (we ought) to do.⁷⁷ If we accept such a view, then a model of good forms of reasoning about obligation also turns out to be a model of what we are in fact obligated to do.

One may take it that the metaphysical model gives us insight into good forms of reasoning about obligation if we apply it to the set of conditional norms that the agent we are modeling accepts (as opposed to the conditional norms that obtain) and the set of inputs that the agent has received (as opposed to the facts).

Though we do not need to decide which of these interpretations is correct in order to see the interest in input/output logic, having a settled interpretations may be helpful: It may make it easier for those in ethics to see what lessons they can draw from the analysis of various deontic phenomena given by input/output logic. It may suggest further ways in which this formalism could be used to

⁷⁷See Williams 1981 [1979], Setiya 2014, and Way 2017

advance our understanding of issues in ethics.⁷⁸⁷⁹

5.3 Contrary-to-Duty Obligation

The logic of contrary-to-duty obligation (obligations conditional on failing to do what one is obligated to do) has been extensively studied in deontic logic. An article in this handbook provides a detailed survey of this work (??? Vol 2). However the topic has largely been neglected by moral philosophy. This may be because moral philosophers believe that moral theory need not give an account of contrary-to-duty obligations: The point of moral theory, some may believe, is to tell us what we are (unconditionally) obligated to do in a given situation. Once we fail to do what we are obligated to do, there may be new things that we ought to do, but there is no reason for an ethical theory to say now what we ought to do given that we fail to do something that we ought to do.⁸⁰

As plausible as this perspective might seem, it does not reflect the richness of human concern. We not only go about trying to do the right thing, but we also go about developing contingencies plans about what to do in case we fail to do the right thing: An alcoholic on the road to recovery will of course know that she ought not to drink tonight and plan not to do so. But a truly wise alcoholic on the road to recovery will also know that she ought to call her sponsor if she does drink and put in place a plan to do so conditional on her drinking. It is therefore a worthwhile area for philosophical investigation.

In recent years, there has been a number of papers that have explored more specific issues in moral theory with an eye to the special features of contrary-to-duty obligations. The actualist/possibilist debate (Kiesewetter 2015, Kiesewetter 2018, White 2017), rational requirements (Comesaña 2015), and moral particularism (Parent 2011) have all been the subject of some recent work connecting these issues to contrary-to-duty requirements. But the area is worth much deeper study by moral philosophers.

⁷⁸van der Torre also suggests to me that Liao, Slavkovik, and Torre n.d. and Benzmüller, Parent, and Torre n.d. are important contributions to ethical theorizing that make use of the input/output framework. The fact that discussion of these approaches from computer science to reasoning about applied problems are not included here, highlights the very specific perspective on ethical theory that we have adopted.

⁷⁹Similar issues of interpretation arise for the so-called “Theory of Joining Systems” (Lindahl and Odelstad 2013) but this theory has not been applied extensively to issues in ethics and its implications for ethics are less well understood (at least by the author of this paper).

⁸⁰Alternatively, moral theorists may believe that once they have given a theory of what we are obligated to do in an arbitrary case, this theory can be trivially generalized to also be a theory of contrary-to-duty obligation. While this may work for certain moral theories, the task is not trivial for reasons akin to the ones mentioned in §2.1.1 and §2.3.2. Notably, political philosophers often discuss “non-ideal theory” (Valentini 2012 provides a survey) and take it to be an important and non-trivial subject matter concerning what we ought to do given that we have failed to live up to the demands of “ideal theory” (which tells us, e.g., what the perfectly just state consists in). It is curious, then, that moral philosophers have taken considerably less interest in this subject matter (Korsgaard 1986 is an important exception).

5.4 Moral Conflicts

This handbook devotes an entire article to cases of moral conflict (Goble 2013). That rich article introduces a variety of important formal systems and compares the forms of reasoning and argument that they sanction. I recommend it to anyone interested in the topic of moral conflicts.

One area however that is under theorized in the formal tradition is the relationship between certain kinds of (reactive) attitudes such as guilt, resentment, indignation, etc. and obligation. While thinking about appropriate attitudes may not be relevant for many topics that deontic logicians are interested in, it is relevant to the topic of moral conflicts: A leading idea from Bernard Williams and Ruth Barcan Marcus has been that it is appropriate to have these reactive attitudes in cases of moral conflict no matter what one does. They have taken this as evidence that these cases involve genuinely conflicting all-things-considered obligations.⁸¹

What lies behind this idea is a commitment to a principled connection between obligations and reactive attitudes. So a worthwhile project for further work is the development of a logic that can model appropriate reactive attitudes and their connection to obligation.⁸² This will allow us to see the assumptions that are made by Williams-style arguments. And it will allow us to explore what the logic of appropriate reactive attitudes is and how it relates to the logic of obligations. We may then compare the different predictions made by those who accept or reject the existence of moral dilemmas.

6 Conclusion

Though deontic logicians and moral philosophers have not interacted at the depth that we might hope and aspire to in the future, there have been a number of fruitful interactions. We have chosen to focus primarily on three topics of interest in moral philosophy and a few representative theories in deontic logic that address these issues. But we have also seen that these topics and theories by no means exhaust the interesting terrain. The field is ripe for further productive interactions that will help us to better understand our reasoning about what is right and wrong and better develop our theories of what is right and wrong.

⁸¹This argument is sometimes known as the argument from “moral residue(s)”. See Williams 1988 [1965] and Marcus 1980.

⁸²A notable exception to the trend of ignoring attitudes in deontic logic is Paul McNamara’s work. McNamara explores the logic of what he calls aretaic attitudes (e.g., praise and blame) in the context of understanding supererogatory action. See McNamara 2011a for the formal theory and McNamara 2011b for further exploration of the philosophical importance of this theory.

References

- Antonelli, G. Aldo (1999). “A Directly Cautious Theory of Defeasible Consequence for Default Logic via the Notion of a General Extension”. In: *Artificial Intelligence* 109 (1-2), pp. 71–109.
- Arrhenius, Gustaf, Jesper Ryberg, and Torbjörn Tännsjö (2017). *The Repugnant Conclusion*. In: *Stanford Encyclopedia of Philosophy*. Ed. by Edward Zalta. Spring 2017. URL: <https://plato.stanford.edu/archives/spr2017/entries/repugnant-conclusion/>.
- Baader, Franz and Bernhard Hollunder (1995). “Priorities on Defaults with Prerequisites, and their Applications in Treating Specificity in Terminological Default Logic”. In: *Journal of Automated Reasoning* 15 (1), pp. 41–68.
- Benzmüller, Christoph, Xavier Parent, and Leendert van der Torre (n.d.). “Designing Normative Theories of Ethical Reasoning: Formal Framework, Methodology, and Tool Support”. In: *CoRR*.
- Boalos, George (1984). “To Be is to be a Value of a Variable (or to be Some Values of Some Variables)”. In: *Journal of Philosophy* 81 (8), pp. 430–449.
- (1985). “Nomalist Platonism”. In: *Philosophical Review* 94 (3), pp. 327–44.
- Bradley, Ben (2006). “Against Satisficing Consequentialism”. In: *Utilitas* 18 (2), pp. 97–108.
- Braham, Matthew and Martin van Hees (2015). “The Formula of Universal Law: A Reconstruction”. In: *Erkenntnis* 80 (2), pp. 243–260.
- Brewka, Gerhard (1994). “Reasoning about Priorities in Default Logic”. In: *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*, pp. 940–945.
- Brewka, Gerhard and Thomas Eiter (2000). “Prioritizing Default Logic”. In: *Intellectics and Computational Logic: Papers in Honor of Wolfgang Bibel*. Ed. by Steffen Hölldobler. Kluwer Academic Publishers, pp. 27–45.
- Bricker, Philip (1989). “Quantified Modal Logic and the Plural De Re”. In: *Midwest Studies in Philosophy* 14 (1), pp. 372–394.
- Broome, John (1999). “Normative Requirements”. In: *Ratio* 12 (4), pp. 398–419.
- (2007). “Wide or Narrow Scope?” In: *Mind* 116 (462), pp. 359–370.
- (2013). *Rationality Through Reasoning*. Oxford: Wiley-Blackwell.
- Brown, Campbell (2018). “Maximalism and the Structure of Acts”. In: *Noûs* 52 (4), pp. 752–771.
- Burgess, John (2009). *Philosophical Logic*. Princeton: Princeton University Press.
- Bykvist, Krister (2002). “Alternative Actions and the Spirit of Consequentialism”. In: *Philosophical Studies* 107 (1), pp. 45–68.
- Cariani, Fabrizio (2013). “‘Ought’ and Resolution Semantics”. In: *Noûs* 47 (3), pp. 534–558.
- (2016a). “Consequence and Contrast in Deontic Semantics”. In: *Journal of Philosophy* 113 (8), pp. 396–416.
- (2016b). “Deontic Modals and Probability: One Theory to Rule Them All”. In: *Deontic Modality*. Ed. by Nate Charlow and Matthew Chrisman. Oxford: Oxford University Press, pp. 11–46.

- Carlsson, Erik (1999a). “Consequentialism, Alternatives, and Actualism”. In: *Philosophical Studies* 96 (3), pp. 253–268.
- (1999b). “The Oughts and Cans of Objective Consequentialism”. In: *Utilitas* 11 (1), pp. 91–96.
- Carr, Jennifer (2015). “Subjective *Ought*”. In: *Ergo* 2 (27), pp. 678–710.
- Castañeda, Hector (1981). “The Paradoxes of Deontic Logic”. In: *New Studies in Deontic Logic*. Ed. by Risto Hilpinen. Dordrecht: Reidel, pp. 37–86.
- Chang, Ruth, ed. (1997). *Incommensurability, Incomparability, and Practical Reason*. Cambridge: Harvard University Press.
- Charlow, Nate (2016). “Decision Theory: Yes! Truth Conditions: No!” In: *Deontic Modality*. Ed. by Nate Charlow and Matthew Chrisman. Oxford: Oxford University Press, pp. 47–81.
- Chellas, Brian (1969). “The Logical Form of Imperatives”. PhD thesis. Philosophy Department, Stanford University.
- Chisholm, Roderick (1964). “The Ethics of Requirement”. In: *American Philosophical Quarterly* 1 (2), pp. 147–153.
- Chrisman, Matthew (2012). “‘Ought’ and Control”. In: *Australasian Journal of Philosophy* 90 (3), pp. 433–451.
- Cohen, Yishai and Travis Timmerman (2016). “Actualism Has Control Issues”. In: *Journal of Ethics and Social Philosophy* 10 (3).
- Comesaña, Juan (2015). “Normative Requirements and Contrary-to-Duty Obligations”. In: *The Journal of Philosophy* 112 (11), pp. 600–626.
- Curran, Angela (1995). “Utilitarianism and Future Mistakes: Another Look”. In: *Philosophical Studies* 78 (1), pp. 71–85.
- Dancy, Jonathan (2004). *Ethics without Principles*. Oxford: Oxford University Press.
- (2017). *Moral Particularism*. In: *Stanford Encyclopedia of Philosophy*. Ed. by Edward Zalta. Winter 2017. URL: <https://plato.stanford.edu/archives/win2017/entries/moral-particularism/>.
- Delgrande, James and Torsten Schaub (2004). “Reasoning with Sets of Defaults in Default Logic”. In: *Computational Intelligence* 20 (1), pp. 56–88.
- Delgrande, James, Torsten Schaub, and W. Ken Jackson (1994). “Alternative Approaches to Default Logic”. In: *Artificial Intelligence* 70 (1-2), pp. 167–237.
- Dietz, Alex (2016). “What We Together Ought to Do”. In: *Ethics* 126, pp. 955–982.
- Dung, Phan Minh (1995). “On the Acceptability of Arguments and Its Fundamental Role in Nonmonotonic Reasoning, Logic Programming, and n-Person Games”. In: *Artificial Intelligence* 77 (2), pp. 321–357.
- Estlund, David (2017). “Prime Justice”. In: *Political Utopias*. Ed. by Kevin Vallier and Michael Weber. Oxford: Oxford University Press, pp. 35–55.
- Feldman, Fred (1980). “The Principle of Moral Harmony”. In: *The Journal of Philosophy* 77 (3), pp. 166–179.
- (1986). *Doing the Best We Can: An Essay in Informal Deontic Logic*. Dordrecht: D. Reidel Publishing Company.

- Finlay, Stephen (2014). *Confusion of Tongues*. New York: Oxford University Press.
- Finlay, Stephen and Justin Snedegar (2014). “One Ought Too Many”. In: *Philosophy and Phenomenological Research* 89 (1), pp. 102–124.
- Gardner, Molly (2015). “A Harm-Based Solution to the Non-Identity Problem”. In: *Ergo* 2 (17).
- Geach, Peter T. (1982). “Whatever Happened to Deontic Logic?” In: *Philosophia* 11, pp. 1–12.
- Goble, Lou (1990a). “A Logic of *good*, *should*, and *would* Part I”. In: *Journal of Philosophical Logic* 19, pp. 169–199.
- (1990b). “A Logic of *good*, *should*, and *would* Part II”. In: *Journal of Philosophical Logic* 19, pp. 253–276.
- (1996). “Utilitarian Deontic Logic”. In: *Philosophical Studies* 82 (3), pp. 317–357.
- (2013). “Prima Facie Norms, Normative Conflicts, and Dilemmas”. In: *Handbook of Deontic Logic and Normative Systems*. Ed. by Dov Gabbay et al. Vol. 1. Milton Keynes: College Publications, pp. 241–352.
- Goldman, Holly (1976). “Dated Rightness and Moral Imperfection”. In: *Philosophical Review* 85, pp. 449–487.
- (1978). “Doing the Best One Can”. In: *Values and Morals*. Ed. by Alvin Goldman and Jaegwon Kim. Dordrecht: D. Reidel Publishing Company, pp. 185–214.
- Gómez Lucero, Mauro, Carlos Chesñevar, and Guillermo Simari (2009). “Modelling Argument Accrual in Possibilistic Defeasible Logic Programming”. In: *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pp. 131–143.
- (2013). “Modelling Argument Accrual with Possibilistic Uncertainty in a Logic Programming Setting”. In: *Information Sciences* 228, pp. 1–25.
- Greaves, Hilary (2017). “Population Axiology”. In: *Philosophy Compass* 12, e12442.
- Greenspan, Patricia (1978). “Oughts and Determinism: A Response to Goldman”. In: *Philosophical Review* 87, pp. 77–83.
- Gustafsson, Johan (2014). “Combinative Consequentialism and the Problem of Act Versions”. In: *Philosophical Studies* 167, pp. 585–596.
- Hansen, Jörg (2008). “Prioritized Conditional Imperatives”. In: *Autonomous Agents and Multi-Agent Systems* 17, pp. 11–35.
- Hansson, Sven Ove (2001). *The Structure of Values and Norms*. New York: Cambridge University Press.
- (2013). “Alternative Semantics for Deontic Logic”. In: *Handbook of Deontic Logic and Normative Systems*. Ed. by Dov Gabbay et al. Vol. 1. Milton Keynes: College Publications, pp. 445–498.
- Hooker, Brad and Margaret Little, eds. (2001). *Moral Particularism*. Oxford: Oxford University Press.
- Horty, John (2001). *Agency and Deontic Logic*. Oxford: Oxford University Press.
- (2007). “Reasons as Defaults”. In: *Philosopher’s Imprint* 7, pp. 1–28.
- (2012). *Reasons as Defaults*. Oxford: Oxford University Press.

- Hurka, Thomas (1990). “Two Kinds of Satisficing”. In: *Philosophical Studies* 59, pp. 107–111.
- Jackson, Frank (1985). “On the Semantics and Logic of Obligation”. In: *Mind* 94, pp. 177–195.
- (2014). “Procrastinate Revisited”. In: *Pacific Philosophical Quarterly* 95, pp. 634–647.
- Jackson, Frank and Robert Pargetter (1986). “Oughts, Actions, and Actualism”. In: *The Philosophical Review* 95 (2), pp. 233–255.
- Jeffrey, Richard (1990). *Logic of Decision*. 2nd ed. Chicago: Chicago University Press.
- Jennings, R.E (1974). “A Utilitarian Semantics for Deontic Logic”. In: *Journal of Philosophical Logic* 3, pp. 445–456.
- Kagan, Shelly (2011). “Do I Make a Difference?” In: *Philosophy and Public Affairs* 39, pp. 105–141.
- Kierland, Brian (2006). “Cooperation, ‘Ought Morally’, and Principles of Moral Harmony”. In: *Philosophical Studies* 128, pp. 381–407.
- Kiesewetter, Benjamin (2015). “Instrumental Normativity: In Defense of the Transmission Principle”. In: *Ethics* 125 (4), pp. 921–946.
- (2017). *The Normativity of Rationality*. Oxford: Oxford University Press.
- (2018). “Contrary-to-Duty Scenarios, Deontic Dilemmas, and Transmission Principles”. In: *Ethics* 129 (1), pp. 98–115.
- Kolodny, Niko (2005). “Why Be Rational?” In: *Mind* 114, pp. 371–385.
- (2007). “How Does Coherence Matter?” In: *Proceedings of the Aristotelian Society* 107, pp. 229–263.
- Korsgaard, Christine (1986). “The Right to Lie”. In: *Philosophy and Public Affairs* 15, pp. 325–349.
- Krantz, David et al. (2007). *Foundations of Measurement*. reprint. Vol. 1. Mineola: Dover Publishing.
- Lewis, David (2001 [1973]). *Counterfactuals*. 2nd ed. Oxford: Blackwell Publishing.
- (1974). “Semantic Analyses for Dyadic Deontic Logic”. In: *Logical Theory and Semantic Analysis*. Ed. by Sören Stenlund. Dordrecht: D. Reidel, pp. 1–14.
- Liao, Beishui, Marija Slavkovik, and Leendert van der Torre (n.d.). “Building Jiminy Cricket: An Architecture for Moral Agreements Among Stakeholders”. In: *CoRR*.
- Lindahl, Lars and Jan Odelstad (2013). “The Theory of Joining-Systems”. In: *Handbook of Deontic Logic and Normative Systems*. Ed. by Dov Gabbay et al. Vol. 1. Milton Keynes: College Publications, pp. 545–634.
- Lindner, Felix and Martin Mose Bentzen (2018). “A Formalization of Kant’s Second Formulation of the Categorical Imperative”. In: *Deontic Logic and Normative Systems: DEON 2018*, pp. 211–225.
- Link, Godehard (2002). “The Logical Analysis of Plurals and Mass Terms”. In: *Formal Semantics*. Ed. by Paul Portner and Barbara Partee. Oxford: Blackwell Publishing, pp. 127–146.

- Linnebo, Øystein (2016). “Plurals and Modals”. In: *Canadian Journal of Philosophy* 86, pp. 654–676.
- Lord, Errol (2017). “What You Are Rationally Required to Do and What You Ought to Do (Are the Same Thing!)” In: *Mind* 126, pp. 1109–1154.
- (2018a). “The Explanatory Problem for Cognitivism about Practical Reason”. In: *Normativity*. Ed. by Conor McHugh, Jonathan Way, and Daniel Whiting. Oxford: Oxford University Press, pp. 137–61.
- (2018b). *The Importance of Being Rational*. New York: Oxford University Press.
- MacFarlane, John (2000). “What Does It Mean to Say Logic Is Formal?” PhD thesis. University of Pittsburgh.
- Makinson, David (2005). *Bridges from Classical to Nonmonotonic Logic*. Milton Keynes: College Publications.
- Makinson, David and Leendert van der Torre (2000). “Input/Output Logics”. In: *Journal of Philosophical Logic* 29, pp. 383–408.
- (2001). “Constraints for Input/Output Logics”. In: *Journal of Philosophical Logic* 30, pp. 155–185.
- Marcus, Ruth Barcan (1980). “Moral Dilemmas and Consistency”. In: *Journal of Philosophy* 77, pp. 121–136.
- McKay, Thomas (2006). *Plural Predication*. Oxford University: Oxford University Press.
- McKeever, Sean and Michael Ridge (2006). *Principled Ethics*. Oxford: Oxford University Press.
- McNamara, Paul (2011a). “Praise, Blame, Obligation, and DWE”. In: *Journal of Applied Logic* 9, pp. 153–170.
- (2011b). “Supererogation, Inside and Out”. In: *Oxford Studies in Normative Ethics*. Ed. by Mark Timmons. Vol. 1. Oxford: Oxford University Press, pp. 202–235.
- Meachem, Christopher and Jonathan Weisberg (2011). “Representations Theorems and the Foundations of Decision Theory”. In: *Australasian Journal of Philosophy* 89 (4), pp. 641–663.
- Modgil, Sanjay and Trevor Bench-Capon (2010). “Integrating Dialectical and Accrual Modes of Argumentation”. In: *Proceedings of the 2010 Conference on Computational Models of Argument*, pp. 335–346.
- Moore, G.E (1962 [1903]). *Principia Ethica*. Cambridge: Cambridge University Press.
- Nair, Shyam (2016). “How Do Reasons Accrue?” In: *Weighing Reason*. Ed. by Errol Lord and Barry Maguire. New York: Oxford University Press, pp. 56–73.
- (2020a). “Fault Lines in Ethical Theory”. In: *The Oxford Handbook of Consequentialism*. Ed. by Douglas Portmore. New York: Oxford University Press, pp. 67–92.
- (2020b). “The Moral Mathematics of ‘Adding Up’ Reasons: Lessons for Reductive and Non-Reductive Approaches”. Unpublished paper.
- Naverson, Jan (1967). *Morality and Utility*. Baltimore: John Hopkins Press.

- Nefsky, Julia (2012). “Consequentialism and the Problem of Collective Harm”. In: *Philosophy and Public Affairs* 39, pp. 364–395.
- Nolt, John (2018). *Free Logic*. In: *Stanford Encyclopedia of Philosophy*. Ed. by Edward Zalta. Fall 2018. URL: <https://plato.stanford.edu/archives/fall2018/entries/logic-free/>.
- Nozick, Robert (1968). “Moral Complications and Moral Structures”. In: *Natural Law Forum* 13, pp. 1–50.
- Oliver, Alex and Timothy Smiley (2001). “Strategies for a Logic of Plurals”. In: *Philosophical Quarterly* 51, pp. 289–306.
- Parent, Xavier (2011). “Moral Particularism in the Light of Deontic Logic”. In: *Artificial Intelligence and Law* 19, pp. 75–98.
- Parent, Xavier and Leendert van der Torre (2013). “Input/Output Logics”. In: *Handbook of Deontic Logic and Normative Systems*. Ed. by Dov Gabbay et al. Vol. 1. Milton Keynes: College Publications, pp. 499–544.
- Parfit, Derek (1986). *Reasons and Persons*. Oxford: Oxford University Press.
- (1988). “What We Together Do”. Unpublished paper. URL: http://individual.utoronto.ca/stafforini/parfit/parfit_-_what_we_together_do.pdf.
- Pinkert, Felix (2015). “What If I Cannot Make a Difference (and Know It)”. In: *Ethics* 125, pp. 971–998.
- Pollock, John (1995). *Cognitive Carpentry*. Cambridge: MIT Press.
- Portmore, Douglas (2011). *Commonsense Consequentialism: Wherein Morality Meets Rationality*. New York: Oxford University Press.
- (2013). “Perform Your Best Option”. In: *Journal of Philosophy* 110, pp. 436–459.
- (2017). “Maximalism vs Omnism about Permissibility”. In: *Pacific Philosophical Quarterly* 98, pp. 427–452.
- (2018). “Maximalism and Moral Harmony”. In: *Philosophy and Phenomenological Research* 96, pp. 318–341.
- (2019). *Opting for the Best*. New York: Oxford University Press.
- Prakken, Henry (2005). “A Study of Accrual of Arguments”. In: *Proceedings of Tenth International Conference on Artificial Intelligence and Law*, pp. 85–94.
- (2010). “An Abstract Framework for Argumentation with Structured Arguments”. In: *Argument and Computation* 1, pp. 93–124.
- Prakken, Henry and John Horty (2011). “An Appreciation of John Pollock’s Work on the Computational Study of Argument”. In: *Argument and Computation* 3, pp. 1–19.
- Ramsey, Frank (1931 [1923]). “Truth and Probability”. In: *The Foundations of Mathematics and other Logical Essays*. Ed. by R.B. Braithwaite. New York: Hartcourt Brace Company, pp. 156–198.
- Rayo, Agustín (2002). “Word and Objects”. In: *Noûs* 36, pp. 436–464.
- (2007). “Plurals”. In: *Philosophy Compass* 2, pp. 411–427.
- Raz, Joseph (2002 [1975]). *Practical Reasoning and Norms*. Oxford: Oxford University Press.

- Rechenauer, Martin and Olivier Roy (2014). “The Logical Structure of Scanlon’s Contractualism”. In: *Deontic Logic and Normative Systems: DEON 2014*, pp. 166–176.
- Regan, Donald (1980). *Utilitarianism and Co-operation*. New York: Oxford University Press.
- Reiter, Raymond (1980). “A Logic for Default Reasoning”. In: *Artificial Intelligence* 13 (1-2), pp. 82–132.
- Roberts, M.A (2015). *The Non-Identity Problem*. In: *Stanford Encyclopedia of Philosophy*. Ed. by Edward Zalta. Winter 2015. URL: <https://plato.stanford.edu/archives/win2015/entries/nonidentity-problem/>.
- Ross, Jacob (2010). “The Irreducibility of Personal Obligations”. In: *Journal of Philosophical Logic* 39, pp. 307–327.
- (2012). “Actualism, Possibilism, and Beyond”. In: *Oxford Studies in Normative Ethics*. Ed. by Mark Timmons. Vol. 2. Oxford: Oxford University Press, pp. 243–282.
- Ross, W. D. (1930). *The Right and the Good*. Oxford: Oxford University Press.
- Sayre-McCord, Geoffrey (1986). “Deontic Logic and the Priority of Moral Theory”. In: *Noûs* 20, pp. 179–197.
- Scanlon, Thomas (1998). *What We Owe to Each Other*. Cambridge: Belknap Press.
- (2014). *Being Realistic About Reasons*. Oxford: Oxford University Press.
- Schein, Barry (1993). *Plurals and Events*. Cambridge: MIT Press.
- (2006). “Plurals”. In: *The Oxford Handbook of Philosophy of Language*. Ed. by Ernest Lepore and Barry Smith. Oxford: Oxford University Press, pp. 716–767.
- Schroeder, Mark (2007). *Slaves of the Passions*. Oxford: Oxford University Press.
- (2009). “Means-End Coherence, Stringency, and Subjective Reasons”. In: *Philosophical Studies* 143, pp. 223–248.
- (2011a). “Holism, Weight, and Undercutting”. In: *Noûs* 45, pp. 328–344.
- (2011b). “Ought, Agents, Actions”. In: *Philosophical Review* 120, pp. 1–41.
- (2015a). “*Being Realistic About Reasons*, by T.M. Scanlon”. In: *Australasian Journal of Philosophy* 93, pp. 195–198.
- (2015b). “Hypothetical Imperatives: Scope and Jurisdiction”. In: *Reason, Value, and Respect*. Ed. by Robert Johnson and Mark Timmons. Oxford: Oxford University Press, pp. 89–100.
- (2018). “Getting Perspective on Objective Reasons”. In: *Ethics* 128, pp. 289–319.
- Sergot, Marek (2013). “Normative Positions”. In: *Handbook of Deontic Logic and Normative Systems*. Ed. by Dov Gabbay et al. Vol. 1. Milton Keynes: College Publications, pp. 353–406.
- Setiya, Kieran (2014). “What is a Reason to Act”. In: *Philosophical Studies* 167 (2), pp. 221–235.
- Sher, Itai (2019). “Comparative Value and the Weight of Reasons”. In: *Economics & Philosophy* 35 (1), pp. 103–158.

- Sidgwick, Henry (1981 [1907]). *The Methods of Ethics*. 7th ed. Indianapolis: Hackett Publishing.
- Slote, Michael (1984). “Satisficing Consequentialism”. In: *Proceedings of the Aristotelian Society* 58, pp. 139–163.
- Snedegar, Justin (2013). “Negative Reason Existentials”. In: *Thought* 2, pp. 108–116.
- (2014). “Contrastive Reasons and Promotion”. In: *Ethics* 125 (1), pp. 39–63.
- (2017). *Contastive Reasons*. Oxford: Oxford University Press.
- Sobel, John Howard (1976). “Utilitarianism and Past and Future Mistakes”. In: *Noûs* 10, pp. 195–219.
- Tempkin, Larry (2015). *Rethinking the Good*. Oxford: Oxford University Press.
- Timmerman, Travis (2015). “Does Scrupulous Securitism Stand-Up to Scrutiny”. In: *Philosophical Studies* 172, pp. 1509–1528.
- Timmerman, Travis and Yishai Cohen (2016). “Moral Obligations: Actualist, Possibilist, or Hybridist”. In: *Australasian Journal of Philosophy* 94, 672–686.
- Tucker, Dustin (2018). “Variable Priorities and Exclusionary Reasons in Input/Output Logic”. In: *Journal of Philosophical Logic* 47, pp. 947–964.
- Uzquiano, Gabriel (2004). “The Supreme Court and the Supreme Court Justices”. In: *Noûs* 38, pp. 135–153.
- (2011). “Plural Quantification and Modality”. In: *Proceedings of the Aristotelian Society* 111, pp. 219–250.
- Valentini, Laura (2012). “Ideal vs. Non-Ideal Theory”. In: *Philosophy Compass* 7, pp. 654–663.
- Verheij, Bart (1995). “Accrual of Arguments in Defeasible Argumentation”. In: *Proceedings of the Second Dutch/German Workshop on Nonmonotonic Reasoning*, pp. 217–224.
- Vessel, John-Paul (2003). “Counterfactuals for Consequentialists”. In: *Philosophical Studies* 112, pp. 103–125.
- (2009). “Defending a Possibilist Insight in Consequentialist Thought”. In: *Philosophical Studies* 142, pp. 183–195.
- (2016). “Against Securitism, the New Breed of Actualism in Consequentialism”. In: *Utilitas* 28, pp. 164–178.
- Way, Jonathan (2017). “Reasons as Premises of Good Reasoning”. In: *Pacific Philosophical Quarterly* 98, pp. 251–270.
- Wedgwood, Ralph (2006). “The Meaning of ‘Ought’”. In: *Oxford Studies in Metaethics*. Ed. by Russ Schafer-Landau. Vol. 1. Oxford: Oxford University Press, pp. 127–160.
- White, Stephen (2017). “Transmission Failures”. In: *Ethics* 127 (3), pp. 719–732.
- Williams, Bernard (1981 [1980]). “Ought and Moral Obligation”. In: *Moral Luck*. Cambridge: Cambridge University Press, pp. 114–123.
- (1981 [1979]). “Internal and External Reasons”. In: *Moral Luck*. Cambridge: Cambridge University Press, pp. 101–113.
- (1988 [1965]). “Ethical Consistency”. In: *Essays on Moral Realism*. Ed. by Geoffrey Sayre-McCord. Ithaca: Cornell University Press, pp. 41–58.

- Williamson, Timothy (2008). *Philosophy of Philosophy*. Oxford: Blackwell Publishing.
- (2010). “Necessitism, Contingentism, and Plural Quantification”. In: *Mind* 119, pp. 657–748.
- (2015). *Modal Logic as Metaphysics*. Oxford: Oxford University Press.
- Woodard, Chris (2008). *Reasons, Patterns, and Cooperation*. London: Routledge.
- Wright, G.H von (1951). “Deontic Logic”. In: *Mind* 60, pp. 1–15.
- Yi, Byeong-uk (1999). “Is Two a Property”. In: *Journal of Philosophy* 96, pp. 163–190.
- Zimmerman, Michael (1996). *The Concept of Moral Obligation*. Cambridge: Cambridge University Press.
- (2006). “The Relevant Risks to Wrongdoing”. In: *The Good, The Right, Life And Death: Essays in Honor of Fred Feldman*. Ed. by Kris McDaniel et al. Aldershot: Ashgate, pp. 151–172.