

---

# DEONTIC MODALS: WHY ABANDON THE CLASSICAL SEMANTICS?

BY

JOHN HORTY

---

**Abstract:** I begin by reviewing classical semantics and the problems presented by normative conflicts. After a brief detour through default logic, I establish some connections between the treatment of conflicts in each of these two approaches, classical and default, and then move on to consider some further issues: priorities among norms, or reasons, conditional oughts, and reasons about reasons.

## 1. *Introduction*

This article is based on an invited presentation at the USC Deontic Modality Conference, held in the spring of 2013. The goal of that conference was to bring together people thinking about deontic modality from a variety of perspectives – ethicists, linguists, psychologists, logicians, and a computer scientist or two – in the hope that we might actually talk to each other. With this goal in mind, I took it as my task to compare the approach to deontic logic set out in my recent book, *Reasons as Defaults* (2012), with the dominant treatment of deontic modality in linguistics and philosophy, drawing on a long tradition in intensional semantics, but especially as it has been developed in a series of influential papers by Angelika Kratzer.<sup>1</sup> It is this dominant treatment that I refer to here as the classical semantics for deontic modals.<sup>2</sup>

One might expect that, in the natural way of things, my comparison between these two treatments of deontic modality would include an effort to demonstrate the superiority of my own approach to that of classical semantics. But that is not how I feel at all. I appreciate the classical

approach, and intensional semantics more generally, for a variety of reasons. Let me list two. First, the classical approach fits into a rich framework of compositional semantics, allowing for the integration of deontic modals into theories of tense and aspect, for example, along with a variety of other constructions. Second, this approach enables us to draw precise comparisons between the ought statements, or oughts, studied within linguistics and philosophy and those generated by some of our best theories of rational action, including decision theory, game theory, and – just around the corner – epistemic game theory.<sup>3</sup> The classical semantics fully deserves its status as the dominant approach.

Nevertheless, like a lot of people whose work on the subject originated in other fields – such as legal theory, or computer science – I have been exploring deontic logics from a standpoint different from that provided by classical semantics. My own approach is based on default logic, a form of nonmonotonic logic first developed in computer science, by Raymond Reiter.<sup>4</sup> Others, of course, are working within different frameworks: a sampling can be found in the recent work by Marek Sergot on normative positions, by Davide Grossi and Andrew Jones on constitutive norms, by David Makinson, Xavier Parent, and Leon van der Torre on input/output logics, by Lars Lindahl and Jan Odelstad on the theory of joining systems and intermediate concepts, and by Dov Gabbay on reactive systems.<sup>5</sup>

There are, evidently, many comparisons to be drawn, but I will concentrate here on those involving my own work. Given the success of the dominant approach, it is natural to ask why anyone would ever consider abandoning this approach – why should we abandon classical semantics? In answering this question, I will consider some of the advantages offered by my alternative approach, based on default logic, and ask: to what extent can these advantages be accommodated within classical semantics. The answers I provide are mixed. Some of what I thought were distinct advantages of my approach can be accommodated very naturally within classical semantics, in ways that I ought to have realized but did not. Others, I am not sure about. Still others, I believe cannot be accommodated in any natural way, though it is possible to argue that these further advantages bear on the theory of normative reasoning more generally, rather than on the meaning of deontic modals alone.

I begin the article by reviewing classical semantics and the problems presented by normative conflicts. After a brief detour through default logic, I establish some connections between the treatment of conflicts in each of these two approaches, classical and default, and then move on to consider some further issues: priorities among norms, or reasons, conditional oughts, and reasons about reasons. A few facts about the logics are appealed to at various points in the article, especially for the purpose of establishing connections. Although no proofs are presented here, there is an appendix noting where verifications of these facts can be found.

## 2. *Classical semantics*

The basic idea underlying classical semantics is straightforward: there is a preference ordering on possible worlds, and what ought to be the case is determined, somehow, by what is the case in the best of the available worlds.

In the simplest theory of this sort, known as *standard deontic logic*, all worlds are available, and the preference ordering divides worlds into only two categories: the good worlds, and the bad worlds. More exactly, standard deontic logic is based on models of the form  $\mathcal{M} = \langle W, h, v \rangle$  – known as *standard deontic models* – with  $W$  a set of possible worlds,  $v$  a valuation mapping sentence letters into sets of worlds at which they are true, and  $h$  a function mapping each world  $\alpha$  into a nonempty set  $h(\alpha)$  of worlds, representing those that are good, or ideal, from the standpoint of  $\alpha$ . The nonemptiness constraint can be seen as requiring that there is always be at least one good world. Alternatively, the set  $h(\alpha)$  can be thought of as a proposition expressing the standard of obligation at work in  $\alpha$ , with nonemptiness requiring that this standard must be satisfiable.

The satisfaction relation  $\models_{SDL}$  for standard deontic logic is defined in the usual way for sentence letters and Boolean connectives, and the evaluation rule for the modal connective  $\bigcirc$ , representing ‘It ought to be that . . . ,’ reads as follows.

**Definition 1 (Standard evaluation rule:  $\bigcirc A$ )** Where  $\mathcal{M} = \langle W, h, v \rangle$  is a standard deontic model and  $\alpha$  is a world from  $W$ ,

- $\mathcal{M}, \alpha \models_{SDL} \bigcirc A$  if and only if  $h(\alpha) \subseteq |A|_{\mathcal{M}}$ .

Here,  $|A|_{\mathcal{M}}$  is the set of worlds  $W$  at which  $A$  holds according to the model  $\mathcal{M}$ , though we will abbreviate this as  $|A|$  when reference to the model is not necessary. The idea behind the rule is that  $\bigcirc A$  holds at a world just in case  $A$  holds in each of the ideal worlds – or, just in case  $A$  is required by the relevant standard of obligation.

It is a striking feature of standard deontic logic, and one we will return to, that the theory rules out the possibility of normative conflicts. To be precise: let us say that a situation gives rise to a normative conflict if it presents each of two conflicting propositions as oughts – if, for example, it supports the truth of both  $\bigcirc A$  and  $\bigcirc B$  where  $A$  and  $B$  are inconsistent, or as an extreme case, if it supports the truth of both  $\bigcirc A$  and  $\bigcirc \neg A$ . It is easy to see why the standard theory rules out conflicts like this. In order for the two statements  $\bigcirc A$  and  $\bigcirc B$  to hold at a world  $\alpha$ , we need both  $h(\alpha) \subseteq |A|$  and  $h(\alpha) \subseteq |B|$ , from which it follows that  $h(\alpha) \subseteq |A| \cap |B|$ . If the statements  $A$  and  $B$  were inconsistent, however, we would have  $|A| \cap |B| = \emptyset$ , from which it would follow that  $h(\alpha) = \emptyset$ ; but in standard

deontic models, the sole requirement on  $h$  is that it should map each world into a nonempty set. Apart from what is built in to the background framework of normal modal logic, the entire content of standard deontic logic is simply that there are no normative conflicts; and in fact, validity in these standard models can be axiomatized by supplementing the usual axioms of normal modal logic with  $\neg(\bigcirc A \wedge \bigcirc \neg A)$  as an additional axiom schema.<sup>6</sup>

Kratzer's approach generalizes standard deontic logic in two ways.<sup>7</sup> First, not all worlds are available, or under consideration, at any given point of evaluation, but only a restricted set of worlds, known as the *modal base*. Formally, this concept is defined through a function  $f$  mapping each world  $\alpha$  into a set of propositions  $f(\alpha)$ , taken to represent some set of conversational assumptions supplied by the background context. It is stipulated that these assumptions should be consistent, so that the modal base can be defined as their nonempty intersection – or, formally, that the modal base at  $\alpha$  is  $\bar{f}(\alpha) = \bigcap f(\alpha)$ . Second, the preference ordering on worlds is more complex than that found in standard deontic logic, and is derived from a function  $g$  mapping each world  $\alpha$  into a set of propositions  $g(\alpha)$ , representing the background set of norms at work in  $\alpha$ , and known as the *ordering source*. Like the propositions from  $f(\alpha)$ , determining the modal base, those belonging to the ordering source are also supposed to be supplied by context; but, in contrast, there is no requirement that the propositions belonging to the ordering source should be consistent – since, after all, we are often faced with inconsistent norms.

We can define a *Kratzer model* as a structure of the form  $\mathcal{M} = \langle W, f, g, v \rangle$ , with  $W$  a set of worlds,  $v$  a valuation, and  $f$  and  $g$  functions from worlds to sets of propositions, as specified above, determining the modal base and the ordering source. For any world  $\alpha$ , the ordering source  $g(\alpha)$  is used to order the worlds through the stipulation that one world  $\beta$  is at least as good as another world  $\gamma$  if  $\beta$  satisfies all those proposition from  $g(\alpha)$  that  $\gamma$  does:

$$\beta \leq_{g(\alpha)} \gamma \text{ if and only if, for all } X \in g(\alpha), \text{ if } \gamma \in X, \text{ then } \beta \in X.$$

Taking  $\models_{KD}$  as the satisfaction relation for Kratzer's logic, the crucial clause, for the deontic operator, is as follows.

**Definition 2 (Kratzer evaluation rule:  $\bigcirc A$ )** Where  $\mathcal{M} = \langle W, f, g, v \rangle$  is a Kratzer model and  $\alpha$  is a world from  $W$ ,

- $\mathcal{M}, \alpha \models_{KD} \bigcirc A$  if and only if, for all  $\beta \in \bar{f}(\alpha)$ , there is a  $\gamma \in \bar{f}(\alpha)$  such that  $\gamma \leq_{g(\alpha)} \beta$  and, for all  $\delta \in \bar{f}(\alpha)$ , if  $\delta \leq_{g(\alpha)} \gamma$ , then  $\delta \in |A|_{\mathcal{M}}$ .

The intuition is this:  $\bigcirc A$  holds just in case, for every world under consideration – that is, in the modal base – there is a world at least as good such that  $A$  then holds in every world at least as good as that one.

This complicated evaluation rule, and the accompanying complicated intuition, can be streamlined under certain simplifying assumptions. One of these – which I adopt throughout this article – is that the Kratzer models under consideration are *norm finite*, in the sense that, at any world  $\alpha$ , the ordering source  $g(\alpha)$  contains only a finite number of propositions as norms. A world  $\beta$  can be defined as strictly better than  $\gamma$ , according to the ordering source, if it is as least as good as  $\gamma$  and the converse does not hold:

$$\beta \leq_{g(\alpha)} \gamma \text{ if and only if } \beta \leq_{g(\alpha)} \gamma \text{ and it is not the case that } \gamma \leq_{g(\alpha)} \beta.$$

And the best worlds from a set  $X$  can then be defined as its undominated members, those members for which the set contains no others that are strictly better:

$$Best_{g(\alpha)}(X) = \{\beta \in X : \neg \exists \gamma \in X (\gamma <_{g(\alpha)} \beta)\}.$$

Given the assumption that the Kratzer models under consideration are norm finite, we can conclude that these models are also *stoppered*, in the sense that: at each world  $\alpha$ , for any  $\beta$  from  $\bar{f}(\alpha)$ , either  $\beta$  itself belongs to  $Best_{g(\alpha)}(\bar{f}(\alpha))$  or there is another world  $\gamma$  from  $Best_{g(\alpha)}(\bar{f}(\alpha))$  such that  $\gamma \leq_{g(\alpha)} \beta$ . What stoppering tells us, in other words, is that each world from the modal base is either itself one of the best available worlds, or is dominated by a best world.<sup>8</sup> And under these conditions, we can define the Kratzer ought by focusing only on the best worlds, stipulating – much as in standard deontic logic – that  $\bigcirc A$  holds whenever  $A$  holds in each of these best worlds.

**Fact 1** If  $\mathcal{M} = \langle W, f, g, v \rangle$  is a stoppered Kratzer model and  $\alpha$  is a world from  $W$ , then  $\mathcal{M}, \alpha \models_{KD} \bigcirc A$  if and only if  $Best_{g(\alpha)}(\bar{f}(\alpha)) \subseteq |A|_{\mathcal{M}}$ .

What is the relation between standard deontic logic and the logic defined here on the basis of norm finite Kratzer models? We begin by noting that there are simple mappings between models of the two kinds, preserving satisfaction. Suppose, first, that  $\mathcal{M} = \langle W, h, v \rangle$  is a standard deontic model. We can then define the *Kratzer transform* of  $\mathcal{M}$  as the model  $\mathcal{M}' = \langle W, f, g, v \rangle$ , with  $W$  and  $v$  as in  $\mathcal{M}$  and, for each world  $\alpha$  from  $W$ , with  $f(\alpha) = \{W\}$  and  $g(\alpha) = \{h(\alpha)\}$ ; the Kratzer transform of a standard model thus accepts the entire set of worlds as its modal base, and takes as the single norm in its ordering source at any given world the notion of

obligation at work at that same world in the standard model. These two models satisfy the same oughts at each world, of course.

**Fact 2** If  $\mathcal{M} = \langle W, h, v \rangle$  is a standard deontic model and  $\mathcal{M}'$  is its Kratzer transform, then  $\mathcal{M}, \alpha \models_{SDL} \bigcirc A$  if and only if  $\mathcal{M}', \alpha \models_{KD} \bigcirc A$ , for each world  $\alpha$  from  $W$ .

In the same way, beginning with a norm finite Kratzer model  $\mathcal{M} = \langle W, f, g, v \rangle$ , we can define the *standard transform* of  $\mathcal{M}$  as the model  $\mathcal{M}' = \langle W, h, v \rangle$  with  $W$  and  $v$  as in  $\mathcal{M}$  and, for each world  $\alpha$ , with  $h(\alpha) = Best_{g(\alpha)}(\bar{f}(\alpha))$ ; here, the proposition expressing the standard of obligation at any given world is identified with the set containing the best worlds from the modal base associated with that world in the Kratzer model, according to the ordering source at that world. Again, the two models satisfy the same oughts.

**Fact 3** If  $\mathcal{M} = \langle W, f, g, v \rangle$  is a norm finite Kratzer model and  $\mathcal{M}'$  is its standard transform, then  $\mathcal{M}, \alpha \models_{KD} \bigcirc A$  if and only if  $\mathcal{M}', \alpha \models_{SDL} \bigcirc A$ , for each world  $\alpha$  from  $W$ .

It follows from Facts 2 and 3 that any statement that can be falsified in a model of either sort, standard or norm finite Kratzer, can be falsified in a model of the other sort. As a result, the two logics support the same set of validities, and so are, in this sense, the same logic.<sup>9</sup>

### 3. Normative conflicts

As we saw, standard deontic logic rules out the possibility of normative conflicts, and of course, since Kratzer's theory validates exactly the same formulas, that theory must deny the possibility as well. The idea that normative conflicts might be excluded on logical grounds, however, is one that many people have found to be troubling.

To begin with, it seems fair to say that there is currently no consensus among moral theorists on the question whether an ideal ethical theory could allow for moral conflicts.<sup>10</sup> Because the question is open, therefore, and the possibility of moral conflicts is a matter for substantive discussion, it would be odd for a negative verdict on the issue to be built into a logic of the subject. And even if it does turn out, ultimately, that research in ethical theory is able to exclude the possibility of moral conflicts, it could be useful all the same for a deontic logic to allow conflicting oughts. One reason for this is that our everyday moral reasoning is surely guided, not by principles from an ideal ethical theory, but by simple rules of thumb – 'Return what you borrow,' or 'Don't cause harm' – and it is not hard to

generate conflicts among these. Another reason is that our normative reasoning more generally is shaped by matters other than morality – etiquette, aesthetics, fun – and of course, these lead to conflicts both among themselves and with the oughts of morality.

For these reasons and others like them, a number of logicians in the 1960's and 70's began to explore ways of weakening standard deontic logic to allow for the possibility of moral conflicts without outright inconsistency. Much of this early work involved an appeal to non-normal modal logics, based on the minimal model, or neighborhood, semantics, in which the accessibility relation maps individual worlds, not into sets of worlds, but into sets of propositions. The clearest example is due to Brian Chellas, who recommends a deontic logic based on *minimal* models of the form  $\mathcal{M} = \langle W, N, v \rangle$ , with  $W$  and  $v$  as before, but with  $N$  a function mapping each world  $\alpha$  into a set  $N(\alpha)$  of propositions; just as in Kratzer semantics, the propositions belonging to  $N(\alpha)$  can be taken to represent the various norms in force at  $\alpha$ .<sup>11</sup> The crucial clause of the satisfaction relation  $\models_M$  for Chellas's minimal logic can be stated as follows

**Definition 3 (Minimal model evaluation rule:  $\bigcirc A$ )** Where  $\mathcal{M} = \langle W, N, v \rangle$  is a minimal model and  $\alpha$  is a world from  $W$ ,

- $\mathcal{M}, \alpha \models_M \bigcirc A$  if and only if there is a proposition  $X$  in  $N(\alpha)$  such that  $X \subseteq |A|_M$ .

And validity is axiomatized by supplementing ordinary propositional logic with

$$A \supset B / \bigcirc A \supset \bigcirc B$$

as an additional rule schema.<sup>12</sup>

In fact, this logic is now weak enough to tolerate normative conflicts: the statements  $\bigcirc A$  and  $\bigcirc B$  are jointly satisfiable even when  $A$  and  $B$  are inconsistent – just consider, for example, a model in which  $N(\alpha)$  contains exactly the two propositions  $|A|$  and  $|\neg A|$ , so that  $\alpha$  supports both  $\bigcirc A$  and  $\bigcirc \neg A$ , without supporting  $\bigcirc(A \wedge \neg A)$ . However, it seems that the logic may now be too weak. Consider the two norms ‘Fight in the army or perform alternative service’ and ‘Don’t fight in the army,’ the first issuing, perhaps, from some legal authority, the second from religion or conscience. Suppose an agent is subject to both of these norms, and only to these norms, at the world  $\alpha$ , so that  $N(\alpha)$  contains exactly the proposition  $|F \vee S|$  and  $|\neg F|$ , where  $F$  is the statement that the agent fights in the army and  $S$  is the statement that the agent performs alternative service. Then both  $\bigcirc(F \vee S)$  and  $\bigcirc \neg F$  hold at  $\alpha$ , as they should. From an intuitive standpoint, however, it seems that  $\bigcirc S$  should hold as well – the agent

ought to perform alternative service. But this latter ought statement fails, since there is no proposition  $X$  in  $N(\alpha)$  such that  $X \subseteq |S|$ .

Let us look at this problem more closely. In Chellas's minimal logic, just as in standard deontic logic and in Kratzer's account, ought statements are closed under logical consequence – that is, if  $\bigcirc A$  holds and  $B$  is a logical consequence of  $A$ , then  $\bigcirc B$  holds. And of course,  $S$  is a logical consequence of  $(F \vee S) \wedge \neg F$ . Therefore, we would have  $\bigcirc S$  if we could somehow merge the individual oughts  $\bigcirc(F \vee S)$  and  $\bigcirc \neg F$  together into a combined ought of the form  $\bigcirc((F \vee S) \wedge \neg F)$ . But how could we get this latter, combined ought? It would follow at once, of course, through a rule of the form

$$\bigcirc A, \bigcirc B / \bigcirc(A \wedge B),$$

dubbed by Bernard Williams as the rule of *agglomeration*.<sup>13</sup> However, this rule is exactly the kind of thing that Chellas's minimal logic is designed to avoid: from  $\bigcirc A$  and  $\bigcirc \neg A$ , agglomeration would lead to  $\bigcirc(A \wedge \neg A)$ , and so to  $\bigcirc B$  for arbitrary  $B$ , due to closure of ought under logical consequence.

The issue of agglomeration is crucial for a proper logical understanding of normative conflicts. We cannot allow unrestricted agglomeration, as in standard deontic logic; this would force us to treat conflicting oughts as incoherent. On the other hand, it seems, we cannot block agglomeration entirely, as in Chellas's logic; this would prevent us from reaching desirable consequences in cases in which there is no threat of conflict. As far as I know, the first appropriate response to this dilemma – the first approach allowing enough agglomeration, but not too much – was presented by Bas van Fraassen, in a paper that is largely devoted to philosophical issues concerning the nature and possibility of moral conflicts, but which presents a new logic at the end.<sup>14</sup>

In van Fraassen's logic, ought statements of the form  $\bigcirc A$  are derived from an underlying set  $\mathcal{I}$  of imperatives, each of the form  $!(B)$ . The account relies on a notion of *score*. Where  $\mathcal{M}$  is an ordinary model of the underlying propositional language, a simple valuation mapping statement letters into truth values, the score of  $\mathcal{M}$  relative to a set  $I$  of imperatives is defined as the subset of imperatives from  $\mathcal{I}$  that are satisfied by  $\mathcal{M}$  – or put formally, with  $\models$  as the satisfaction relation between models and statements, the score of  $\mathcal{M}$  relative to  $\mathcal{I}$  is

$$\text{score}_{\mathcal{I}}(\mathcal{M}) = \{!(B) \in \mathcal{I} : \mathcal{M} \models B\}.$$

With this notion in hand, Van Fraassen's satisfaction relation  $\models_F$  between a set of imperatives and the ought statements it supports can be defined as follows.



**Definition 4 (van Fraassen evaluation rule:  $\bigcirc A$ )** Where  $\mathcal{I}$  is a set of imperatives,

- $\mathcal{I} \models_F \bigcirc A$  if and only if there is a model  $\mathcal{M}_1 \models A$  for which there is no model  $\mathcal{M}_2 \models \neg A$  such that  $score_{\mathcal{I}}(\mathcal{M}_1) \subseteq score_{\mathcal{I}}(\mathcal{M}_2)$ .

The idea behind this definition is that  $\bigcirc A$  holds, given some background set of imperatives, just in case the truth of  $A$  is a necessary condition for achieving a maximal score based on those imperatives. Van Fraassen's satisfaction relation can also be characterized in terms of derivability from a consistent subset of imperatives. Taking  $\vdash$  to indicate the ordinary derivability relation, we have:

**Fact 4** Where  $\mathcal{I}$  is a set of imperatives,  $\mathcal{I} \models_F \bigcirc A$  if and only if there some consistent set  $\mathcal{G}$  such that  $\mathcal{G} \subseteq \{B : !(B) \in \mathcal{I}\}$  and  $\mathcal{G} \vdash A$ .

As in Chellas's minimal logic, this approach tolerates normative conflicts. The background imperative set  $\mathcal{I} = \{!(A), !(\neg A)\}$ , for example, will satisfy both  $\bigcirc A$  and  $\bigcirc \neg A$ , without allowing for the agglomeration of these individual oughts into the joint  $\bigcirc(A \wedge \neg A)$ . However, unlike the minimal logic, van Fraassen's approach does allow what seems to be the right degree of agglomeration. The imperative set  $\mathcal{I} = \{!(F \vee S), !(\neg F)\}$  will support, not only  $\bigcirc(F \vee S)$  and  $\bigcirc(\neg F)$ , but also the agglomerated  $\bigcirc((F \vee S) \wedge \neg F)$ , and so  $\bigcirc S$ .

Although this proposal of van Fraassen's captures an intuitively attractive and stable account of reasoning with conflicting norms, and although, by the time of van Fraassen's paper, the general topic of normative conflict had been an issue of intense concern in philosophy for well over a decade, it is hard to find any discussion of van Fraassen's logical proposal in the literature of the period.<sup>15</sup> Why? When I first began to think about these issues, my hypothesis was that this neglect resulted from the fact that both philosophers and logicians were accustomed to approaching deontic logic as a form of modal logic – basically, from the perspective of classical semantics – and, as I thought, van Fraassen's proposal did not fit naturally within that framework. Instead, I argued that the proposal was best understood within the general framework of nonmonotonic logic, and I offered two particular interpretations: one within default logic, and one within a simple model preference logic.<sup>16</sup> Of these, the interpretation within default logic has proved to be more fruitful, and I will now describe this formalism in just enough detail to be able to show how that interpretation works.

#### 4. *A simple default logic*

Where  $A$  and  $B$  are statements, let us take  $A \rightarrow B$  as the *default rule* that allows us to conclude  $B$ , by default, once  $A$  has been

established. To illustrate: if we suppose that  $B$  is the statement that Tweety is a bird and  $F$  the statement that Tweety can fly, then  $B \rightarrow F$  is the rule that allows us to conclude that Tweety can fly, by default, once it has been established that Tweety is a bird. We assume two functions – *Premise* and *Conclusion* – that pick out the premises and conclusions of default rules: if  $r$  is the default  $A \rightarrow B$ , then  $Premise(r)$  is the statement  $A$  and  $Conclusion(r)$  is the statement  $B$ . The latter function is lifted from individual defaults to sets of defaults in the obvious way: where  $\mathcal{S}$  is a set of defaults,  $Conclusion(\mathcal{S}) = \{Conclusion(r) : r \in \mathcal{S}\}$  is the set of conclusions of those defaults belonging to  $\mathcal{S}$ .

Default rules can be thought of as expressing the *reason relation*. In the case of our example, what the default  $B \rightarrow F$  indicates is that the premise that Tweety is a bird functions as a reason for the conclusion that Tweety flies.<sup>17</sup>

Some defaults, as well as their corresponding reasons, have greater weight, or higher priority, than others. This information is represented through an ordering  $<$  on default rules, where the statement  $r < r'$  means that the default  $r'$  has a higher priority than the default  $r$ . Suppose, for example, that  $P$  is the statement that Tweety is a penguin, so that  $P \rightarrow \neg F$  is the default allowing us to conclude that Tweety cannot fly once it is established that Tweety is a penguin. Then if we take  $r_1$  as the earlier default  $B \rightarrow F$  and  $r_2$  as this new default, it is natural to assume that  $r_1 < r_2$ .

I will focus to begin with on *fixed priority* default theories – theories, that is, in which all priorities among default rules are fixed in advance. Such a theory is a structure of the form  $\Delta = \langle \mathcal{W}, \mathcal{D}, < \rangle$ , in which  $\mathcal{W}$  is a set of ordinary statements,  $\mathcal{D}$  is a set of default rules, and  $<$  is a strict partial ordering on  $\mathcal{D}$ , representing priority.

Defaults are often thought of as special rules of inference that can be used to extend the conclusions derivable from a body of hard information beyond its ordinary logical consequences, and for this reason, the conclusion sets supported by default theories are generally referred to as *extensions*. These will be defined in terms of scenarios, where a *scenario* based on a default theory  $\Delta = \langle \mathcal{W}, \mathcal{D}, < \rangle$  is simply some subset  $\mathcal{S}$  of the set  $\mathcal{D}$  of defaults contained in that theory. From an intuitive standpoint, a scenario is supposed to represent the particular subset of default rules that have actually been selected by the reasoning agent as providing sufficient support for their conclusions. Not every scenario based on a default theory is intuitively acceptable, of course; some might contain what seems to be the wrong selection of defaults. The goal, therefore, is to characterize the *proper scenarios* – those sets of defaults that could be accepted by an ideal reasoning agent based on the information contained in the original theory. Once a proper scenario  $\mathcal{S}$  based on the theory

$\Delta = \langle \mathcal{W}, \mathcal{D}, < \rangle$  has been identified, an extension for this theory can then be defined as the set

$$\mathcal{E} = Th(\mathcal{W} \cup Conclusion(\mathcal{S})),$$

containing the statements that result from combining the hard information from the default theory with the conclusions of the defaults from the proper scenario, and then closing under logical consequence.

These ideas can be illustrated by returning to our example, with  $r_1$  and  $r_2$  as the defaults  $B \rightarrow F$  and  $P \rightarrow \neg F$ . If we suppose that Tweety is both a bird and a penguin, the information from this example can be captured by the default theory  $\Delta_1 = \langle \mathcal{W}, \mathcal{D}, < \rangle$ , where  $\mathcal{W} = \{P, P \supset B\}$ , where  $\mathcal{D} = \{r_1, r_2\}$ , and where  $r_1 < r_2$ . The set  $\mathcal{W}$  contains the basic information that Tweety is a penguin, and that this entails the fact that he is a bird; the set  $\mathcal{D}$  contains the two defaults; and the ordering tells us that the default about penguins has higher priority than the default about birds. This theory allows four possible scenarios –  $\mathcal{S}_1 = \emptyset$ ,  $\mathcal{S}_2 = \{r_1\}$ ,  $\mathcal{S}_3 = \{r_2\}$ , and  $\mathcal{S}_4 = \{r_1, r_2\}$  – corresponding to the situations in which the reasoning agent endorses neither of the two available defaults, only the first default, only the second, or both. From an intuitive standpoint, though, it seems that the agent should endorse the default  $r_2$ , and only that default, leading to the conclusion that Tweety does not fly. Therefore, only the third of these four scenarios,  $\mathcal{S}_3 = \{r_2\}$ , should be classified as proper.

How, then, can we define the proper scenarios? The definition I offer depends on three initial concepts – triggering, conflict, and defeat.

The *triggered* defaults represent those that are applicable in the context of a particular scenario; they are defined as the defaults whose premises are entailed by that scenario – those defaults, that is, whose premises follow from the hard information belonging to the default theory together with the conclusions of the defaults already endorsed. More exactly, if  $\mathcal{S}$  is a scenario based on the theory  $\Delta = \langle \mathcal{W}, \mathcal{D}, < \rangle$ , the defaults triggered in this scenario are those belonging to

$$Triggered_{\mathcal{W}, \mathcal{D}}(\mathcal{S}) = \{r \in \mathcal{D} : \mathcal{W} \cup Conclusion(\mathcal{S}) \vdash Premise(r)\}.$$

To illustrate by returning to the Tweety example, suppose  $\mathcal{S}_1 = \emptyset$ . In the context of this scenario, both  $r_1$  and  $r_2$  are triggered, since  $\mathcal{W} \cup Conclusion(\mathcal{S}_1) \vdash Premise(r_1)$  and  $\mathcal{W} \cup Conclusion(\mathcal{S}_1) \vdash Premise(r_2)$ .

A default will be classified as *conflicted* in the context of a scenario if the agent is already committed to the negation of its conclusion – that is, the conflicted defaults in the context of the scenario  $\mathcal{S}$ , based on the theory  $\Delta = \langle \mathcal{W}, \mathcal{D}, < \rangle$ , are those belonging to

$$Conflicted_{\mathcal{W}, \mathcal{D}}(\mathcal{S}) = \{r \in \mathcal{D} : \mathcal{W} \cup Conclusion(\mathcal{S}) \vdash \neg Conclusion(r)\}.$$

This idea can be illustrated through another example. Suppose that  $Q$ ,  $R$ , and  $P$  are the statements that Nixon is a Quaker, that Nixon is a

Republican, and that Nixon is a pacifist; and let  $r_1$  and  $r_2$  be the defaults  $Q \rightarrow P$  and  $R \rightarrow \neg P$ , instances for Nixon of the generalizations that Quakers tend to be pacifists and that Republicans tend not to be pacifists. Then, since Nixon was, in fact, both a Quaker and a Republican, we can represent an agent's information through the theory  $\Delta_2 = \langle \mathcal{W}, \mathcal{D}, < \rangle$ , where  $\mathcal{W} = \{Q, R\}$ , where  $\mathcal{D} = \{r_1, r_2\}$ , and where  $<$  is empty, since neither default has a higher priority than the other. Now imagine that, on whatever grounds, the agent decides to endorse one of these two defaults – say  $r_1$ , supporting the conclusion  $P$  – and is therefore reasoning in the context of the scenario  $\mathcal{S}_1 = \{r_1\}$ . In this context, the other default –  $r_2$ , supporting the conclusion  $\neg P$  – will be conflicted, since  $\mathcal{W} \cup \text{Conclusion}(\mathcal{S}_1) \vdash \neg \text{Conclusion}(r_2)$ .

Although the concept of *defeat* is surprisingly difficult to define in full generality, the basic idea is simple enough, and can serve as the basis of a preliminary definition.<sup>18</sup> Very roughly, a default will be classified as defeated in the context of a scenario  $\mathcal{S}$ , based on the theory  $\Delta = \langle \mathcal{W}, \mathcal{D}, < \rangle$ , whenever there is a stronger triggered default that supports a conflicting conclusion – whenever, that is, the default belongs to the set

$\text{Defeated}_{\mathcal{W}, \mathcal{D}, <}(\mathcal{S}) = \{r \in \mathcal{D} : \text{there is a default } r' \in \text{Triggered}_{\mathcal{W}, \mathcal{D}}(\mathcal{S}) \text{ such that}$

$$(1) r < r',$$

$$(2) \mathcal{W} \cup \{\text{Conclusion}(r')\} \vdash \neg \text{Conclusion}(r)\}.$$

This idea can be illustrated by returning to the Tweety example, the theory  $\Delta_1 = \langle \mathcal{W}, \mathcal{D}, < \rangle$ , where  $\mathcal{W} = \{P, P \supset B\}$ , where  $\mathcal{D} = \{r_1, r_2\}$  with  $r_1$  and  $r_2$  as the defaults  $B \rightarrow F$  and  $P \rightarrow \neg F$ , and where  $r_1 < r_2$ . We can suppose once again that the agent has not yet endorsed either of the two defaults, so that the initial scenario is  $\mathcal{S}_1 = \emptyset$ . In this situation, the default  $r_1$  is defeated, since  $r_2$  is triggered, and we have both (1)  $r_1 < r_2$  and (2)  $\mathcal{W} \cup \text{Conclusion}(r_2) \vdash \neg \text{Conclusion}(r_1)$ .

Once the underlying notions of triggering, conflict, and defeat are in place, we can define the notion of a default that is *binding* in the context of the scenario  $\mathcal{S}$ , based on the theory  $\Delta = \langle \mathcal{W}, \mathcal{D}, < \rangle$ , as one that is triggered in this context, but neither conflicted nor defeated – as a default, that is, belonging to the set

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

The concept can again be illustrated with Tweety, under the assumption that the agent's scenario is  $\mathcal{S}_1 = \emptyset$ . Here, the default  $r_1$ , supporting the conclusion  $F$ , is triggered in the context of this scenario, and it is not conflicted, but as we have just seen, it is defeated by the default  $r_2$ ; and

so it is not binding. By contrast, the default  $r_2$ , supporting the conclusion  $\neg F$ , is likewise triggered, not conflicted, and in this case not defeated either. This default is, therefore, binding.

With these initial concepts in place, we can now turn to the notion of a proper scenario. In fact, there are again some complexities involved in the definition of this idea, though these need not concern us here.<sup>19</sup> We can therefore work with a preliminary definition of a proper scenario, based on a theory  $\Delta = \langle \mathcal{W}, \mathcal{D}, < \rangle$ , as one containing all and only the defaults that are binding in the context of that very scenario – a definition, that is, according to which a scenario  $\mathcal{S}$  is classified as proper just in case

$$\mathcal{S} = \text{Binding}_{\mathcal{W}, \mathcal{D}, <}(\mathcal{S}).$$

We can think of the defaults from a proper scenario as presenting, not just reasons, but good reasons, in the context of that scenario. An agent who has accepted a set of defaults forming a proper scenario, then, is in an enviable position. Such an agent has already accepted all and only those defaults that it recognizes as presenting good reasons; the agent, therefore, has no incentive either to abandon any of the defaults already accepted, or to accept any others.

This concept can be illustrated by returning to the Tweety example – once again, the theory  $\Delta_1 = \langle \mathcal{W}, \mathcal{D}, < \rangle$  where  $\mathcal{W} = \{P, P \supset B\}$ , where  $\mathcal{D} = \{r_1, r_2\}$  with  $r_1$  as  $B \rightarrow F$  and  $r_2$  as  $P \rightarrow \neg F$ , and where  $r_1 < r_2$ . We noted earlier that, of the four possible scenarios based on this theory, only the third – that is,  $\mathcal{S}_3 = \{r_2\}$  – seemed attractive from an intuitive point of view; and with our definitions in place, the reader can now verify that only  $\mathcal{S}_3$  is proper. Following our earlier recipe for constructing extensions from proper scenarios, we can now see that this default theory yields the unique extension

$$\begin{aligned} \mathcal{E}_3 &= Th(\mathcal{W} \cup Conclusion(\mathcal{S}_3)) \\ &= Th(\{P, P \supset B, \neg F\}), \end{aligned}$$

supporting the conclusion  $\neg F$ , that Tweety cannot fly.

Things do not always work out so nicely, however. In contrast to the situation in ordinary logic, where any premise set leads to a unique set of conclusions, default theories can yield multiple extensions. This is illustrated by our Nixon example: the theory  $\Delta_2 = \langle \mathcal{W}, \mathcal{D}, < \rangle$  where  $\mathcal{W} = \{Q, R\}$ , where  $\mathcal{D} = \{r_1, r_2\}$  with  $r_1$  as  $Q \rightarrow P$  and  $r_2$  as  $R \rightarrow \neg P$ , and where  $<$  is empty. As the reader can verify, this theory allows two proper scenarios, both  $\mathcal{S}_1 = \{r_1\}$  and  $\mathcal{S}_2 = \{r_2\}$ , leading to  $\mathcal{E}_1 = Th(\{Q, R, P\})$  and  $\mathcal{E}_2 = Th(\{Q, R, \neg P\})$  as extensions. Both of these extensions contain  $Q$  and  $R$ , the initial information that Nixon is a Quaker and a Republican, but

the first contains  $P$ , the statement that he is a pacifist, while the second contains  $\neg P$ , the statement that he is not. In light of these two extensions, what should the reasoning agent actually conclude from the original default theory: is Nixon a pacifist or not?

In cases like this, when a default theory leads to multiple extensions, it is hard to decide what conclusions a reasoner should actually draw from the information contained in the theory. Two broad strategies have been suggested in the literature. According to the first, sometimes described as the *credulous* strategy, the reasoner should arbitrarily select one of the theory's several extensions and endorse the conclusions contained in that extension; according to the second, often described as the *skeptical* strategy, the reasoner should endorse a conclusion only if it is contained in the intersection of the theory's extensions.<sup>20</sup> For the purpose of modeling commonsense reasoning, the multiple extensions associated with default theories can sometimes seem like an embarrassment: what we really want is a unique conclusion set, and so we are forced either to select nondeterministically from among these various extensions provided by the theory, or else to combine them, somehow, into a unique set. When it comes to interpreting deontic ideas, however, the multiple extensions provided by default logic are no longer embarrassing. They give us, as we will see, exactly what is needed.

## 5. Connections

Often, default rules seem to represent something like commonsense probabilistic generalizations. The defaults concerning birds or Quakers, for instance, seem to mean simply that a large majority of birds can fly, or that a large majority of Quakers are pacifists. The connection between defaults and generalizations of this kind has suggested to many that default reasoning can best be understood as a kind of qualitative probabilistic reasoning.

There are also, however, important examples of default reasoning that do not fit so naturally into a probabilistic framework. The presumption of innocence in a legal system, for example, is a kind of default that overrides probabilistic considerations: even if the most salient reference class to which an individual belongs is one among which the proportion of criminals is very high, we are to presume that the individual has committed no crime unless there is conclusive evidence to the contrary.<sup>21</sup> Or on a more prosaic level, if you have promised to meet a friend for dinner, it seems right to conclude, by default, that meeting your friend is something you ought to do, unless exceptional circumstances interfere; and it is hard to see how probabilistic considerations could have anything to do with this conclusion.

What these examples suggest is that default rules can be used to represent *norms* quite generally. When the norms involved have a probabilistic basis, as many do, it is natural to expect default reasoning to resemble probabilistic reasoning. But default rules can also be used to represent other kinds of norms – such as legal or ethical norms – and in that case, any relation with probabilistic reasoning will be more distant.

It is this understanding of defaults, as representing norms in general, that motivated the deontic interpretation of default logics, leading to the initial observation that, if the norms carried by a set of imperatives are represented through default rules, then default logic can be used to interpret van Fraassen's account of conflicting oughts. Formally, where  $\mathcal{I}$  is a set of imperatives, we can define the *default transform* of  $\mathcal{I}$  as the theory  $\Delta_{\mathcal{I}} = \langle \mathcal{W}, \mathcal{D}, < \rangle$  where  $\mathcal{W} = \emptyset$ , where  $\mathcal{D} = \{ \top \rightarrow B : \neg(B) \in \mathcal{I} \}$  with  $\top$  a trivial truth, and where  $<$  is empty. We then have:

**Fact 5** If  $\mathcal{I}$  is a set of imperatives and  $\Delta_{\mathcal{I}}$  is its default transform, then  $\mathcal{I} \models_F \bigcirc A$  if and only if  $A \in \mathcal{E}$  for some extension  $\mathcal{E}$  of  $\Delta_{\mathcal{I}}$ .

Once a set of imperatives is transformed into a default theory, that is, the extensions of that default theory contain exactly those statements that van Fraassen's account supports as oughts, on the basis of the original set of imperatives. The default theory yields multiple extensions just in case the original set of imperatives supports conflicting oughts.

We can describe the default theories that result from interpreting sets of imperatives as *imperative default theories*. These theories are very simple, of course, but the normative interpretation can be generalized to richer theories as well – theories of the form  $\Delta = \langle \mathcal{W}, \mathcal{D}, < \rangle$  in which the hard information from  $\mathcal{W}$  might not be empty, the defaults from  $\mathcal{D}$  might have nontrivial premises, and there might be real priority relations among them. Based on these richer default theories, we can define two different evaluation rules for the deontic operator. The first, indicated by the satisfaction relation  $\vdash_C$  and generalizing van Fraassen's theory, is known as the *conflict* account; this interpretation classifies a statement as an ought as long as it is present in some extension of the default theory. The second, indicated by the satisfaction relation  $\vdash_D$ , is known as the *disjunctive* account; this interpretation does not classify a statement as an ought unless it is present in each extension.

**Definition 5 (Default evaluation rules:  $\bigcirc A$ )** Let  $\Delta$  be a default theory. Then

- $\Delta \vdash_C \bigcirc A$  if and only if  $A \in \mathcal{E}$  for some extension  $\mathcal{E}$  of  $\Delta$ ,
- $\Delta \vdash_D \bigcirc A$  if and only if  $A \in \mathcal{E}$  for each extension  $\mathcal{E}$  of  $\Delta$ .

The differences between these two interpretations can be illustrated by considering a situation in which I am – due to inadvertently conflicting



promises – committed to having a private dinner with each of two twins, but cannot have dinner with both. Suppose  $D_1$  and  $D_2$  are the statements that I dine with the first twin, or the second. The situation can then be represented by the theory  $\Delta_3 = \langle \mathcal{W}, \mathcal{D}, < \rangle$  where  $\mathcal{W} = \{ \neg(D_1 \wedge D_2) \}$ , where  $\mathcal{D} = \{r_1, r_2\}$ , with  $r_1$  and  $r_2$  as  $\top \rightarrow D_1$  and  $\top \rightarrow D_2$ , and where the ordering  $<$  is empty. Here, the defaults from  $\mathcal{D}$  represent the two norms, or imperatives, urging a dinner with each of the twins;  $\mathcal{W}$  contains the hard information that I cannot dine with both; and  $<$  is empty, since neither default has higher priority. This theory yields two proper scenarios  $\mathcal{S}_1 = \{\delta_1\}$  and  $\mathcal{S}_2 = \{\delta_2\}$ , generating the two extensions

$$\begin{aligned}\mathcal{E}_1 &= Th(\{ \neg(D_1 \wedge D_2), D_1 \}), \\ \mathcal{E}_2 &= Th(\{ \neg(D_1 \wedge D_2), D_2 \}).\end{aligned}$$

Since  $D_1$  belongs to the first of these extensions and  $D_2$  to the second, the conflict account supports both of the conflicting oughts  $\bigcirc D_1$  and  $\bigcirc D_2$ , telling me that I ought to dine with the first twin, and with the second as well, in spite of the fact that I cannot dine with both. According to the disjunctive account, on the other hand, neither  $\bigcirc D_1$  nor  $\bigcirc D_2$  is supported, since neither  $D_1$  nor  $D_2$  belongs to each extension of the theory. But of course, since each extension does contain either  $D_1$  or  $D_2$ , and since extensions are closed under consequence, each extension also contains  $D_1 \vee D_2$ . As a result, the disjunctive account yields  $\bigcirc(D_1 \vee D_2)$ , telling me that I ought to have dinner with one twin or another, at least. This particular example indicates the general pattern: where the conflict account yields normative conflicts, the disjunctive account yields only disjunctive oughts.

Default logic, then, offers a framework within which van Fraassen's account of reasoning with normative conflicts can be interpreted and also generalized – in the ways we have seen, and in other ways to be considered shortly. The framework is based, furthermore, on a very different picture, or ideology, than that provided by classical semantics for the foundations of deontic logic, and one with important connections to recent work in ethical theory: if we accept the interpretation of defaults as representing norms, or reasons more generally, then default logic can be taken to show how ought statements are grounded, not in a preference ordering among works, but in facts about reasons and their interactions.<sup>22</sup>

Do we now have, therefore, a convincing motive for abandoning classical semantics – since the classical framework cannot provide a sensible treatment of normative conflicts, while default logic can?

Well, no, because it turns out that van Fraassen's account can also be interpreted within the classical framework. To begin with, while continuing to work with Kratzer models, let us replace Kratzer's original evaluation rule for the ought operator, set out in Definition 2, with the



following variant, specifying the new satisfaction relation  $\models_{KC}$ , which allows conflicting oughts.

**Definition 6 (Kratzer conflict evaluation rule:  $\bigcirc A$ )** Where  $M = \langle W, f, g, v \rangle$  is a Kratzer model and  $\alpha$  is a world from  $W$ ,

- $\mathcal{M}, \alpha \models_{KC} \bigcirc A$  if and only if there is a  $\gamma \in \bar{f}(\alpha)$  such that, for all  $\delta \in \bar{f}(\alpha)$ , if  $\delta \leq_{g(\alpha)} \gamma$ , then  $\delta \in |A|_{\mathcal{M}}$ .

The intuition behind this new rule is that  $\bigcirc A$  holds just in case there is some world from the modal base such that  $A$  then holds in every world from the modal base at least as good as that one.<sup>23</sup>

There is a wrinkle when it comes to interpreting imperative sets into Kratzer models. Van Fraassen sets out his theory in terms of classical models, rather than possible worlds within a modal model, and so works with a notion of possibility as logical consistency. To avoid any difficulties resulting from this choice, we require our models to be built around *rich* sets of worlds, where these are defined as those containing, for each subset of atomic formulas from the background language, a world in which all and only the formulas from that subset hold; a modal model built on a rich set of worlds thus contains possible worlds corresponding to each of the classical models of the underlying language. Given a set of imperatives  $\mathcal{I}$ , then, a *Kratzer transform* of  $\mathcal{I}$  can be defined as a Kratzer model  $\mathcal{M}_{\mathcal{I}} = \langle W, f, g, v \rangle$  in which  $W$  is a rich set of worlds,  $v$  is a valuation,  $f(\alpha) = \{W\}$ , and  $g(\alpha) = \{|B|_{\mathcal{M}} : (B) \in \mathcal{I}\}$  for each  $\alpha$  from  $\mathcal{W}$  – that is, the modal base  $\bar{f}(\alpha) = \bigcap f(\alpha)$  is the entire set of worlds, and the ordering source  $g(\alpha)$  contains a norm proposition corresponding to each imperative from  $\mathcal{I}$ .<sup>24</sup>

Since the usual language of deontic logic allows for nested oughts, and van Fraassen's theory does not, comparisons can be drawn only for nonnested oughts – those of the form  $\bigcirc A$  in which  $A$  does not itself contain an ought operator. If we limit ourselves to statements like this, it then follows that, when ought statements are evaluated in accord with the Kratzer conflict evaluation rule, any Kratzer transform of a set of imperatives supports the same ought statements as the original imperative set:

**Fact 6** *Given a set  $\mathcal{I}$  of imperatives, if  $\mathcal{M}_{\mathcal{I}}$  is a Kratzer transform of this set with  $\alpha$  a world from this model, and  $\bigcirc A$  is nonnested, then  $\mathcal{I} \models_F \bigcirc A$  if and only if  $\mathcal{M}_{\mathcal{I}}, \alpha \models_{KC} \bigcirc A$ .*

Combining Facts 5 and 6, we can see that, where  $\mathcal{I}$  is a set of imperatives,  $\Delta_{\mathcal{I}}$  is its default transform,  $\mathcal{M}_{\mathcal{I}}$  is a Kratzer transform with  $\alpha$  a

world from that model, and limiting ourselves to nonnested oughts, the following three statements are equivalent:

$$\begin{aligned}\mathcal{I} &\models_F \bigcirc A, \\ \Delta_{\mathcal{I}} &\vdash_C \bigcirc A, \\ \mathcal{M}_{\mathcal{I}}, \alpha &\models_{KC} \bigcirc A.\end{aligned}$$

So these three perspectives on normative conflict – van Fraassen’s original theory, the conflict interpretation of default logic, and classical semantics under the Kratzer conflict evaluation rule – all coincide. Furthermore, for theories resulting from imperative sets, the oughts supported by the disjunctive interpretation of default logic coincide with those supported by the standard Kratzer evaluation rule:

**Fact 7** *Given a set  $\mathcal{I}$  of imperatives, if  $\Delta_{\mathcal{I}}$  is the default transform of this set,  $\mathcal{M}_{\mathcal{I}}$  is a Kratzer transform with  $\alpha$  a world from this model, and  $\bigcirc A$  is nonnested, then  $\Delta_{\mathcal{I}} \vdash_D \bigcirc A$  if and only if  $\mathcal{M}_{\mathcal{I}}, \alpha \models_{KD} \bigcirc A$ .*

Combining this fact with the previous Fact 3, we can now see that, where  $\mathcal{I}$  is a set of imperatives,  $\Delta_{\mathcal{I}}$  is its default transform,  $\mathcal{M}_{\mathcal{I}}$  is a Kratzer transform with  $\alpha$  a world from that model, and  $\mathcal{M}_{\mathcal{I}}'$  is the standard transform of  $\mathcal{M}_{\mathcal{I}}$ , the following statements are equivalent as well:

$$\begin{aligned}\mathcal{M}_{\mathcal{I}}', \alpha &\models_{SDL} \bigcirc A, \\ \Delta_{\mathcal{I}} &\vdash_D \bigcirc A, \\ \mathcal{M}_{\mathcal{I}}, \alpha &\models_{KD} \bigcirc A.\end{aligned}$$

Again, these three perspectives coincide – the disjunctive interpretation of default logic, classical semantics under the standard Kratzer evaluation rule, and standard deontic logic applied to standard transforms of Kratzer models.

We can now draw some preliminary conclusions. First, there appear to be two broad strategies for reasoning with conflicting norms: one sees conflicting norms as generating conflicting oughts, while the other avoids direct conflicts in favor of disjunctive oughts. Second, although the frameworks provided by default logic and classical semantics reflect different pictures, or ideologies, concerning the foundations of deontic logic – centered around reasons and their interaction, or around an ordering on possible worlds – each of these two broad strategies regarding conflicting norms can be formulated in either framework. The upshot is that, as long as we restrict ourselves to simple imperative default theories – and setting aside any concerns with foundational matters, or ideology – it is hard to see, at this point, what advantages could be derived from abandoning the familiar framework of classical semantics for the new framework of default logic.

## 6. *Priorities and conditional oughts*

But of course, default theories can be more complex than the simple imperative theories considered so far. We begin here by considering theories in which, although each default rule still contains a trivial premise, the ordering among defaults is no longer required to be empty, so that there is a real prioritization on norms; next, we will relax the requirement that the premises of default rules must be trivial, and consider conditional oughts. To illustrate the first kind of case, imagine that my Chair asks me to attend the faculty meeting while my Dean asks me to attend the senate meeting, but because the meetings are at the same time, I cannot attend both. Let  $F$  and  $S$  be the statements that I attend the faculty meeting and the senate meeting, respectively. The situation can then be represented by the theory  $\Delta_4 = \langle \mathcal{W}, \mathcal{D}, < \rangle$  where  $\mathcal{W} = \{ \neg(F \wedge S) \}$  and where  $\mathcal{D} = \{r_1, r_2\}$ , with  $r_1$  as  $\top \rightarrow F$  and  $r_2$  as  $\top \rightarrow S$ . So far this looks like a simple case of conflicting norms, but if we suppose also that  $r_1 < r_2$  – since the Dean outranks the Chair, so that a request from the Dean carries more weight than a request from the Chair – then it seems to follow that I ought to attend the senate meeting, rather than the faculty meeting. To illustrate the second kind of case, imagine that my Chair, who is afraid I will embarrass the Department, asks me not to attend the Provost's reception, that my Dean, who would like the College to be represented, asks me to attend, and that the Provost, who does not care whether I attend but would like the event to be classy, asks me to wear a suit if I do attend. Then it seems to follow that I ought to attend, that I ought to wear a suit if I attend, and that I ought to wear a suit.

Now we must ask, do the more complex default theories necessary for representing information like this provide any motive for abandoning classical semantics – do they allow us to understand any new aspects of normative reasoning that cannot be represented equally well within the classical approach?

Beginning with the first kind of case, we can recall that the treatment of priorities is already built into prioritized default logic. Even the simple default logic sketched in Section 4 allows us to conclude that  $\Delta_4$  supports  $\bigcirc S$  but not  $\bigcirc F$  – that I ought to go to the senate meeting, rather than the faculty meeting. Of course, the treatment of priorities in this simple logic is oversimplified in various ways, but a number of more sophisticated theories have been proposed.<sup>25</sup> Although the merits and demerits of these various proposals are still a matter of dispute, it is clear that default logic offers a rich framework in which different ideas concerning prioritized norms, or reasons, can be investigated.

Now, what about classical semantics? From this perspective, a priority ranking among defaults with trivial premises is like a priority ranking among propositions belonging to the ordering source; and in fact, several

writers from the classical tradition have explored the idea that the propositions, or norms, in the ordering source should be prioritized.<sup>26</sup> This work, however, tends to differ in two ways from that pursued within the framework of default logic. First, in the case of deontic logics based on default logic, there is, so far, only a single deontic modality at issue, while the work on prioritization in the classical tradition is generally developed to help us understand the relations among different modalities – such as weak and strong oughts, or gradable modalities. Second, while prioritized defaults are thought of as arranged only in a partial order, the prioritization of norms from an ordering source, in the classical tradition, is often taken, for one reason or another, to be total.

Still, in spite of these differences in the body of existing work – surely resulting only from historical accidents and local disciplinary concerns – there is nothing to prevent the development of a classical theory that parallels prioritized default logic exactly, with a partially ordered set of norms all bearing on a single ought.<sup>27</sup> To illustrate, let us define a *prioritized Kratzer model* as a structure like a Kratzer model but where the propositions, or norms, in the ordering source are partially ordered – as a structure  $\mathcal{M} = \langle W, f, g, v \rangle$ , that is, with  $W, f$ , and  $v$  as before, but where, for any world  $\alpha$ , we now have  $g(\alpha) = \langle \mathcal{G}, < \rangle$ , with  $\mathcal{G}$  a set of propositions and  $<$  a strict partial ordering on this set. Now, how can we derive an ordering on worlds from the ordered set of propositions  $\langle \mathcal{G}, < \rangle$ , which constitutes the new ordering source?

One natural option – there are many others – is to order the worlds lexically, with respect to the ordered set of propositions, systematically favoring worlds that satisfy more important propositions. To implement this idea, we first define, for  $\mathcal{H}$  a set of propositions ordered by  $<$ , the set

$$\max_{<}(\mathcal{H}) = \{X \in \mathcal{H} : \neg \exists Y \in \mathcal{H} (X < Y)\}$$

containing those propositions that are maximal, or undominated – that is, the most important propositions from  $\mathcal{H}$  with respect to  $<$ . Beginning, then, with the set  $\mathcal{G}$  of propositions from the prioritized ordering source  $g(\alpha) = \langle \mathcal{G}, < \rangle$ , we can define a sequence  $\mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_2, \dots$  of sets, with  $\mathcal{G}$  itself as the first member of the sequence and each succeeding member containing all but the most important propositions from its predecessor:

$$\mathcal{G}_0 = \mathcal{G},$$

$$\mathcal{G}_{i+1} = \mathcal{G}_i - \max_{<}(\mathcal{G}_i).$$

We can then define a sequence  $\preceq_0, \preceq_1, \preceq_2, \dots$  of ordering relations according to which the world  $\beta$  is at least as good as the world  $\gamma$  at the initial stage just in case  $\beta$  satisfies all of the most important propositions that  $\gamma$  does from the first set in the previous sequence, and at each

successor stage, just in case  $\beta$  is at least as good as  $\gamma$  at the predecessor stage and, in addition,  $\beta$  satisfies all of the most important propositions  $\gamma$  does that remain in the set corresponding to that stage:

$$\beta \preceq_{\mathcal{G}_0} \gamma \text{ if and only if } \beta \leq_{\max_{<}(\mathcal{G}_0)} \gamma,$$

$$\beta \preceq_{\mathcal{G}_{i+1}} \gamma \text{ if and only if } \beta \preceq_{\mathcal{G}_i} \gamma \text{ and } \beta \leq_{\max_{<}(\mathcal{G}_{i+1})} \gamma.$$

Finally, we can stipulate that  $\beta$  is at least as good as  $\gamma$  according to the prioritized ordering source  $g(\alpha) = \langle \mathcal{G}, < \rangle$  just in case that relation holds for each member of the previous sequence of orderings, defined in terms of this ordering source:

$$\beta \preceq_{g(\alpha)} \gamma \text{ if and only if } \beta \preceq_{\mathcal{G}_i} \gamma \text{ for each } i.$$

Once we see how the worlds can be ordered on the basis of a partial ordering of propositions, it is easy enough to adapt the previous Kratzer evaluation rules to fit the prioritized setting. Although the changes are routine – merely replacing the original ordering relation  $\leq_{g(\alpha)}$  in these rules with the new relation  $\preceq_{g(\alpha)}$ , derived from a prioritized ordering source – I will nevertheless display the crucial clauses for the new satisfaction relations  $\models_{PKD}$  and  $\models_{PKC}$ , prioritized versions of the previous disjunctive and conflict Kratzer evaluation rules, from Definitions 2 and 6 respectively.

**Definition 7 (Prioritized Kratzer evaluation rules:  $\circ A$ )** Where  $\mathcal{M} = \langle W, f, g, v \rangle$  is a prioritized Kratzer model and  $\alpha$  is a world from  $W$ ,

- $\mathcal{M}, \alpha \models_{PKD} \circ A$  if and only if, for all  $\beta \in \bar{f}(\alpha)$ , there is a  $\gamma \in \bar{f}(\alpha)$  such that  $\gamma \preceq_{g(\alpha)} \beta$  and, for all  $\delta \in \bar{f}(\alpha)$ , if  $\delta \preceq_{g(\alpha)} \gamma$ , then  $\delta \in |A|_{\mathcal{M}}$ ,
- $\mathcal{M}, \alpha \models_{PKC} \circ A$  if and only if there is a  $\gamma \in \bar{f}(\alpha)$  such that, for all  $\delta \in \bar{f}(\alpha)$ , if  $\delta \preceq_{g(\alpha)} \gamma$ , then  $\delta \in |A|_{\mathcal{M}}$ .

Of course, since the new relation  $\preceq_{g(\alpha)}$  collapses into the previous  $\leq_{g(\alpha)}$  in case the ordering  $<$  from the ordering source  $g(\alpha) = \langle \mathcal{G}, < \rangle$  happens to be empty, these two definitions are conservative generalizations of the previous versions.

Again, this lexical proposal illustrates just one way of handling priorities among norms from an ordering source within the classical framework. There are others as well, and it would be an interesting project, not just to explore the range of these prioritized classical theories, but to compare them with some of the different accounts developed within the extensive literature on prioritization in default logics.<sup>28</sup> My purpose in presenting

this particular proposal is not to endorse it, but only to show that, as long as we restrict ourselves to defaults with trivial premises, the kinds of situations representable in prioritized default logics can be handled in the classical framework as well – with the result that, once again, we have found no real reason for abandoning classical semantics.

Let us now, however, drop the restriction that defaults must have trivial premises, and focus, in this more general setting, on conditional ought statements of the form  $\bigcirc(A/B)$ , taken to mean that  $A$  ought to be the case in the circumstances  $B$ .

In the case of default logic, conditional oughts work like this. Start with a default theory  $\Delta = \langle \mathcal{W}, \mathcal{D}, < \rangle$ . To evaluate a simple ought  $\bigcirc A$ , we looked to see if  $A$  is contained in the extensions of this theory, either some or all, depending on whether we were working with the conflict or disjunctive account. With a conditional ought  $\bigcirc(A/B)$ , we now look to see if  $A$  is contained in the extensions, not of the original default theory, but of the new theory

$$\Delta[B] = \langle \mathcal{W} \cup \{B\}, \mathcal{D}, < \rangle,$$

arrived at by supplementing the hard information  $\mathcal{W}$  from the original theory  $\Delta$  with the antecedent  $B$  of the conditional. The default treatment of conditional oughts, both disjunctive and conflict, can now be presented by generalizing the treatment of categorical oughts from Definition 5 as follows.

**Definition 8 (Default evaluation rules:  $\bigcirc(A/B)$ )** Let  $\Delta$  be a default theory. Then

- $\Delta \vdash_D \bigcirc(A/B)$  if and only if  $\Delta[B] \vdash_D \bigcirc A$ ,
- $\Delta \vdash_C \bigcirc(A/B)$  if and only if  $\Delta[B] \vdash_C \bigcirc A$ .

Moving to the classical setting, and starting with a Kratzer model  $\mathcal{M} = \langle W, f, g, v \rangle$ , we recall that, to see if  $\bigcirc A$  held at a particular world  $\alpha$  from  $W$ , we focused – in different ways, depending on whether we were working with the original Kratzer evaluation rule or its conflict variant – on the worlds from the modal base  $\bar{f}(\alpha)$  that were classified as favorable according to the ordering  $\leq_{g(\alpha)}$ , derived from the ordering source. To evaluate a conditional ought  $\bigcirc(A/B)$ , we now focus – in exactly the same ways – not on the favorable worlds from the entire modal base, but only on those favorable worlds from the modal base in which  $B$  is true as well. This idea is implemented by taking  $\mathcal{M}[B] = \langle W, f_B, g, v \rangle$  with  $W, g$ , and  $v$  as in  $\mathcal{M}$  and with  $f_B$  defined so that

$$f_B(\alpha) = f(\alpha) \cup \{B|_{\mathcal{M}}\}$$

for each world  $\alpha$  from  $W$ ; as a result, we also have  $\overline{f_B}(\alpha) = \overline{f}(\alpha) \cap |B|_{\mathcal{M}}$ . The original and conflict Kratzer evaluation rules can now be generalized to cover conditional oughts as follows.

**Definition 9 (Kratzer evaluation rules:  $\bigcirc(A/B)$ )** Where  $\mathcal{M} = \langle W, f, g, v \rangle$  is a Kratzer model and  $\alpha$  is a world from  $W$ ,

- $\mathcal{M}, \alpha \models_{KD} \bigcirc(A/B)$  if and only if  $\mathcal{M}[B], \alpha \models_{KD} \bigcirc A$ ,
- $\mathcal{M}, \alpha \models_{KC} \bigcirc(A/B)$  if and only if  $\mathcal{M}[B], \alpha \models_{KC} \bigcirc A$ .

Set out in this way, the approaches to conditional oughts in these two frameworks – default and classical – seem to follow the same path. The hard information from a default theory is like the modal base in a Kratzer model, registering the set of fixed assumptions against which reasoning about oughts takes place. In evaluating a conditional ought, each of these two approaches first supplements this set of fixed assumptions with the information carried by the antecedent of that ought; the conditional ought itself then holds if its consequent holds as a simple, or categorical, ought against the background of this supplemented set of assumptions. Each approach, that is, reduces the evaluation of conditional oughts to the evaluation of categorical oughts in the same way – and since, as we have seen, the treatment of categorical oughts is so similar in the default and classical frameworks, it may seem that their treatments of conditional oughts must coincide as well.

But this is not so, as a simple example shows. Suppose that, having read *Miss Manners*, I realize that I am subject to the two imperatives ‘Don’t eat with your fingers’ and ‘If you are served cold asparagus, eat it with your fingers.’<sup>29</sup> Taking  $F$  and  $A$  as the respective statements that I eat with my fingers and that I am served cold asparagus, the situation can be represented through the default theory  $\Delta_6 = \langle \mathcal{W}, \mathcal{D}, < \rangle$  where  $\mathcal{W} = \emptyset$ , where  $\mathcal{D} = \{r_1, r_2\}$  with  $r_1$  as  $\top \rightarrow \neg F$  and  $r_2$  as  $A \rightarrow F$ , and where  $r_1 < r_2$ . It is easy to see that this initial theory  $\Delta_6$  has the unique extension  $\mathcal{E}_1 = Th(\{\neg F\})$ , and also that the supplemented theory  $\Delta_6[A]$  has the unique extension  $\mathcal{E}_2 = Th(\{A, F\})$ . From Definitions 5 and 8 we can conclude that  $\Delta_6$  supports both the conclusions  $\bigcirc(\neg F)$  and  $\bigcirc(F/A)$  – that I ought not to eat with my fingers, but that, if I am served cold asparagus, I ought to eat it with my fingers.<sup>30</sup> On the other hand, the theory  $\Delta_6$  does not support  $\bigcirc(\neg A)$ , which seems right – it does not seem to follow from the imperatives I learned in the etiquette book that I ought not to be served cold asparagus.

In the classical framework, however, we cannot have  $\bigcirc(\neg F)$  and  $\bigcirc(F/A)$  without  $\bigcirc(\neg A)$ , because the first two statements entail the third. The entailment holds for all variants of the Kratzer evaluation rule considered in this article, but I will illustrate by appeal only to the original form, set



out in Definition 2, under the further assumption that we are working in a stoppered Kratzer model, so that ought statements can be evaluated in accord with Fact 1. Suppose, then, that  $\bigcirc(\neg F)$  and  $\bigcirc(F/A)$  hold at a world  $\alpha$  from a stoppered Kratzer model  $\mathcal{M} = \langle W, f, g, v \rangle$ , but that  $\bigcirc(\neg A)$  does not. Since  $\bigcirc(\neg F)$  holds, we have (1)  $Best_{g(\alpha)}(\bar{f}(\alpha)) \subseteq |\neg F|$ . Since  $\bigcirc(F/A)$  holds, and drawing on Definition 9 for the reduction of conditional to categorical oughts, we have (2)  $Best_{g(\alpha)}(\bar{f}_A(\alpha)) \subseteq |F|$ . But since  $\bigcirc(\neg A)$  fails, it is not the case that  $Best_{g(\alpha)}(\bar{f}(\alpha)) \subseteq |\neg A|$ , from which we can conclude that (3)  $Best_{g(\alpha)}(\bar{f}(\alpha)) \cap |A| \neq \emptyset$ . It is easy to verify that, when (3) holds, we have (4)  $Best_{g(\alpha)}(\bar{f}_A(\alpha)) = Best_{g(\alpha)}(\bar{f}(\alpha)) \cap |A|$ . But (4) leads to (5)  $Best_{g(\alpha)}(\bar{f}(\alpha)) \cap |A| \subseteq |F|$  when it is taken together with (2), and to (6)  $Best_{g(\alpha)}(\bar{f}(\alpha)) \cap |A| \subseteq |\neg F|$  when it is taken together with (1), since  $Best_{g(\alpha)}(\bar{f}(\alpha)) \cap |A| \subseteq Best_{g(\alpha)}(\bar{f}(\alpha))$ . And (5) and (6) entail (7)  $Best_{g(\alpha)}(\bar{f}(\alpha)) \cap |A| = \emptyset$ , which contradicts (3).

Another difference between the default and classical treatments of conditional oughts can be seen if we imagine that, through further study of *Miss Manners*, I learn that I am subject also to the imperative 'Put your napkin on your lap.' If we take  $N$  as the statement that I put my napkin on my lap, the situation can be represented through the default theory  $\Delta_7 = \langle \mathcal{W}, \mathcal{D}, < \rangle$ , in which  $\mathcal{W}$  and  $<$  are just as in  $\Delta_6$ , but where  $\mathcal{D} = \{r_1, r_2, r_3\}$ , with  $r_1$  and  $r_2$  as before and with  $r_3$  as the new default  $\top \rightarrow N$ . This initial theory  $\Delta_7$  has the unique extension  $\mathcal{E}_3 = Th(\{\neg F, N\})$ , while the supplemented theory  $\Delta_7[A]$  has the unique extension  $\mathcal{E}_4 = Th(\{A, F, N\})$ . We can thus conclude that  $\Delta_7$  supports  $\bigcirc(\neg F)$  and  $\bigcirc(F/A)$  as before, but now also  $\bigcirc(N)$  and  $\bigcirc(N/A)$ .

Again, this result seems to be correct: I ought to put my napkin on my lap whether or not I am served cold asparagus. But it is hard to see how to reach exactly this result within the classical framework. The problem is general. It does not depend on any particular mechanism for relating norms, or imperatives, to conditional oughts, but can be stated entirely in the language of dyadic deontic logic. To see this, suppose the defaults from  $\Delta_7$  are themselves coded as conditional oughts, yielding as premises the three statements  $\bigcirc(\neg F/\top)$ ,  $\bigcirc(F/A)$ , and  $\bigcirc(N/\top)$ . From these premises, we wish to derive the statement  $\bigcirc(N/A)$ , but not the statement  $\bigcirc(\neg F/A)$ . Now, any logic developed within the classical framework must either admit a rule of *antecedent strengthening*, of the form

$$\bigcirc(A/B) / \bigcirc(A/B \wedge C),$$

or not. If a logic admits this rule, then we can derive  $\bigcirc(N/A)$  by applying strengthening to the third premise, but in that case, strengthening applied to the first premise yields  $\bigcirc(\neg F/A)$ , which we do not want. If the logic fails to admit strengthening, on the other hand, then we are not



forced to conclude that  $\bigcirc(\neg F/A)$ , but without any strengthening at all, it is hard to see how the logic could yield  $\bigcirc(N/A)$ . This difficulty is reminiscent of that described earlier in our discussion of normative conflicts, where we noted the need for a certain amount of agglomeration, in order to support appropriate conclusions, but not too much, on pain of triviality. In this case, it seems that what we need is a certain amount of antecedent strengthening, but not too much: we want to allow norms formulated explicitly only for general circumstances to apply also in more specific situations, unless they are overridden in those situations.<sup>31</sup>

Here, then, with conditional oughts, we at last have a real difference between the approaches to deontic logic developed within the default and classical frameworks. The default approach allows combinations such as  $\bigcirc(\neg F)$  and  $\bigcirc(F/A)$  without  $\bigcirc(\neg A)$ , while the classical approach does not. The default approach allows for only a limited degree of antecedent strengthening, while the classical approach presents the stark options of unrestricted strengthening, or none at all.

What are we to make of these differences – should we say that one of these two approaches, default or classical, is right while the other is wrong? I do not think so. Instead, I think we should say that conditional oughts can be understood in two senses. There is, first of all, what I will call the *constrained optimality* sense, according to which the statement  $\bigcirc(A/B)$  is taken to mean, very roughly, that  $A$  holds in the best worlds in which  $B$  holds, even if these worlds are not among the best overall.<sup>32</sup> It is this sense of the conditional ought that was originally explored by Bengt Hansson, in early papers by van Fraassen, Chellas, and David Lewis, and in an extensive later literature, including, particularly, the work of Henry Prakken and Marek Sergot.<sup>33</sup> And it is this sense, also, that is captured in Kratzer's work.<sup>34</sup> The constrained optimality sense of the conditional ought provides a reading on which the asparagus inference really does seem to be correct: if the best worlds are those in which I do not eat with my fingers, but the best worlds in which I am served cold asparagus are those in which I eat with my fingers, then it follows at once that the worlds in which I am served cold asparagus are not among the best. And it provides a reading on which – as in the traditional dyadic deontic logics – antecedent strengthening is entirely ruled out, so that the napkin inference fails: even if the best worlds are those in which I put a napkin on my lap, since, as we have seen, the worlds in which I am served cold asparagus are not among the best, there is no way to conclude that I put a napkin on my lap in the best of those.

There is also, however, a different sense of the conditional ought, which I will refer to here, drawing on the ideas of W. D. Ross, as the *resultant* sense, and according to which, the statement  $\bigcirc(A/B)$  is taken to mean, again very roughly, that the various *prima facie* norms – or imperatives, or reasons – at work under the condition  $B$  interact in a way that results in

overall support for *A*.<sup>35</sup> And it is this sense that is captured by the framework of default logic. Rather than using the underlying norms, as in Kratzer's theory, only to define an ordering on worlds, and then evaluating oughts with respect to that ordering, default logic allows us to explore ways in which these norms, or reasons, might interact to support ought statements directly. The resultant sense of the conditional ought provides a reading on which it seems natural that the asparagus inference should fail: I have a generally applicable reason not to eat with my fingers, but once I am served cold asparagus – although there is nothing nonideal about that situation, and no reason why I should not be served cold asparagus – this fact triggers a reason for eating with my fingers that both defeats my previous reason not to eat with my fingers, and supports the conclusion that I ought to do so. And the resultant sense provides us, also, with a reading that supports limited antecedent strengthening, so that we can understand how the napkin inference is allowed: although being served cold asparagus triggers a reason that defeats previous my reason not to eat with my fingers, it does not interfere with my previous reason to put my napkin on my lap, so that the overall interaction among my various reasons results in continued support for the conclusion that I ought to do so, even under the condition that I am served cold asparagus.

Two questions now arise. First, even if the resultant sense of the conditional ought differs from the constrained optimality sense, is there any reason to think that the resultant sense cannot be captured within the classical framework? In fact, the two conditional oughts share certain logical properties, most notably failure of antecedent strengthening as a strict logical rule. Because of this, some writers have attempted to analyze the resultant ought using ideas introduced in the study of constrained optimization oughts – basically, a more complex preference ordering on worlds.<sup>36</sup> Although I can offer no argument to show that this cannot be done, I have not seen it done successfully, and I am skeptical: being served cold asparagus just does not seem like the sort of thing that should force us to consider nonideal worlds. Alternatively, it may be possible to enrich the ordering source from Kratzer models with something like pairs of propositions functioning as conditional norms, not just individual propositions as norms, and then to order worlds on the basis of these conditional norms in a way that yields attractive results. I am less skeptical of this approach, but would have to see the idea worked out in detail before being convinced of its viability. In any case, I think it is fair to say that the resultant sense of conditional oughts now provides us with a real reason for exploring the framework of default logic, certainly not in place of, but at least in addition to classical semantics.

The second question concerns language. Supposing I am right that conditional oughts can be taken in two senses – constrained optimization, or resultant – how is this difference registered linguistically? It is

implausible to suppose there is any semantic ambiguity in the statements we use to express conditional oughts. But then, if the matter is pragmatic, we would need a systematic account of the conditions under which an ought statement is to be interpreted in one way, rather than the other.<sup>37</sup> How can we tell, for example, that, when I say ‘Given that I have missed my plane, and cannot now visit my mother on her birthday, I ought to send flowers,’ my statement is to be interpreted in the constrained optimization sense – as offering a judgment about what is best in a nonideal situation – while my statement ‘Given that I have been served cold asparagus, I ought to eat it with my fingers’ is to be interpreted in the resultant sense?

## 7. Reasoning about reasons

I have suggested that a deontic logic developed within the framework of default logic captures a resultant sense of the conditional ought, according to which the truth or falsity of ought statements results from the interaction among reasons. Although we have considered, so far, only very simple relations among reasons – conflict, defeat – their interactions can be more complex. Sometimes, we reason about the weights, or priorities, to assign to other reasons; sometimes we conclude, on the basis of reasons, that other reasons must be removed from consideration entirely. In order to illustrate how this kind of reasoning about reasons can be accommodated within default logic, I sketch an account of the first of these phenomenon – reasoning about the priorities among other reasons.<sup>38</sup>

If reasons are provided by defaults, then reasons about the priorities among reasons must be modeled through defaults about the priority relations among defaults. To make sense of defaults like this, we must extend our previous treatment of default logic to show how the priorities that constrain the selection of a proper scenario can be established within that scenario itself. This may sound complicated, perhaps circular, but in fact it is straightforward, and can be explained in four steps.

The first step is to enrich our background language with the resources to enable formal reasoning about priorities among defaults: a new set of individual constants, to be interpreted as names of defaults, together with a relation symbol representing priority. For the sake of simplicity, we will assume that each of these new constants has the form  $n_X$ , for some subscript  $X$ , and that each such constant refers to the default  $r_X$ . And we will assume also that our language now contains the relation symbol  $\prec$ , representing priority among defaults.

To illustrate, suppose that  $r_1$  and  $r_2$  are the defaults  $A \rightarrow B$  and  $C \rightarrow \neg B$ , respectively, and that  $r_3$  is the priority default  $D \rightarrow r_1 \prec r_2$ . Then what  $r_3$  says is that  $D$  functions as a reason for assigning  $r_2$  a higher priority than

$r_1$ . As a result, we would expect that, when all three of these defaults are triggered – that is, when  $A$ ,  $C$ , and  $D$  all hold – the default  $r_1$  will generally be defeated by  $r_2$ , since the two defaults have conflicting conclusions. Of course, since  $r_3$  is itself a default, the information it provides concerning the priority between  $r_1$  and  $r_2$  is defeasible, and could likewise be defeated.

The second step is to shift our attention from theories of the form  $\langle \mathcal{W}, \mathcal{D}, < \rangle$  – that is, from fixed priority default theories – to theories containing a set  $\mathcal{W}$  of ordinary propositions as well as a set  $\mathcal{D}$  of defaults, but no priority relation on the defaults that is fixed in advance. Instead, both  $\mathcal{W}$  and  $\mathcal{D}$  may contain initial information concerning priority relations among defaults, and then conclusions about these priorities, like any other conclusions, are arrived at through default reasoning. Because conclusions about the priorities among defaults might themselves vary depending on other conclusions drawn by the reasoning agent, theories like this, of the form  $\Delta = \langle \mathcal{W}, \mathcal{D} \rangle$ , are known as *variable priority* default theories; it is stipulated as part of the definition that the set  $\mathcal{W}$  of ordinary propositions must contain each instance of the irreflexivity and transitivity schemata

$$\neg(n < n),$$

$$(n < n' \wedge n' < n'') \supset n < n'',$$

in which the variables are replaced with names of the defaults belonging to  $\mathcal{D}$ .

Now suppose the agent accepts some particular scenario  $\mathcal{S}$  based on a variable priority theory; the third step, then, is to lift the priority ordering implicit in the agent's scenario to an explicit ordering that can be used in default reasoning. This is done in the simplest possible way, through the introduction of a *derived* priority ordering  $<_{\mathcal{S}}$ , defined as follows:

$$r <_{\mathcal{S}} r' \text{ just in case } \mathcal{W} \cup \text{Conclusion}(\mathcal{S}) \vdash n < n'.$$

The statement  $r <_{\mathcal{S}} r'$  is taken to mean that  $r'$  has a higher priority than  $r$  according to the scenario  $\mathcal{S}$ . The force of the definition, then, is that this relation holds just in case  $n < n'$  can be derived from the conclusions of the defaults belonging to  $\mathcal{S}$ , taken together with the hard information from  $\mathcal{W}$ . Because  $\mathcal{W}$  contains all instances of transitivity and irreflexivity, the derived priority relation  $<_{\mathcal{S}}$  is guaranteed to be a strict partial ordering.

The fourth and final step is to define the notion of a proper scenario for such a variable priority default theory. This is accomplished by leveraging our previous definition of proper scenarios for fixed priority theories of the form  $\langle \mathcal{W}, \mathcal{D}, < \rangle$ , where  $<$  can be any strict partial ordering whatsoever. Using this previous definition, we can now stipulate that  $\mathcal{S}$  is a proper scenario for the variable priority theory  $\Delta = \langle \mathcal{W}, \mathcal{D} \rangle$  just in case  $\mathcal{S}$  is a

proper scenario, in the previous sense, for the particular fixed priority theory  $\langle \mathcal{W}, \mathcal{D}, <_S \rangle$ , where  $\mathcal{W}$  and  $\mathcal{D}$  are carried over from the variable priority theory  $\Delta$ , and where  $<_S$  is the priority relation derived from the scenario  $S$  itself. The intuitive picture is this. In searching for a proper scenario, the agent arrives at some scenario  $S$ , which then entails conclusions about various aspects of the world, including priority relations among the agent's own defaults. If these derived priority relations can be used to justify the agent in accepting exactly the scenario  $S$  that the agent began with, then that scenario is proper.

These various definitions can be illustrated through a variant of the previous Nixon example in which it is useful to adopt, not the epistemic perspective of a third party trying to decide whether or not Nixon is a pacifist, but instead, the practical perspective of a young Nixon trying to decide whether or not to become a pacifist. As before, we take  $r_1$  and  $r_2$  as the defaults  $Q \rightarrow P$  and  $R \rightarrow \neg P$ , where  $P$ ,  $Q$ , and  $R$  are the propositions that Nixon is a pacifist, a Quaker, and a Republican. Given our current perspective, these two defaults should now be interpreted as providing practical, rather than epistemic, reasons:  $r_1$  corresponds to the fact that, as a Quaker, Nixon has reason to become a pacifist, and  $r_2$  to the fact that, as a Republican, he has reason not to become a pacifist.

In light of his conflicting reasons, let us imagine that Nixon seeks advice, first, from an elder of his Friends Meeting, who tells him that his religious reason should be given more weight than his political reason, but second, from an official of the Republican Party, who tells him exactly the opposite. If we take  $A$  and  $B$  as the respective statements of the religious and political figures, then the advice of these two authorities can be represented through the defaults  $r_3$  and  $r_4$ , where  $r_3$  is  $A \rightarrow n_2 < n_1$  and  $r_4$  is  $B \rightarrow n_1 < n_2$ . Nixon is now faced with his initial conflicting reasons, as well as further conflicting reasons as to how that initial conflict should be resolved. Finally, though, let us suppose that he seeks further counsel, perhaps from his wife, Pat, who tells him that the advice of the religious figure is to be preferred to that of the party official. If we take  $C$  as Pat's statement, her advice can be represented through the default  $r_5$ , where  $r_5$  is  $C \rightarrow n_4 < n_3$ .

The variable priority default theory that provides the background for Nixon's reasoning is  $\Delta_8 = \langle \mathcal{W}, \mathcal{D} \rangle$ , where  $\mathcal{W}$  contains the propositions  $A$ ,  $B$ ,  $C$ ,  $Q$ , and  $R$  – according to which Nixon is a Quaker and a Republican, and his religious and political advisors, as well as his wife, said what they did – and where  $\mathcal{D}$  now contains  $r_1$ ,  $r_2$ ,  $r_3$ ,  $r_4$ , and  $r_5$ .<sup>39</sup> As the reader can verify, this theory allows the unique proper scenario  $S_1 = \{r_1, r_3, r_5\}$ , leading to

$$E_1 = Th(\{A, B, C, Q, R, P, n_4 < n_3, n_2 < n_1\})$$

as its extension, the logical closure of the set containing the hard information from  $\mathcal{W}$  together with conclusions of the defaults from  $S_1$ . Appealing

to our Definition 5 treatment of deontic logic in terms of default logic, we can now see that  $\Delta_8$  supports the ought statements  $O(n_4 \prec n_3)$ ,  $O(n_2 \prec n_1)$ , and  $O.P$ . What the theory tells us, in other words, is that Nixon ought to take the advice of the religious figure more seriously than that of the political official, so that he ought to give more weight to his religious reason than to his political reasons, and therefore, that he ought to become a pacifist.

This form of reasoning – in which reasons are appealed to in the process of assessing the importance, or even the applicability, of other reasons – underlies many of our most significant normative judgments. But, while I am open to the possibility that the resultant sense of the conditional ought, reflecting interactions among what might be called first-order reasons, might be analyzed within the classical framework, I see no way at all for classical semantics to model the kind of higher-order reasoning illustrated here.<sup>40</sup> What would such a model even look like? The ordering source would have to contain, not just prioritized norms, but norms favoring priority relations among other norms, still further norms favoring priorities among those norms, possibly further norms about those – continuing up in a tangled hierarchy, but all cashed out, somehow, in an ordering on possible worlds, which could then be used in evaluating ought statements. It is hard to imagine that any such theory would not be hopelessly complex, or entirely artificial, or both.

## 8. Conclusion

I have compared two approaches to the semantics of ought statements. The first is the classical approach, according to which the evaluation of oughts depends on an ordering among worlds – and particularly the development of this approach in the hands of Kratzer, for whom the ordering on worlds is itself derived from an underlying set of norms: the ordering source. The second is an approach developed within the framework of default logic, according to which the underlying norms, or reasons, are interpreted as defaults, and the logic then shows how these reasons interact to determine which oughts are supported. The advantages of the classical semantics are evident. My goal in this article has been to explore the apparent advantages of the approach based on default logic, to see which of these could be accommodated within the classical framework.

I considered four apparent advantages. The first is that default logic allows sensible ought statements to be derived from conflicting norms. Although it is this feature of default logic that originally suggested its interpretation as a deontic logic, it turns out that this initial advantage is only apparent. As long as we limit ourselves to categorical norms and oughts, exactly the same pattern of input norms yielding output oughts

can be defined within the classical framework, without changing any structural features of the underlying models, but simply appealing to a different evaluation rule for ought statements. The second advantage is that default logic allows for a natural treatment of oughts derived from a prioritized set of norms, or reasons. Once again, though, this advantage is only apparent, since the norms from an ordering source can be prioritized as well, and then worlds ordered in a way that reflects this prioritization on norms.

Next, I turned to conditional oughts, and argued that these statements could be understood in two senses: first, a constrained maximization sense, which is captured within classical semantics, and second, a resultant sense, which is captured by default logic. Although I have seen no successful attempt at analyzing the resultant sense of conditional oughts within the classical framework, I also have no argument against the possibility: perhaps someone could enrich the ordering source from Kratzer models with something like conditional norms, and then order worlds in the basis of these conditional norms. Whether or not this feature of the default approach – its treatment of resultant oughts – can be accommodated within the classical framework therefore remains an open question.

Finally, I illustrated the way in which default logic allows oughts to depend, not only on interactions among first-order norms, or reasons, but also on higher-order reasons about other reasons – both about the relative importance of various reasons, and about which reasons should be excluded from consideration. While it may be possible for the resultant ought to be described within the classical framework, I see no plausible way for classical techniques to model this kind of higher-order reasoning – so here we have, I believe, an advantage of default logic that cannot be accommodated within classical semantics. Still, it is possible to wonder whether an expressive limitation of this kind can actually be taken as a criticism of the classical approach as theory of linguistic semantics. In showing how oughts can depend on reasoning about reasons – which reasons are more important than others, which should be excluded – the framework of default logic provides, in effect, a model of the entire process through which an ought statement might be supported by some complex pattern of interactions among reasons. And surely we do not need to understand all of normative reasoning in order to understand, simply, the semantics of ought statements. On the other hand, if the truth or falsity of oughts depends on an interplay among norms, or reasons, then a semantic theory of ought statements must take at least some aspects of this interplay into account, and it is hard to know exactly where to draw the line.<sup>41</sup>

Philosophy Department  
University of Maryland



## APPENDIX: VERIFICATIONS OF FACTS

Fact 1 was verified in Goble (2013). Facts 2 and 3 are obvious. Fact 4 is also obvious, but was verified in Horty (1993). Fact 5 was first established in Horty (1994). The key observation behind Fact 6 – that the Kratzer conflict evaluation rule from Definition 6 coincides with van Fraassen's rule, from Definition 4 – was first noted in Horty (1993), though in the context of model preference logic, rather than modal logic. Fact 7 was floating in the air, but follows from several observations of Goble (2013).

## NOTES

<sup>1</sup> These papers have recently been collected in Kratzer (2012).

<sup>2</sup> The label is not new; see von Fintel (2012). Gillies (2014) describes this dominant approach as the 'canon,' and points out that, with von Fintel and Heim, 2011, it has now achieved textbook status.

<sup>3</sup> A development of deontic logic within the framework of decision theory under uncertainty can be found in Horty (2001). This account was recently extended to richer decision theoretic frameworks by Bartha (2014), and to game theory by Kooi and Tamminga (2008) and Tamminga (2013).

<sup>4</sup> See Reiter (1980).

<sup>5</sup> Summaries of this work, and extensive references, can be found in Gabbay *et al.* (2014), and in future volumes of this handbook.

<sup>6</sup> See, for example, Chellas (1980), for axiomatizations of normal modal logics, including standard deontic logic. A brief, classic history of deontic logic is presented in Føllesdal and Hilpinen (1971); see Hilpinen and McNamara (2014), for a more recent survey, of significantly greater scope.

<sup>7</sup> In addition to Kratzer's own papers, helpful surveys of the approach can be found in von Fintel and Heim (2011), Hacquard (2011), and Portner (2009).

<sup>8</sup> The concept of stoppering was first introduced, I believe, by Makinson (1989) in the context of work on model preference logics, and has since been studied extensively; see especially Makinson (1994). It is, essentially, the appropriate form of the limit assumption for preorders that are not total, ruling out the possibility of unbounded chains of increasingly better worlds. The importance of this concept in the present setting was urged by Lou Goble, in personal correspondence, and especially in Goble (2013), which explores the relation between Kratzer semantics and a variety of deontic logics in detail.

<sup>9</sup> The label *KD* in the satisfaction relation  $\models_{KD}$  is meant to indicate that this satisfaction relation incorporates the 'Kratzer disjunctive' evaluation rule, a term that will make more sense once this rule is contrasted with the 'Kratzer conflict' evaluation rule, in Section 5 of this article. It is a happy notational accident that, as Fact 3 shows, this basic Kratzer logic coincides with standard deontic logic, which is itself often referred to as *KD*, since it results from supplementing the basic modal logic *K* with the *D* axiom; see Chellas (1980), for details.

<sup>10</sup> A useful collection containing many classic papers on the issue is Gowans (1987).

<sup>11</sup> See Chellas (1974), and Sections 6.5 and 10.2 of Chellas (1980). Chellas recommends also the further condition that  $\emptyset \notin N(\alpha)$ , according to which no individual norm can be inconsistent.

<sup>12</sup> If we were to accept Chellas's additional constraint that  $\emptyset \notin N(\alpha)$ , we would also require the axiom schema  $\neg \bigcirc(A \wedge \neg A)$ .



<sup>13</sup> See Williams (1965).

<sup>14</sup> See van Fraassen (1973).

<sup>15</sup> After about twenty years, beginning in the early 1990s, both van Fraassen's proposal and the logical problems presented by conflicting norms more generally began to attract serious interest among logicians. This more recent work is surveyed in Goble (2014), which also presents new ideas of its own, and which is the paper with which anyone interested in the topic should now begin.

<sup>16</sup> The interpretation within default logic first appears in Horty (1994), the interpretation within a model preference logic in Horty (1993).

<sup>17</sup> I argue in Horty (2012), that reasons are provided, not by defaults in general, but only by defaults that are *triggered* – a concept that will be defined shortly. This refinement is not necessary for the current discussion, and I will ignore it.

<sup>18</sup> See Chapter 8 of Horty (2012), for a discussion of the difficulties involved in arriving at a general definition.

<sup>19</sup> See Appendix A.1 of Horty (2012), for the full definition.

<sup>20</sup> These two broad strategies, along with some variants, are discussed in Chapters 1 and 7 Horty (2012).

<sup>21</sup> The notion of presumption is discussed in detail by Ullman-Margalit (1973), who argues that specific presumptions are justified by a mixture of probabilistic and 'value-related' considerations, and cites the presumption of innocence as one in whose justification the value-related considerations seem to outweigh those of probability.

<sup>22</sup> The canonical source for the view that oughts are grounded in reasons is Ross (1930), but the idea has been advanced, in one form or another, by a number of more modern writers, including Baier (1958), Chisholm (1964), Dancy (2004), Harman (1975), Nagel (1970), Parfit (2011), Raz (1975), Scanlon (1998), and Schroeder (2007).

<sup>23</sup> This evaluation rule can be found in Horty (1993), though in the context of model preference logic, rather than modal logic. The rule is discussed also by Goble (2013) who points out that it results simply by removing all reference to the initial world  $\beta$  from Kratzer's original evaluation rule, from Definition 2. Both von Fintel (2012) and Gillies (2014) likewise argue that conflicts can be accommodated within classical semantics by modifying Kratzer's account in minor ways, though they suggest different modifications from that proposed here.

<sup>24</sup> Note that this transform relation, unlike the others defined so far, need not be a function, since any number of Kratzer models, differing in their underlying sets of worlds, might count as Kratzer transforms of the same set of imperatives.

<sup>25</sup> Some of the problems with the Section 4 approach, and my own tentative proposal for correcting these problems, are discussed in Chapter 8 of Horty (2012). Other proposals can be found in, for example, Baader and Hollunder (1995), Brewka (1994a,b), Brewka and Eiter (2000), Delgrande and Schaub (2000), Rintanen (1998), and Hansen (2006, 2008).

<sup>26</sup> See, for example, von Fintel and Iatridou (2008), Katz *et al.* (2012), Rubinstein (2013), and Silk (2012).

<sup>27</sup> Indeed, Kratzer seems to have had a theory of just this sort in mind in the final few paragraphs of Kratzer (1981); these paragraphs, however, were removed from the version of that paper later reprinted in Kratzer (2012).

<sup>28</sup> To illustrate, we can see that the lexical treatment of classical priorities sketched here does not coincide with the treatment in the simple default logic presented earlier. Consider the theory  $\Delta_5 = \langle \mathcal{W}, \mathcal{D}, < \rangle$  where  $\mathcal{W} = \emptyset$ , where  $\mathcal{D} = \{r_1, r_2, r_3\}$  with  $r_1$  as  $\top \rightarrow B$ , with  $r_2$  as  $\top \rightarrow \neg(A \vee B)$ , and with  $r_3$  as  $\top \rightarrow A$ , and where  $r_1 < r_2 < r_3$ . The reader can verify that  $S_1 = \{r_3\}$  is the unique proper scenario based on this theory, leading to  $\mathcal{E}_1 = Th(\{A\})$  as its unique extension, so that the theory supports  $\bigcirc A$  but not  $\bigcirc B$ . However, the natural transform of  $\Delta_5$  into a prioritized Kratzer model would involve an ordering source in which the

propositions corresponding to each of the three defaults are ordered so that  $|B| < |\neg(A \vee B)| < |A|$ ; and, as the reader can again verify, the lexical treatment of classical priorities would support both  $\bigcirc A$  and  $\bigcirc B$ . The issue highlighted by this example is that of *reinstatement* – whether, for instance, the fact that  $r_3$  defeats  $r_2$  prevents  $r_2$  from defeating  $r_1$ , and so reinstates  $r_1$ . This issue is discussed at length in Section 8.2.2 of Horty (2012); other issues presented by priorities even among defaults with trivial premises, or among norms, are discussed in Goble, 2014, and the work cited there.

<sup>29</sup> See Martin (1982, p. 143).

<sup>30</sup> The disjunctive and conflict accounts coincide when applied to the theories  $\Delta_6$  and  $\Delta_6[A]$ , since each has only one extension.

<sup>31</sup> The argument about antecedent strengthening in this paragraph is drawn from Horty, 1994, but I am grateful to Goble for pointing out the parallel between this argument and the earlier argument about agglomeration.

<sup>32</sup> Conditional oughts of this sort are often referred to as *contrary to duty obligations*, but this label is misleading since they can be studied in situation in which there are no duties, no obligations, and indeed no agency at all – merely optimization under constraints expressed by an antecedent statement.

<sup>33</sup> See Chellas (1974), Hansson (1971), Lewis (1974), Prakken and Sergot (1996, 1997), and van Fraassen (1972).

<sup>34</sup> Although, as we have seen, the unconditional fragment of Kratzer's logic coincides with standard deontic logic, it is currently an open question whether Kratzer's analysis of the conditional deontic operator, presented in the first clause of Definition 9, coincides with any of the familiar conditional deontic logics.

<sup>35</sup> See, of course, Ross (1930). I have heard this sense of the conditional ought referred to as the *prima facie* sense, but this phrase is misleading because, on Ross's picture, it is the underlying norms that are supposed to be *prima facie*, rather than the resulting all-things-considered oughts.

<sup>36</sup> See, for example, Alchourrón (1993), and Loewer and Belzer (1983). There is also the converse question of whether constrained optimality oughts can be represented in a framework, such as that of default logic, designed for the treatment of resultant oughts. It is possible, in retrospect, to see a proposal along these lines in the work of McCarty (1994), and Ryu and Lee (1997), for example, but I am convinced by Prakken and Sergot's (1996) arguments that this proposal is mistaken.

<sup>37</sup> I am grateful to Frank Veltman for highlighting this concern.

<sup>38</sup> Both phenomena are explored at length in Chapter 5 of Horty (2012).

<sup>39</sup> The set  $\mathcal{W}$  must also contain appropriate instances of the asymmetry schema, but since these can be generated automatically from the defaults contained in that theory, I will not mention them explicitly.

<sup>40</sup> Describing reasons that do not bear on other reasons as 'first-order,' and reasons about reasons as 'higher-order,' is meant to be helpful and suggestive, but is not strictly accurate, since the representation of these reasons in default logic need not be stratified; see Horty (2012, p. 130), for discussion.

<sup>41</sup> I am grateful to Stephen Finlay and Mark Schroeder for inviting me to give the presentation that led to this article, to Fabrizio Cariani for suggesting a topic for the presentation, and to Fabrizio and Valentine Hacquard for helping me find my way through the literature on deontic modality. In preparing my presentation, I engaged in an extensive correspondence with Lou Goble, the effects of which can be seen on every page of this article, and indeed beyond, since it provoked him into producing Goble (2013). I am grateful also to Thony Gillies, whose comments on my book at an APA symposium, recorded in Gillies (2014), showed that he knew all along much of what I had just figured out about the relation

between default logic and classical semantics. After completing a draft of this paper, I was able to discuss it with Frank Veltman, who provided a helpful sanity check, but also brought several problems into sharper focus. And I received further written corrections and suggestions from Goble, and also from Shyam Nair, so substantial that, if I had tried to do justice to them all, this article would have been twice as long, and twice as late.

## REFERENCES

- Alchourrón, C. (1993). 'Philosophical Foundations of Deontic Logic and the Logic of Defeasible Conditionals,' in J-J. Meyer and R. M. Wieringa (eds) *Deontic Logic in Computer Science: Normative System Specifications*. Oxford: John Wiley and Sons, pp. 43–84.
- Baader, F. and Hollunder, B. (1995). 'Priorities on Defaults with Prerequisites, and their Applications in Treating Specificity in Terminological Default Logic,' *Journal of Automated Reasoning*, 15, pp. 41–68.
- Baier, K. (1958). *The Moral Point of View: A Rational Basis of Ethics*. Ithaca, NY: Cornell University Press.
- Bartha, P. (2014). 'Decisions in Branching Time,' in T. Müller (ed.) *Nuel Belnap on Indeterminism and Free Action*. Dordrecht: Springer Verlag, pp. 29–56.
- Brewka, G. (1994a). 'Adding Priorities and Specificity to Default Logic,' in *Proceedings of the European Workshop on Logics in Artificial Intelligence (JELIA-94)*, Springer Verlag Lecture Notes in Artificial Intelligence. Dordrecht: Springer Verlag, pp. 247–260.
- Brewka, G. (1994b). 'Reasoning About Priorities in Default Logic,' in *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*. Cambridge, MA: AAAI/MIT Press, pp. 940–945.
- Brewka, G. and Eiter, T. (2000). 'Prioritizing Default Logic,' in S. Hölldobler (ed.) *Intellectics and Computational Logic: Papers in Honor of Wolfgang Bibel*. Dordrecht: Kluwer Academic Publishers.
- Chellas, B. (1974). 'Conditional Obligation,' in S. Stenlund (ed.) *Logical Theory and Semantic Analysis*. Dordrecht: D. Reidel, pp. 23–33.
- Chellas, B. (1980). *Modal Logic: An Introduction*. Cambridge: Cambridge University Press.
- Chisholm, R. (1964). 'The Ethics of Requirement,' *American Philosophical Quarterly* 1, pp. 147–153.
- Dancy, J. (2004). *Ethics Without Principles*. Oxford: Oxford University Press.
- Delgrande, J. and Schaub, T. (2000). 'Expressing Preferences in Default Logic,' *Artificial Intelligence* 123, pp. 41–87.
- Føllesdal, D. and Hilpinen, R. (1971). 'Deontic Logic: An Introduction,' in R. Hilpinen (ed.) *Deontic Logic: Introductory and Systematic Readings*. Dordrecht: D. Reidel, pp. 1–35.
- van Fraassen, B. (1972). 'The Logic of Conditional Obligation,' *The Journal of Philosophical Logic* 72, pp. 417–438.
- van Fraassen, B. (1973). 'Values and the Heart's Command,' *The Journal of Philosophy* 70, pp. 5–19.
- von Fintel, K. (2012). 'The Best We Can (Expect) to Get? Challenges to the Classic Semantics for Deontic Modals'. Unpublished manuscript. Available online at: <http://web.mit.edu/fintel/fintel-2012-apa-ought.pdf>
- von Fintel, K. and Heim, I. (2011). 'Intensional Semantics'. Unpublished manuscript. Available online at: <http://web.mit.edu/fintel/fintel-heim-intensional.pdf>
- von Fintel, K. and Iatridou, S. (2008). 'How to Say *Ought* in Foreign: The Composition of Weak Necessity Modals,' in J. Guéron and J. Lecarme (eds) *Time and Modality*. Dordrecht: Springer, pp. 115–141.

- Gabbay, D., Horty, J., Parent, X., van der Meyden, R. and van der Torre, L. (eds) (2014). *Handbook of Deontic Logic and Normative Systems*. London: College Publications.
- Gillies, A. (2014). 'Default Logic and *Hafta*'. Unpublished manuscript.
- Goble, L. (2013). 'Notes on Kratzer Semantics for Modality with Application to Simple Deontic Logic'. Unpublished manuscript.
- Goble, L. (2014). 'Prima Facie Norms, Normative Conflicts, and Dilemmas,' in D. Gabbay, J. Horty, X. Parent, R. van der Meyden and L. van der Torre (eds) *Handbook of Deontic Logic and Normative Systems*. London: College Publications, pp. 241–351.
- Gowans, C. (ed.) (1987). *Moral Dilemmas*. Oxford: Oxford University Press.
- Hacquard, V. (2011). 'Modality,' in C. Maienborn, K. von Heusinger and P. Portner (eds) *Semantics: An International Handbook of Natural Language Meaning*, vol. 1. Berlin: De Gruyter Mouton, pp. 1484–1515.
- Hansen, J. (2006). 'Deontic Logic for Prioritized Imperatives,' *Artificial Intelligence and Law*, 14, pp. 1–34.
- Hansen, J. (2008). 'Prioritized Conditional Imperatives: Problems and a New Proposal,' *Autonomous Agents and Multi-agent Systems* 109, pp. 11–35.
- Hansson, B. (1971). 'An Analysis of Some Deontic Logics,' in R. Hilpinen (ed.) *Deontic Logic: Introductory and Systematic Readings*. Dordrecht: D. Reidel, pp. 121–147.
- Harman, G. (1975). 'Reasons,' *Critica* 7, pp. 3–18. Reprinted in J. Raz (ed.) (1978). *Practical Reasoning*. Oxford: Oxford University Press; pagination refers to this version.
- Hilpinen, R. and McNamara, P. (2014). 'Deontic Logic: A Historical Survey and Introduction,' in Gabbay *et al.*, 2014, pp. 3–136.
- Horty, J. (1993). 'Deontic Logic as Founded on Nonmonotonic Logic,' *Annals of Mathematics and Artificial Intelligence* 9, pp. 69–91.
- Horty, J. (1994). 'Moral Dilemmas and Nonmonotonic Logic,' *Journal of Philosophical Logic* 23, pp. 35–65. A preprint appears in J.-J. Meyer and R. M. Wieringa (eds) *Proceedings of the First International Workshop on Deontic Logic in Computer Science*. Amsterdam: Free University of Amsterdam, 1991, pp. 212–231.
- Horty, J. (2001). *Agency and Deontic Logic*. Oxford: Oxford University Press.
- Horty, J. (2012). *Reasons as Defaults*. Oxford: Oxford University Press.
- Katz, G., Portner, P. and Rubinstein, A. (2012). 'Ordering Combination for Modal Comparison,' in A. Chermes (ed.) *Proceedings of the 22nd Semantics and Linguistic Theory Conference (SALT 22)*. Ithaca, NY: Cornell/CLC Publications, pp. 488–507.
- Kooi, B. and Tamminga, A. (2008). 'Moral Conflicts Between Groups of Agents,' *Journal of Philosophical Logic* 37, pp. 1–21.
- Kratzer, A. (1981). 'The Notional Category of Modality,' in H. Eikmeyer, J. Hans and H. Rieser (eds) *Words, Worlds, and Contexts*. Berlin: De Gruyter Mouton, pp. 38–74. Reprinted in Kratzer, 2012.
- Kratzer, A. (2012). *Modals and Conditionals: New and Revised Perspectives*. Oxford: Oxford University Press.
- Lewis, D. (1974). 'Semantic Analyses for Dyadic Deontic Logic,' in S. Stenlund (ed.) *Logical Theory and Semantic Analysis*. Dordrecht: D. Reidel, pp. 1–14.
- Loewer, B. and Belzer, M. (1983). 'Dyadic Deontic Detachment,' *Synthese* 54, pp. 295–318.
- Makinson, D. (1989). 'General Theory of Cumulative Inference,' in M. Reinfrank, J. de Kleer, M. Ginsberg and E. Sandewall (eds) *Proceedings of the Second International Workshop on Nonmonotonic Reasoning*, vol. 346 of Springer Verlag Lecture Notes in Artificial Intelligence. Dordrecht: Springer Verlag, pp. 1–18.
- Makinson, D. (1994). 'General Patterns in Nonmonotonic Reasoning,' in D. Gabbay, C. Hogger and J. Robinson (eds) *Handbook of Logic in Artificial Intelligence and Logic*

- Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning*. Oxford: Oxford University Press, pp. 35–110.
- Martin, J. (1982). *Miss Manners' Guide to Excruciatingly Correct Behavior*. New York: Atheneum Publishers.
- McCarty, L. T. (1994). 'Defeasible Deontic Reasoning,' *Fundamenta Informaticae* 21, pp. 125–148.
- Nagel, T. (1970). *The Possibility of Altruism*. Princeton, NJ: Princeton University Press.
- Parfit, D. (2011). *On What Matters*. Oxford: Oxford University Press.
- Portner, P. (2009). *Modality*. Oxford: Oxford University Press.
- Prakken, H. and Sergot, M. (1996). 'Contrary-to-duty Obligations,' *Studia Logica* 57, pp. 91–115.
- Prakken, H. and Sergot, M. (1997). 'Dyadic Deontic Logic and Contrary to Duty Obligations,' in D. Nute (ed.) *Defeasible Deontic Logic*. Dordrecht: Kluwer Academic Publishers, pp. 233–262.
- Raz, J. (1975). *Practical Reasoning and Norms*. London: Hutchinson and Company, 2nd edn. with new Postscript printed in 1990 by Princeton University Press, and reprinted by Oxford University Press in 2002; pagination refers to the Oxford edition.
- Reiter, R. (1980). 'A Logic for Default Reasoning,' *Artificial Intelligence* 13, pp. 81–132.
- Rintanen, J. (1998). 'Lexicographic Priorities in Default Logic,' *Artificial Intelligence*, 106, pp. 221–265.
- Ross, W. D. (1930). *The Right and the Good*. Oxford: Oxford University Press.
- Rubinstein, A. (2013). 'Figuring Out What We Ought to Do: The Challenge of Delineating Priorities,' *University of Pennsylvania Working Papers in Linguistics* 19(1), pp. 169–178.
- Ryu, Y. U. and Lee, R. M. (1997). 'Deontic Logic Viewed as Defeasible Reasoning,' in D. Nute (ed.) *Defeasible Deontic Logic*. Dordrecht: Kluwer Academic Publishers, pp. 123–137.
- Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Schroeder, M. (2007). *Slaves of the Passions*. Oxford: Oxford University Press.
- Silk, A. (2012). 'Modality, Weights, and Inconsistent Premise Sets,' in A. Chermes (ed.) *Proceedings of the 22nd Semantics and Linguistic Theory Conference (SALT 22)*. Ithaca, NY: Cornell/CLC Publications, pp. 43–64.
- Tamminga, A. (2013). 'Deontic Logic for Strategic Games,' *Erkenntnis* 78, pp. 183–200.
- Ullman-Margalit, E. (1973). 'On Presumption,' *The Journal of Philosophy* 70, pp. 143–163.
- Williams, B. (1965). 'Ethical Consistency,' *Proceedings of the Aristotelian Society* 39 (supp.), pp. 103–124. A revised version appears in *Problems of the Self: Philosophical Papers 1956–1972*. Cambridge: Cambridge University Press, 1973, pp. 166–186.