Nonmonotonic Foundations for Deontic Logic

John F. Horty

Philosophy Department and
Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742
(Email: horty@umiacs.umd.edu)

*** DRAFT ***
Printed on: July 29, 1998

Abstract

Ever since its inception in the work of G. H. von Wright, deontic logic has been developed primarily as a species of modal logic. I argue in this paper, however, that, at least for certain purposes, the techniques recently developed within the field of nonmonotonic logic may provide a better theoretical framework for the formal study of normative reasoning than the usual modal treatment. After reviewing some standard approaches to deontic logic, I focus on two areas in which nonmonotonic techniques promise improved understanding: reasoning in the presence of conflicting obligations, and reasoning with conditional obligations.

1 Introduction

Ever since its inception in the work of G. H. von Wright [35], deontic logic has been developed primarily as a species of modal logic. I argue in this paper, however, that, at least for certain purposes, the techniques recently developed within the field of nonmonotonic logic may provide a better theoretical framework for the formal study of normative reasoning than the usual modal treatment.

The two subjects of deontic and nonmonotonic logic have evolved within different disciplines. Deontic logic was developed, for the most part, by philosophers concerned with valid patterns of ethical and legal reasoning. The study of nonmonotonic logic, on the other hand, was initiated, much more recently, by researchers in artificial intelligence who felt that the logical theories then available could not be used to represent the kind of defeasible generalizations that constitute so much of our commonsense knowledge. Still, even though they have evolved within such different disciplines, it is not really surprising that there might be illuminating connections between these two subjects, for it is natural to view the rules underlying much of our ethical and legal reasoning as themselves carrying the kind of defeasible quality that motivated the development of nonmonotonic logic.

In the past few years, much of the research in nonmonotonic logic has concentrated on conflicts among defeasible generalizations, and also on situations in which one of two conflicting generalization might be taken to override another. Accordingly, I focus here on two particular areas of normative reasoning in which the techniques of nonmonotonic logic promise improved understanding: the first is the logic of conflicting oughts statements; the second is the logic governing those conditionals taken as expressing prima facie oughts.

The paper is organized as follows. Section 2 reviews three logical approaches to conflicting oughts: the accounts provided by standard modal logic and by a weaker modal logic, and then an account set out by Bas van Fraassen in [34]. I argue that only van Fraassen's account is intuitively adequate; but as it turns out, this approach cannot be interpreted in any natural way within the framework of modal logic. Section 3 introduces one of the most familiar formalisms for nonmonotonic reasoning, Raymond Reiter's default logic [27], and then reviews a result, first presented in my [16], that justifies an interpretation of van Fraassen's deontic logic into Reiter's default logic. Section 4 explores some technical points concerning van Fraassen's account, and then develops three variant approaches. I believe that these variants approaches, like van Fraassen's original account, likewise reflect stable and coherent strategies for reasoning with conflicting oughts; but again, they cannot be interpreted naturally within modal logic.

Section 5 reviews two ways in which conditional ought statements have been formalized within within the framework of modal logic, and then points out a shared difficulty that renders each of these treatments inadequate as a formalization of prima facie oughts. Section 6 begins by sketching a way in which this problem might be addressed within the framework of nonmonotonic logic. A number of issues concerning conditional reasoning in the nonmonotonic framework are still unsettled, however, and so the section concludes by surveying some of the ways in which these open questions bear upon the preliminary sketch.

2 Normative conflicts

2.1 Standard deontic logic

On the usual approach to deontic logic, obligation is interpreted as a kind of necessity, which can be modeled using possible worlds techniques. The most familiar theory of this kind, known as *standard deontic logic*, is based on models of the form $\mathcal{M} = \langle W, f, v \rangle$, with W a set of possible worlds, v a valuation mapping sentence letters into sets of worlds at which they are true, and f a function mapping each world into a nonempty set of worlds. Where α is an individual world, $f(\alpha)$ can be thought of as the set of worlds ideal from the standpoint of α, those in which all the oughts in force at α are satisfied; or if we follow the common practice of identifying propositions with sets of worlds, $f(\alpha)$ can then be viewed as a proposition expressing the standard of obligation at work in α.

Against the background of these standard deontic models, the evaluation rule for the connective $\bigcirc$, representing 'It ought to be the case that ...', is given as

$$\mathcal{M}, \alpha \models \bigcirc A \text{ if and only if } f(\alpha) \subseteq |A|,$$

with $|A|$ representing the set of worlds in which A is true. The idea is that $\bigcirc A$ should hold just in case A is a necessary condition for things turning out as they should—just in case A is entailed by the relevant standard of obligation.

Let us say that a situation gives rise to a *normative conflict* if it presents each of two conflicting propositions as obligatory—if, for example, it supports the truth of both $\bigcirc A$ and $\bigcirc B$ where A and B cannot hold jointly; or, as an extreme case, if it supports the truth of both $\bigcirc A$ and $\bigcirc \neg A$. We often seem to face conflicts like this in everyday life, and there are a number of vivid examples in philosophy and literature. Perhaps the best known of these is Sartre's description in [28] of a student during the Second World War who felt for reasons of patriotism and vengeance (his brother had been killed by the Germans) that he ought to leave home in order to join the Free French, but who felt also, for reasons of sympathy and personal devotion, that he ought to stay at home in order to care for his mother.

Sartre presents this student's situation in a compelling way that really does make it seem as if he had been confronted with conflicting, and perhaps irreconcilable, moral principles. However, if standard deontic logic is correct, Sartre is mistaken: the student did not face a moral conflict—no one ever does, because according to standard deontic logic, such a conflict is impossible. This is easy to see. In order for the two statements $\bigcirc A$ and $\bigcirc B$ to be supported at a world α, we need both $f(\alpha) \subseteq |A|$ and $f(\alpha) \subseteq |B|$, from which it follows, of course, that $f(\alpha) \subseteq |A| \cap |B|$. If the statements A and B were inconsistent, however, we would have $|A| \cap |B| = \emptyset$, from which it would then follow that $f(\alpha) = \emptyset$ as well; but in standard deontic models, the sole requirement on f is that it should map each world into a nonempty set. Apart from what is presupposed by the background framework of normal modal logic, then, the entire content of standard deontic logic seems to be simply that there are no normative conflicts; and in fact, validity in these standard models can be axiomatized by supplementing the basic modal logic K with the statement

$$\neg(\bigcirc A \wedge \bigcirc \neg A)$$

as an additional axiom schema. The resulting system is known as KD.

Now this feature of standard deontic logic—that it rules out normative conflicts—has received extensive discussion in the philosophical literature. There is currently no consensus among moral theorists on the question whether an ideal ethical theory could actually be structured in such a way that moral dilemmas might arise.[1] Still, it can seem like an objectionable feature of standard deontic logic that it rules out this possibility. Because the question is open, and the possibility of moral dilemmas is a matter for substantive ethical discussion, it seems to be inappropriate for a position on this issue to be built into the logic of the subject. And even if it does turn out, ultimately, that research in ethics is able to exclude the possibility of conflicts in a correct moral theory, it may be useful all the same to have a logic that allows for conflicting oughts.

One reason for this is that the task of actually applying a correct moral theory to each of the ethical decisions we face every day would be difficult and time-consuming; and it seems unlikely, for most of us, that such a theory could have any more bearing upon our day to day ethical reasoning than physics has upon our everyday reasoning about objects in the world. Most of our commonsense ethical thinking seems to be guided instead, not by the dictates of moral theory, but by simple rules of thumb—'Return what you borrow', 'Don't cause harm'—and it is not hard to generate conflicts among these.[2] Moreover, normative reasoning more generally is conditioned by a number of oughts, many of which are founded in a concern with matters other than morality—etiquette, aesthetics, fun—and of course, these lead to other conflicts both among themselves and with the oughts of morality. Even if we do eventually conclude, then, that there can be no clashes among the oughts generated by a correct ethical theory, it would still seem necessary to allow for conflicting oughts in any logic that aims to represent either our everyday moral thinking or our normative reasoning more broadly.

Finally, the need for a deontic logic that tolerates conflicting norms can be motivated from a different perspective if we imagine an intelligent system that is designed to reason about and achieve certain goals supplied to it by its users, and that represents those goals as ought statements. It is always possible for different users (or even for the same user) to supply the system with conflicting goals; and in such a case, we would not want the mechanisms for reasoning about goals to break down entirely, as it would if it were guided by standard deontic logic. This kind of situation is analogous to that envisioned by Nuel Belnap [2, 3] as a way of motivating the applicability of a contradiction tolerating logic (a relevance logic, as it happens) in the area of automated reasoning. Belnap imagines a computer designed to reason from data supplied by its users; and he argues that there are situations in which, even if the users inadvertently supply the machine with inconsistent information—say, A and $\neg A$—we would not want it to conclude that everything is true. In the same way, we can easily imagine a situation in which, even if a machine happens to be supplied by its users with inconsistent goals—say, $\bigcirc A$ and $\bigcirc \neg A$—we would not want it to conclude, as in standard

[1] The issue has been addressed, for example, by Alan Donagan [9], Philippa Foot [10], E. J. Lemmon [18], Ruth Barcan Marcus [22], and Bernard Williams [36].

[2] The relation between moral theory and the rules of thumb that guide everyday ethical decisions has recently been discussed by Dennett [8].

deontic logic, that it should regard every proposition as a goal.

2.2 A weak modal logic

One strategy for adapting deontic logic to reason sensibly in the face of conflicting norms is to continue the attempt to develop the subject within a modal framework, but simply to move to a weaker, non-normal modal logic. The clearest example of this is Brian Chellas's suggestion, in [5] and [6, Sections 6.5 and 10.2], that we base our deontic logic on a class of minimal models for modal logic, in which the accessibility relation maps individual worlds, not into sets of worlds, but into sets of propositions—sets of sets of worlds. More exactly, Chellas recommends a deontic logic based on models of the form $\mathcal{M} = \langle W, N, v \rangle$, with W and v as before, but with N a function from W into $\mathcal{P}(\mathcal{P}(W))$, subject to the condition that, for each of the propositions X and Y in $\mathcal{P}(W)$, if $X \in N(\alpha)$ and $X \subseteq Y$, then $Y \in N(\alpha)$.[3] Intuitively, the various propositions belonging to $N(\alpha)$ can be thought of as expressing the variety of different ways in which things ought to turn out at α, the variety of different normative standards in force at α.

In these models, the truth conditions for ought statements can be presented through the rule

$$\mathcal{M}, \alpha \models \bigcirc A \text{ if and only if } |A| \in N(\alpha);$$

the idea is that $\bigcirc A$ should hold just in case A is entailed by some normative standard at work in α. And validity is axiomatized by the system *EM*, which results from supplementing ordinary propositional logic with the rule schema

$$\frac{A \supset B}{\bigcirc A \supset \bigcirc B.}$$

In fact, this logic is weak enough to tolerate normative conflicts: the statements $\bigcirc A$ and $\bigcirc \neg A$ are jointly satisfiable, without entailing $\bigcirc B$. However, in weakening standard deontic logic to allow conflicts, it seems that we have now arrived at a system that is too weak: it fails to validate intuitively desirable inferences. Suppose, for example, that an agent is subject to the following two norms, the first issuing perhaps from some legal authority, the second from religion or conscience:

> You ought either to fight in the army or perform alternative service,
> You ought not to fight in the army.

We can represent these norms through the formulas $\bigcirc(F \vee S)$ and $\bigcirc \neg F$. Now it seems intuitively that the agent should conclude from these premises that he ought to perform alternative service. However, the inference from $\bigcirc(F \vee S)$ and $\bigcirc \neg F$ to $\bigcirc S$ is not valid in the logic *EM*.

[3]Chellas recommends also the further condition that $\emptyset \notin N(\alpha)$. I ignore this condition because it seems like an overly strong constraint for many application areas, particularly the case in which the oughts of a deontic logic represent goals supplied to an intelligent system by its users. We would not want to rule out the possibility that a fallible user might present an intelligent system with an impossible goal ("Find a rational root for this equation"), or to abandon sensible reasoning in such a case.

Let us look at this problem a bit more closely. Any logical consequence of an ought derivable in *EM* is itself derivable as an ought in this system; and of course, S is a logical consequence of $(F \vee S) \wedge \neg F$. Therefore, we would be able to derive $\bigcirc S$ from our premise set if we could somehow merge the individual oughts $\bigcirc(F \vee S)$ and $\bigcirc\neg F$ together into a joint ought of the form

$$\bigcirc((F \vee S) \wedge \neg F).$$

But how could we get this latter statement? It seems possible to derive it from our premises only through a rule of the form

$$\frac{\bigcirc A \qquad \bigcirc B}{\bigcirc(A \wedge B),}$$

dubbed by Bernard Williams [36] as the rule of *agglomeration*. However, such a rule is not admissible in *EM*, and in fact, it is exactly the kind of thing that this logic is designed to avoid: from $\bigcirc A$ and $\bigcirc\neg A$, agglomeration would allow us to conclude $\bigcirc(A \wedge \neg A)$, and so $\bigcirc B$ for arbitrary B, due to closure of ought under logical consequence.

Evidently, the issue of agglomeration is crucial for a proper logical understanding of normative conflicts. We do not want to allow unrestricted agglomeration, as in the standard deontic logic *KD*; this would force us to treat conflicting oughts as incoherent. On the other hand, we do not want to block agglomeration entirely, as in the weak deontic logic *EM*; we would then miss some desirable consequences in cases in which conflict is not a problem.

2.3 Van Fraassen's proposal

As far as I know, the first intuitively adequate account of reasoning in the presence of normative conflicts was presented in Bas van Fraassen's [34], a paper that is largely devoted to more broadly philosophical issues. Suppose that Γ is a set of oughts, possibly conflicting, and let us say that a statement of the form $\bigcirc B$ is fulfilled in some situation just in case B is true in that situation. The basic idea behind van Fraassen's suggestion, then, is that the statement $\bigcirc A$ should follow from Γ just in case the truth of A is a necessary condition for fulfilling some maximal set of the ought statements contained in Γ.

As van Fraassen presents it, the account relies formally on a notion of *score*. Where $\mathcal{M}$ is an (ordinary, classical) model of an underlying, ought-free language, the score of $\mathcal{M}$, relative to a set of ought statements Γ, is defined as the set of statements from Γ that $\mathcal{M}$ fulfills: $score_\Gamma(\mathcal{M}) = \{\bigcirc B \in \Gamma : \mathcal{M} \models B\}$. In this non-modal framework, we now let $|A|$ represent the ordinary model class of A: $|A| = \{\mathcal{M} : \mathcal{M} \models A\}$. Van Fraassen's notion of deontic consequence, which we represent as the relation $\vdash_F$, is then defined as follows.

Definition 1 $\Gamma \vdash_F \bigcirc A$ *if and only if there is a model* $\mathcal{M}_1 \in |A|$ *for which there is no model* $\mathcal{M}_2 \in |\neg A|$ *such that* $score_\Gamma(\mathcal{M}_1) \subseteq score_\Gamma(\mathcal{M}_2)$.

As in the logic *EM*, this notion of consequence is weak enough that conflicting oughts do not imply arbitrary oughts: we cannot derive $\bigcirc B$ from $\bigcirc A$ and $\bigcirc\neg A$. However, unlike *EM*, this way of characterizing deontic consequence does allow what seems to be the right degree of agglomeration: we can agglomerate individual oughts as long as this does not lead to the introduction of an inconsistent formula within the scope of an ought. For example, although we do not get

$$\bigcirc A, \bigcirc \neg A \vdash_F \bigcirc (A \wedge \neg A),$$

we do have

$$\bigcirc (F \vee S), \bigcirc \neg F \vdash_F \bigcirc ((F \vee S) \wedge \neg F);$$

and then, since any logical consequence of an ought is itself an ought, this tells us that

$$\bigcirc (F \vee S), \bigcirc \neg F \vdash_F \bigcirc S.$$

Although this proposal of van Fraassen's does appear to capture an intuitively attractive and stable account of reasoning in the presence of conflicting norms, and although the general topic of normative conflict has been an issue of intense concern in philosophy for well over a decade, it is hard to find any discussion of the proposal in either the philosophical or the logical literature on the topic. I feel that part of the reason for this neglect is that both philosophers and logicians are accustomed to approaching deontic logic from the perspective of modal logic; and as we will see, van Fraassen's proposal does not fit naturally within this framework. It turns out, however, that the proposal can be interpreted in a straightforward way within the theoretical framework provided by nonmonotonic logic.

3 An interpretation within default logic

3.1 Default logic

Although nonmonotonic logics have found applications in areas as diverse as database theory and automated diagnosis, an important initial motive in their development was the need felt within artificial intelligence for a formalism more naturally suited than ordinary logical systems to model the tentative nature of commonsense reasoning. Often, it seems, we want to draw conclusions from a given body of data that we would be willing to abandon if that data were supplemented with further information. To take a standard example, if we were told that Tweety is a bird, most of us would conclude that Tweety can fly—since we believe that, as a general rule, birds can fly. However, we would abandon this conclusion, and we would not feel that we had been presented with any kind of inconsistency, if we were then told in addition that Tweety cannot fly.

By now, a number of different formalisms have evolved with the field of nonmonotonic reasoning, but we focus here on one of the most familiar: Raymond Reiter's default logic [27], which supplements a standard classical logic with new rules of inference, known as *default rules*. In order to characterize the conclusion sets of theories involving these new default rules, Reiter then modifies the standard, monotonic notion of logical consequence.

An ordinary rule of inference (with a single premise) can be depicted simply as a premise-conclusion pair, such as (A/B). This rule commits a reasoner to B once A has been established. By contrast, a default rule is a triple, such as $(A : C \ / \ B)$. Very roughly, this rule can be thought of as committing the reasoner to B once A has been established and, in addition, C is consistent with the reasoner's conclusion set. The formula A is referred to as the *prerequisite* of this default rule, B as its *consequent*, and C as its *justification*. A *default theory* is a pair $\Delta = \langle \mathcal{W}, \mathcal{D} \rangle$, in which is $\mathcal{W}$ is a set of ordinary formulas and $\mathcal{D}$ is a set of default rules.

Before going on to set out the new concept of a conclusion set defined by Reiter for default theories, let us see how the information given above about Tweety might be represented in default logic. The first case, in which we are told only that Tweety is a bird, can be represented by the default theory $\Delta_1 = \langle \mathcal{W}_1, \mathcal{D}_1 \rangle$, where $\mathcal{W}_1 = \{Bt\}$ and $\mathcal{D}_1 = \{(Bt : Ft \ / \ Ft)\}$. Here the default rule says that if we know Tweety is a bird, and it is consistent with what we know that Tweety can fly, then we should conclude that Tweety can fly. (The generic statement 'Birds fly' can be taken to mean that, once we learn of some object that it is a bird, we should conclude that it flies, unless we happen to know that it does not. The default rule can then be thought of as an instantiation for Tweety of this generic truth.) In this case, because we do know that Bt, and there is no reason to think that Ft is inconsistent with what we know, the default rule yields Ft as a conclusion. Where Cn is a function mapping any set of formulas to its logical closure, then, the appropriate conclusion set based on Δ_1 seems to be $Cn[\{Bt, Ft\}]$, the logical closure of what we are told to begin with together with the conclusions of the applicable defaults. In the second case, however, when we are told in addition that Tweety does not fly, we move to the default theory $\Delta_2 = \langle \mathcal{W}_2, \mathcal{D}_2 \rangle$, with $\mathcal{D}_2 = \mathcal{D}_1$ and $\mathcal{W}_2 = \mathcal{W}_1 \cup \{\neg Ft\}$. Here the default rule cannot be applied, because its justification is inconsistent with what we know. So the appropriate conclusion set based on Δ_2 seems to be $Cn[\mathcal{W}_2]$.

These two examples illustrate, in some simple and natural cases, the kind of conclusion sets desired from given default theories. The task of arriving at a general definition of this notion, however, is not trivial; the trick is to find a way of capturing the intended meaning of the new component—the justification—present in default rules. A default rule is supposed to be applicable only if its justification is consistent with the conclusion set; but what can consistency mean in this setting? Consistency is usually defined in terms of logical consequence (a set is consistent if there is no explicit contradiction among its consequences), and so there is a danger of circularity here. In fact, the very application of a default rule might undermine its own justification, or the justification of some other rule that has already been applied. As an example, consider the theory $\Delta_3 = \langle \mathcal{W}_3, \mathcal{D}_3 \rangle$, with $\mathcal{W}_3 = \{A, B \supset \neg C\}$ and $\mathcal{D}_3 = \{(A : C \ / \ B)\}$. Before any new conclusions are drawn from this information, the rule $(A : C \ / \ B)$ seems to be applicable, since its prerequisite already belongs to the initial data set $\mathcal{W}_3$, and its justification is consistent with this set. The effect of applying this rule, though, is to introduce B into the conclusion set; just a bit of additional reasoning then shows that the conclusion set must contain $\neg C$ as well, and so the applicability of the default rule is undermined.

Of course, a chain of reasoning like this showing that some default rule is undermined can be arbitrarily long; and so we cannot really be sure that a default rule is applicable in some context until we have applied it, along with all the other rules that seem applicable, and then surveyed the logical closure of the result. Because of this, the conclusion set associated with a default theory cannot be defined in the usual iterative way, by successively adding to the original data the conclusions of the applicable rules of inference, and then taking the limit of this process.

Instead, Reiter is forced to adopt a fixed point approach in specifying the conclusion sets of default theories. He first defines an operator Φ that uses the information from a particular default theory Δ to map each formula set $\mathcal{S}$ into the formula set $\Phi_\Delta(\mathcal{S})$, as follows.

Definition 2 *Where $\Delta = \langle \mathcal{W}, \mathcal{D} \rangle$ is a default theory and $\mathcal{S}$ is some set of formulas, $\Phi_\Delta(\mathcal{S})$ is the minimal set satisfying the following three conditions:*

1. $\mathcal{W} \subseteq \Phi_\Delta(\mathcal{S})$,
2. $Cn[\Phi_\Delta(\mathcal{S})] = \Phi_\Delta(\mathcal{S})$,
3. *For each* $(A : B \;/\; C) \in \mathcal{D}$, *if* $A \in \Phi_\Delta(\mathcal{S})$ *and* $\neg B \notin \mathcal{S}$, *then* $C \in \Phi_\Delta(\mathcal{S})$.

The first two conditions in this definition tell us simply that $\Phi_\Delta(\mathcal{S})$ contains the information provided by the original theory, and that it is closed under logical consequence; the third condition tells us that it contains the conclusions of the default rules applicable in $\mathcal{S}$; and the minimality constraint prevents unwarranted conclusions from creeping in.

Where $\Delta = \langle \mathcal{W}, \mathcal{D} \rangle$ is a default theory, the operator Φ_Δ maps any formula set $\mathcal{S}$ into the minimal superset of $\mathcal{W}$ that is closed under both ordinary logical consequence and the default rules from $\mathcal{D}$ that are applicable in $\mathcal{S}$. The appropriate conclusion sets of default theories—known as *extensions*—are then defined as the fixed points of this operator.

Definition 3 *The set $\mathcal{E}$ is an extension of the default theory Δ if and only if $\Phi_\Delta(\mathcal{E}) = \mathcal{E}$.*

As the reader can verify, the default theories Δ_1 and Δ_2 above have the advertised conclusion sets as their extensions. In addition, it should be clear that the notion of an extension defined here is a conservative generalization of the corresponding notion of a conclusion set from ordinary logic: the extension of a default theory $\langle \mathcal{W}, \mathcal{D} \rangle$ in which $\mathcal{D}$ is empty is simply $Cn[\mathcal{W}]$.

3.2 Multiple extensions

In contrast to the situation in ordinary logic, however, not every default theory leads to a single set of appropriate conclusions. Some default theories, such as Δ_3 above, can be shown to have no extensions; these theories are often viewed as incoherent. More interesting, for our purposes, some default theories lead to multiple extensions. A standard example arises when we try to encode within default logic the following set of facts:

Nixon is a Quaker,
Nixon is a republican,
Quakers tend to be pacifists,
Republicans tend not to be pacifists.

If we instantiate for Nixon the general statements expressed here about Quakers and republicans, the resulting theory is $\Delta_4 = \langle \mathcal{W}_4, \mathcal{D}_4 \rangle$, with $\mathcal{W}_4 = \{Qn, Rn\}$ and $\mathcal{D}_4 = \{(Qn : Pn \;/\; Pn), (Rn : \neg Pn \;/\; \neg Pn)\}$.

This theory allows both $Cn[\mathcal{W}_4 \cup \{Pn\}]$ and $Cn[\mathcal{W}_4 \cup \{\neg Pn\}]$ as extensions. Initially, before we draw any new conclusions, both of the default rules from $\mathcal{D}_4$ are applicable, but once we adopt the conclusion of either, the applicability of the other is blocked.

In cases like this, when a default theory leads to more than one extension, it is hard to decide what conclusions a reasoner should actually draw from the information contained in the theory. Two broad reasoning 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, now generally described as the *skeptical* strategy, the reasoner should endorse a conclusion only if it is contained in the intersection of the theory's extensions.[4] 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, or else to combine them somehow into a unique set. As we shall see, however, the multiple extensions provided by default logic are no longer embarrassing when it comes to interpreting deontic ideas; they give us exactly what we need.

3.3 Oughts as defaults

Often, and in all of our examples so far, 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, a view that is most thoroughly developed by Judea Pearl [24].

There are, however, some important examples of default reasoning that do not seem to fit so naturally into the probabilistic framework. In driving along a narrow country road, for instance, it is best, whenever one approaches the crest of a hill, to adopt the default that there will be traffic in the oncoming lane, even if the road is deserted and the actual likelihood of traffic is low. Again, the presumption of innocence in a legal system 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 he has committed no crime unless there is conclusive evidence to the contrary.[5]

[4]The use of the *credulous/skeptical* terminology to characterize these two broad reasoning strategies was first introduced in Touretzky et al. [30], but the distinction is older than this; it was noted already in Section 2.2 of Reiter's [27], and was described in McDermott [23] as the distinction between *brave* and *cautious* reasoning.

[5]The notion of presumption is discussed in detail by Edna Ullman-Margalit [32], 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.

Those who favor a probabilistic understanding of defaults can attempt to account for discrepancies like these between defaults and commonsense generalizations by supposing that default rules might reflect, in addition, information concerning utilities of the outcomes. (For example, it could be argued that the default concerning oncoming traffic is reasonable, even though the likelihood is low, because the cost of a false negative in this case is potentially so high.) But there is also another explanation of the differences here between defaults and commonsense probabilistic generalizations. What these examples suggest is that default rules can be used to represent *norms* quite generally. When the norms involved have a probabilistic basis, it is natural to expect default reasoning to resemble probabilistic reasoning. But default rules can be used also, it seems, 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 reading of defaults as representing norms in general that motivates the connection, first established in [16], between default and deontic logics: if the norms generated by ought statements are represented through default rules, it turns out that van Fraassen's theory of oughts can be interpreted in a straightforward way within Reiter's default logic.

Formally, the interpretation is developed as follows. Where Γ is some set of ought statements, we first define the corresponding default theory as $\Delta_\Gamma = \langle \mathcal{W}, \mathcal{D} \rangle$, with $\mathcal{W} = \emptyset$ and $\mathcal{D} = \{(\top : B \;/\; B) : \bigcirc B \in \Gamma\}$, and with $\top$ representing the universal truth. The interpretation of ought statements as defaults is then justified by the following result.

Theorem 1 $\Gamma \vdash_F \bigcirc A$ *if and only if* $A \in \mathcal{E}$ *for some extension* $\mathcal{E}$ *of* Δ_Γ.

This result suggests a way of understanding the extensions of a default theory as descriptions of situations in which the norms expressed by the defaults are fulfilled; a default theory that contains conflicting norms, which cannot all be fulfilled at once, will then give rise to multiple extensions.

It is interesting to note that this particular way of defining a deontic logic within default logic, aimed at imterpreting van Fraassen's theory, relies upon a credulous treatment of multiple extensions: a statement of the form $\bigcirc A$ is said to follow from a set of oughts Γ just in case A belongs to any extension of the corresponding default theory. A deontic consequence relation that relies upon a skeptical treatment of multiple extensions will be motivated and defined in Section 4.3.

4 Exploring the theory

4.1 The consequence relation

Although, as we have seen, van Fraassen's notion of deontic consequence fits naturally within the framework of nonmonotonic logic, the consequence relation $\vdash_F$ is itself monotonic: from the fact that $\Gamma \vdash_F \bigcirc A$ we can conclude that $\Gamma \cup \Gamma' \vdash_F \bigcirc A$. This result follows at once from our Theorem 1 together with Theorem 3.2 of Reiter [27], and also, more directly, from Theorem 2 below; what it suggests is that, in relating van Fraassen's account of oughts to default logic, we are not yet relying on the actual nonmonotonicity of this theory, but only on its ability to yield multiple, mutually inconsistent sets of sentences as consequence sets for a given set of premises. This will change in Section 4.3, where we consider some alternatives to the present treatment, and also in in Section 6, where we extend the present treatment to deal with conditional oughts.

It is easy to see both that the logical truths follow as oughts from any premise set, and also, as mentioned earlier, that any logical consequence of a generated ought is itself generated as an ought: $\vdash A$ implies $\Gamma \vdash_F \bigcirc A$; and $\Gamma \vdash_F \bigcirc A$ and $A \vdash B$ together imply $\Gamma \vdash_F \bigcirc B$. Moreover, van Fraassen's consequence relation allows us to derive only consistent formulas as oughts (a form of ought implies can), no matter what ought statements it is supplied with as premises: if $\Gamma \vdash_F \bigcirc A$, then A is consistent.

Because only consistent formulas are derivable as oughts, we can see at once that the consequence relation $\vdash_F$ is not reflexive. Although an inconsistent ought might appear among some set of premises, it cannot appear as a conclusion of those premises; and so we do not have

$$\bigcirc(A \wedge \neg A) \vdash_F \bigcirc(A \wedge \neg A),$$

for example. From this, it follows that van Fraassen's theory cannot be captured in a natural way within a conventional modal logic, since any such logic carries a reflexive consequence relation.

In addition, the $\vdash_F$ relation fails to satisfy the cut rule; for example, although we have

$$\bigcirc(A \wedge B) \vdash_F \bigcirc A$$

and

$$\bigcirc A, \bigcirc\neg B \vdash_F \bigcirc(A \wedge \neg B),$$

we do not have

$$\bigcirc(A \wedge B), \bigcirc\neg B \vdash_F \bigcirc(A \wedge \neg B).^6$$

4.2 Some comparisons

As might be expected, van Fraassen's consequence relation $\vdash_F$ generally lies between $\vdash_{EM}$ and $\vdash_{KD}$, the consequence relations associated with *EM* and *KD*; it generally allows us to derive more oughts from a given set of premises than *EM* and fewer than *KD*. But there are exceptions to this general rule, and we need to introduce some technical vocabulary in order to describe the situation exactly.

First, let us officially characterize an *ought statement* as a statement of the form $\bigcirc A$ in which A is $\bigcirc$-free. Since the modal theories allow for iterated deontic operators and van Fraassen's theory does not, we must restrict ourselves in comparisons to the shared sub-language of ought statements. Where Γ is a set of ought statements, we will let $\overline{\Gamma} = \{B : \bigcirc B \in \Gamma\}$ represent the content of the oughts contained in that set. We can then define a set of ought statements Γ as *unit consistent* if each individual ought belonging to the set is itself satisfiable—that is, if B is consistent for each $B \in \overline{\Gamma}$. And we will say that Γ is not just unit consistent but *consistent* if the oughts belonging to Γ are jointly satisfiable—that is, if $\overline{\Gamma}$ itself is consistent.

Before working out the exact relations among these different deontic logics, we offer yet another characterization, perhaps the most straightforward, of the consequence relation $\vdash_F$.

[6] I owe this observation, along with the example, is due to Johan van Benthem (personal correspondence).

Theorem 2 *Let Γ be a set of ought statements. Then $\Gamma \vdash_F \bigcirc A$ if and only if there a consistent subset $\mathcal{G}$ of $\overline{\Gamma}$ such that $\mathcal{G} \vdash A$.*

Proof First, suppose $\Gamma \vdash_F \bigcirc A$. Let $\mathcal{M}_1$ be as in Definition 1, and let $\mathcal{G} = Th(\mathcal{M}_1) \cap \overline{\Gamma}$. Clearly, $\mathcal{G}$ is consistent and a subset of $\overline{\Gamma}$; and it is clear also that $score_\Gamma(\mathcal{M}) = score_\Gamma(\mathcal{M}')$ for any $\mathcal{M}, \mathcal{M}' \in |\mathcal{G}|$. To see that $\mathcal{G} \vdash A$, suppose otherwise: then there exists a model $\mathcal{M}_2 \in |\mathcal{G}| \cap |\neg A|$; but in that case we have $score_\Gamma(\mathcal{M}_2) = score_\Gamma(\mathcal{M}_1)$, contrary to the definition of $\vdash_F$. Next, suppose $\mathcal{G} \vdash A$ for some consistent subset $\mathcal{G}$ of $\overline{\Gamma}$. Standard techniques allow us to define a maximal consistent subset $\mathcal{G}^*$ of $\overline{\Gamma}$ containing $\mathcal{G}$. Since $\mathcal{G}^*$ is consistent, and since it must also entail A, we have some model $\mathcal{M}_1 \in |\mathcal{G}^*| \subseteq |A|$; and then since $\mathcal{G}^*$ is maximal, it is easy to see that there can be no $\mathcal{M}_2 \in |\neg A|$ such that $score_\Gamma(\mathcal{M}_1) \subseteq score_\Gamma(\mathcal{M}_2)$. So $\Gamma \vdash_F \bigcirc A$. ∎

We consider first the relations between van Fraassen's theory and *EM*. If a set of ought statements Γ is not even unit consistent, we must have $\Gamma \vdash_{EM} \bigcirc A$ for every A; and so *EM* is stronger than van Fraassen's theory, since this theory allows us to derive only consistent oughts. As we have seen from the army example discussed above in Sections 2.2 and 2.3, however, van Fraassen's theory does allow us to draw conclusions from certain unit consistent sets that cannot be derived in *EM*; and together with the following theorem, this shows that the theory is properly stronger than *EM* for unit consistent sets of oughts.

Theorem 3 *Let Γ be a unit consistent set of ought statements. Then if $\Gamma \vdash_{EM} \bigcirc A$, it follows that $\Gamma \vdash_F \bigcirc A$.*

Proof We begin by constructing a model for the modal language in which the possible worlds are ordinary models of the underlying classical language. Let $\mathcal{M} = \langle W, N, v \rangle$, where W is the set of models of the underlying classical language, and in which $N(\alpha) = \{X : |B| \subseteq X \text{ and } \bigcirc B \in \Gamma\}$ for each $\alpha \in W$, and $v(p) = |p|$ for each proposition letter p. It is clear that $\mathcal{M}$ is a minimal model satisfying the condition that, if $X \in N(\alpha)$ and $X \subseteq Y$, then $Y \in N(\alpha)$; and clear also that $\mathcal{M}, \alpha \models \Gamma$ for each $\alpha \in W$. Therefore, since $\Gamma \vdash_{EM} \bigcirc A$, we know that $\mathcal{M} \models \bigcirc A$; that is, $|A| \in N(\alpha)$ for each $\alpha \in W$. From this and the definition of N, we can conclude that $|B| \subseteq |A|$ for some $\bigcirc B \in \Gamma$. However, since Γ is unit consistent, $\{B\}$ is then a consistent subset of $\overline{\Gamma}$ that entails A; and so we can conclude that $\Gamma \vdash_F \bigcirc A$ from Theorem 2. ∎

We turn now to *KD*. Of course, anything can be derived in *KD* from an inconsistent set of oughts; and so, together with the following theorem, this shows that, as expected, *KD* is properly stronger than van Fraassen's theory.

Theorem 4 *Let Γ be a set of ought statements. Then if $\Gamma \vdash_F \bigcirc A$, it follows that $\Gamma \vdash_{KD} \bigcirc A$.*

Proof Suppose $\Gamma \vdash_F \bigcirc A$. By Theorem 2, it follows that $\mathcal{G} \vdash A$ for some subset $\mathcal{G}$ of $\overline{\Gamma}$; and so $\vdash (B_1 \wedge \ldots \wedge B_n) \supset A$, for some $B_1, \ldots, B_n \in \overline{\Gamma}$. Since *KD* is a normal modal logic, we can conclude from this that $\vdash_{KD} (\bigcirc B_1 \wedge \ldots \wedge \bigcirc B_n) \supset \bigcirc A$; and so $\Gamma \vdash_{KD} \bigcirc A$, since $\bigcirc B_1, \ldots, \bigcirc B_n \in \Gamma$. ∎

It is reassuring to see, however, that, unlike *EM*, van Fraassen's theory differs from *KD* only when applied to an inconsistent set of ought statements; otherwise, the two theories yield exactly the same results.

Theorem 5 *Let Γ be a consistent set of ought statements. Then if $\Gamma \vdash_{KD} \bigcirc A$, it follows that $\Gamma \vdash_F \bigcirc A$.*

Proof As in the proof of Theorem 3, we construct a model for the modal language with the ordinary models of the underlying classical language as its possible worlds. Let $\mathcal{M} = \langle W, f, v \rangle$, with W and v as before, but in which $f(\alpha) = |\overline{\Gamma}|$ for each $\alpha \in W$. Since Γ is consistent, $f(\alpha)$ is always a nonempty set; and so $\mathcal{M}$ is a standard deontic model. Moreover, $\mathcal{M} \models \Gamma$, and so since $\Gamma \vdash_{KD} \bigcirc A$, we have $\mathcal{M} \models \bigcirc A$; that is, $f(\alpha) \subseteq |A|$ for each $\alpha \in W$. From this and the definition of f, we can conclude that $|\overline{\Gamma}| \subseteq |A|$; and since $\overline{\Gamma}$ is itself consistent, Theorem 2 allows us to conclude that $\Gamma \vdash_F \bigcirc A$. ∎

4.3 Some variations

Although van Fraassen's account embodies an intuitively coherent and stable approach to reasoning in the presence of normative conflicts, it is not the only such approach. In this section, I simply mention some variations on van Fraassen's original account—one based on a skeptical reasoning strategy, one based a strategy of articulating the premise set, and one that combines these two ideas.

Suppose, first, that an agent is given $\bigcirc A$ and $\bigcirc \neg A$ as premises. We have assumed so far that these premises should yield as conclusions both $\bigcirc A$ and $\bigcirc \neg A$, though not the agglomerate $\bigcirc(A \wedge \neg A)$. But there is another option. It seems possible to imagine that, in this case, a sensible agent might want to resist the conclusion $\bigcirc A$ precisely because he has reason to believe that $\bigcirc \neg A$, and that he might likewise want to resist the conclusion that $\bigcirc \neg A$ because he has reason to believe that $\bigcirc A$. More generally, it seems possible to imagine that a sensible agent might want to conclude that a proposition ought to hold just in case he has reason for thinking that it ought to hold, and no reason for thinking otherwise.

There are, in fact, several different ways of making precise logical sense of this suggestion, but the one that seems most attractive can be set out quite simply, by modifying the idea underlying Theorem 1 to reflect a skeptical reasoning strategy. Rather than supposing that $\bigcirc A$ should follow from the premise set Γ just in case A belongs to some extension of corresponding default theory Δ_Γ, we might choose to conclude $\bigcirc A$ only if A belongs to each of these extensions. The resulting skeptical notion of deontic consequence, represented through the relation $\vdash_S$, is defined as follows.

Definition 4 $\Gamma \vdash_S \bigcirc A$ *if and only if $A \in \mathcal{E}$ for each extension $\mathcal{E}$ of Δ_Γ.*

It can then be shown, in parallel with Theorem 2, that $\Gamma \vdash_S \bigcirc A$ just in case $\mathcal{G} \vdash A$ for each consistent subset $\mathcal{G}$ of $\overline{\Gamma}$.

To illustrate this skeptical theory, let us consider an example suggested by Ruth Barcan Marcus [22] involving symmetrical but conflicting oughts: we suppose that an agent is faced with a situation in which he has equally weighty reasons for saving the lives of two identical twins, but is able to save only one. The premises conditioning the agent's deliberations in this situation can be represented through the set $\Gamma_1 = \{\bigcirc A, \bigcirc B\}$, where A represents the statement that the agent saves the first twin and B represents the statement that he saves the second, and where we take A and B, therefore, to be individually but not jointly consistent.

According to van Fraassen's theory, the premise set Γ_1 yields as consequences both $\bigcirc A$ and $\bigcirc B$—and then also, since in the situation A entails $\neg B$ and B entails $\neg A$—both $\bigcirc\neg A$ and $\bigcirc\neg B$. On this theory, then, no matter which of the twins the agent chooses to save, he is both neglecting to fulfill certain oughts that are generated by his premises and actually violating other generated oughts. According to the skeptical theory, on the other hand, none of the above ought statements follows from Γ_1, since none of the formulas A, B, $\neg A$, or $\neg B$ are found in all extensions of Δ_{Γ_1}. Instead, the strongest conclusion that can be drawn from Γ_1 is the statement $\bigcirc(A \vee B)$. The skeptical theory, then, will still constrain the agent's actions: he cannot simply walk away, saving neither twin. But as long as the agent saves one twin or the other, he will satisfy all the oughts generated by his premises.[7]

Like van Fraassen's consequence relation $\vdash_F$, the skeptical consequence relation $\vdash_S$ fails to satisfy the structural rules of reflexivity and cut; the examples supplied in Section 4.1 to establish these failures for $\vdash_F$ suffice here as well. However, unlike $\vdash_F$, the skeptical consequence relation $\vdash_S$ is itself nonmonotonic: although we might have

$$\bigcirc A \vdash_S \bigcirc A,$$

for example, we cannot have

$$\bigcirc A, \bigcirc\neg A \vdash_S \bigcirc A.$$

It is interesting to note, also, that this consequence relation validates a rule of agglomeration: given that $\Gamma \vdash_S \bigcirc A$ and $\Gamma \vdash_S \bigcirc B$, we can conclude that $\Gamma \vdash_S \bigcirc(A \wedge B)$. With the skeptical reasoning strategy, then, implications of the form

$$\bigcirc A, \bigcirc\neg A \vdash_S \bigcirc(A \wedge \neg A)$$

are avoided not, as before, through the failure of agglomeration, but instead, through the failure of monotonicity.

To motivate our second variation on van Fraassen's account, consider a situation in which an agent is issued the following command by some authoritative but confused source:

Square the circle, and go to the store for some milk.

Such a command might enter into the agent's reasoning through the premise set

$$\Gamma_2 = \{\bigcirc((A \wedge \neg A) \wedge B)\},$$

in which the formula $A \wedge \neg A$ represents the inconsistent first conjunct of the command, and the formula B represents its consistent second conjunct.

Now, what should the agent conclude? As we have noted, van Fraassen's suggestion is simply to ignore inconsistent oughts that occur among premises; according to this policy, no oughts at all could be concluded from Γ_2 other than the logical truths. But there seems to be another coherent option. We can imagine that an agent provided with an inconsistent ought statement as a premise, which cannot be satisfied entirely, might still wish to satisfy "as much" of this statement as possible. Returning to our example, the agent here

[7] An approach to Marcus's twins problem along the lines suggested here was sketched in Donagan [9].

might want to draw from the premise set Γ_2 at least the conclusion $\bigcirc B$, that he ought to go to the store for some milk.

In order to develop a reasoning strategy that embodies this idea—satisfying as much as possible even of inconsistent premises—we adapt to the present setting a procedure developed for another purpose by Belnap, and defended in detail in Anderson et al. [1, Section 82.4]. Given a set of premises, rather than drawing conclusions immediately, we first articulate the premise set in a way that is supposed to represent its intended meaning more explicitly, and then apply van Fraassen's approach to this articulated set of premises in order to reach the appropriate conclusions.

Adapting the procedure described in [1], we define the *articulation* of a premise set of ought statements as follows. Implication is first eliminated, so that the resulting formulas are written in $\wedge$, $\vee$, and $\neg$; and an occurrence of a subformula in an ought statement is defined as *positive* or *negative* depending on whether it lies within the scope of an even or odd number of negations. Given a premise set Γ, the articulated set Γ^* is then defined as the smallest superset of Γ that contains both $\bigcirc(\ldots B \ldots)$ and $\bigcirc(\ldots C \ldots)$ whenever it contains either a formula of the form $\bigcirc(\ldots(B \wedge C)\ldots)$ with the occurrence of the conjunction positive, or a formula of the form $\bigcirc(\ldots(B \vee C)\ldots)$ with the occurrence of the disjunction negative. As an example, the articulated set corresponding to Γ_2 would be

$$\Gamma_2^* = \Gamma \cup \{\bigcirc(A \wedge B), \bigcirc(\neg A \wedge B), \bigcirc A, \bigcirc\neg A, \bigcirc B\}.$$

It is now a straightforward matter to define the consequence relation $\vdash_{FA}$, representing an articulated variant of van Fraassen's original notion.

Definition 5 $\Gamma \vdash_{FA} \bigcirc A$ *if and only if* $\Gamma^* \vdash_F \bigcirc A$.

This new consequence notion would allow us to draw the following conclusions from Γ_2, none of which is generated by Fraassen's original definition: $\bigcirc(A \wedge B)$, $\bigcirc(\neg A \wedge B)$, $\bigcirc A$, $\bigcirc\neg A$, and $\bigcirc B$.

Of course, the two variations suggested here—the skeptical variation and the articulation variation—run in orthogonal directions, and they can be combined without interference: in reasoning from a premise set Γ, an agent might first extend this to the articulated set Γ^*, and then, reasoning skeptically, draw only those conclusions contained in each extension of the corresponding default theory Δ_{Γ^*}. The notion $\vdash_{SA}$ of articulated skeptical consequence that reflects this reasoning strategy can be defined in the obvious way.

Definition 6 $\Gamma \vdash_{SA} \bigcirc A$ *if and only if* $\Gamma^* \vdash_S \bigcirc A$.

Focusing again on Γ_2, an agent reasoning with this new notion of consequence would have to abandon all the conclusions listed above except $\bigcirc B$.

5 Conditional oughts

Much of our normative reasoning involves ought statements that are conditional rather than absolute, as in 'Given A, it ought to be that B', which we represent through the standard notation $\bigcirc(B/A)$.

Within the framework of modal logic, two general styles of analysis have been proposed for conditional oughts. Some writers have advanced an analysis involving a combination of an ordinary ought and an ordinary material conditional. Von Wright [35] originally suggested, for example, that a conditional ought statement should be analyzed through a formula of the form $\bigcirc(A \supset B)$, and A. N. Prior [26] suggested $A \supset \bigcirc B$; these two suggestions are compared by Jakko Hintikka in [13]. Other writers—such as Bengt Hansson [12], David Lewis [19, Section 5.1], and van Fraassen [33]—have suggested that a conditional ought operator should be analyzed instead as a primitive dyadic construction within the general framework of conditional logic; a number of proposals along these lines are surveyed by Lewis in [20]. As usual in conditional logics, this kind of analysis relies on a background ordering of the possible worlds, here intended to represent some relation of comparative goodness, or value; the basic idea is that $\bigcirc(B/A)$ should be true at a world just in case, with respect to the value ordering that is operative there, B is true at the best worlds in which A is true.

As it turns out, neither of these two general lines of approach seems promising as an analysis of prima facie ought statements.[8] At least when statements of the form $\bigcirc(B/A)$ are taken to express a prima facie oughts, problems then arise for both approaches concerning the degree of strengthening, or monotonicity, to be allowed in the antecedent of the conditional.

Consider the first approach. If $\bigcirc(B/A)$ is analyzed either as $A \supset \bigcirc B$ or as $\bigcirc(A \supset B)$, then conditional ought statements allow unrestricted strengthening in their antecedents; on either of these analyses, a rule of the form

$$\frac{\bigcirc(B/A)}{\bigcirc(B/A \wedge C)}$$

is admissible. (This is easy to see. The statement $(A \wedge C) \supset \bigcirc B$ is an immediate consequence of $A \supset \bigcirc B$; and since $(A \wedge C) \supset B$ is a consequence of $A \supset B$ and oughts are closed under consequence, $\bigcirc((A \wedge C) \supset B)$ follows from $\bigcirc(A \supset B)$.) If we follow the second general approach, on the other hand, analyzing conditional oughts within the general framework of conditional logic, the rule of antecedent strengthening must then be abandoned entirely. There is no way to conclude $\bigcirc(B/A \wedge C)$ from $\bigcirc(B/A)$; there is no reason to think, just because B holds at all the best worlds in which A is true, that it should hold also at the best worlds in which $A \wedge C$ is true.

It seems, however, that, at least for an analysis of prima facie reasoning, neither of these extreme approaches to strengthening in the antecedent of a conditional ought is correct: we do not want the rule of antecedent strengthening to hold without restriction, but neither do we want to abandon this rule entirely. This can be seen through an example drawn from [16]. Suppose that an agent, having studied the proprieties,

[8] It is often unclear to what extent the various writers concerned with conditional oughts intend for their analyses to apply to prima facie ought statements, rather than some other style of deontic conditional. However, at least Hintikka [13] explicitly suggests that statements of the form $\bigcirc(A \supset B)$ should be taken to represent prima facie oughts; and Robert Stalnaker and Richmond Thomason conjecture in [29] that the general techniques of conditional logic can be used to analyze prima facie obligation.

believes that his behavior at mealtime should be governed by the following three prima facie oughts:

> You ought not to eat with your fingers,
>
> You ought to put your napkin on your lap,
>
> If you are served asparagus, you ought to eat it with your fingers.

Taking an unconditional ought, in the usual way, as an ought conditional on the universal truth $\top$, we can then represent the premises governing the agent's deliberations as the set

$$\Gamma_3 = \{\bigcirc(\neg F/\top), \bigcirc(N/\top), \bigcirc(F/A)\}.$$

Now it seems, intuitively, that the third of these oughts should override the first when asparagus is served, so that, in this special case, the agent should not conclude that he ought not to eat with his fingers; but even if asparagus is served, nothing interferes with the second of these oughts, and so, even in this special case, the agent should still conclude that he ought to put his napkin on his lap. That is: from the premise set Γ_3, we would want to derive $\bigcirc(N/A)$, but not $\bigcirc(\neg F/A)$.

The only plausible way to derive $\bigcirc(N/A)$ in this situation, it seems, is by strengthening the antecedent of the second premise; a treatment of conditional oughts that simply rules out this kind of strengthening, such as those based on conditional logic, will not allow us to derive this conclusion. On the other hand, a treatment that allows unrestricted strengthening, such as those suggested by von Wright and Prior, will incorrectly yield $\bigcirc(\neg F/A)$ from the first premise. What is needed, apparently, is a certain amount of strengthening, but not too much: in order to model prima facie reasoning, we want to allow oughts formulated explicitly only for very general circumstances to apply also by default in more specific situations, unless they are overridden in those situations.

As far as I know, no treatment of conditional oughts based on any of the standard philosophical logics is able to model this kind of reasoning. It appears, for example, that the consequence relation associated with any appropriate theory would have to be nonmonotonic. To see this, suppose the statement

> If you are served asparagus, you ought to eat it with your fingers

were deleted from the list of rules displayed above, resulting in the new premise set

$$\Gamma_4 = \Gamma_3 - \{\bigcirc(F/A)\}.$$

In that case, since the general injunction against eating with one's fingers is not explicitly overridden in the particular situation in which asparagus is served, it should apply here by default also; and so we would now want to derive $\bigcirc(\neg F/A)$ from Γ_4. But as we have seen, with $\bigcirc(F/A)$ present as a premise, the general injunction is overridden, and so $\bigcirc(\neg F/A)$ is no longer acceptable. Even though Γ_4 yields $\bigcirc(F/A)$, then, and Γ_4 is a subset of Γ_3, we do not want Γ_3 to yield $\bigcirc(F/A)$; adding a premise leads us to withdraw a conclusion.

The idea of analyzing conditional oughts within the general semantic framework of conditional logic led to certain departures from the earlier treatment that involved mixing ordinary oughts with material conditionals; but at least as an analysis of prima facie oughts, these departures now seem to be both too radical and too conservative. The way in which the departures seem to be too radical is by forcing us entirely to abandon a rule of strengthening in the antecedent of a conditional ought; for it appears that we do want to admit a certain amount of strengthening as a default. But the departures also seem to be too conservative because, although these conditional logics do abandon the rule of strengthening, or antecedent monotonicity, within conditional ought statements, they nevertheless treat conditional oughts within an ordinary logical framework, with a monotonic consequence relation; and it appears that the consequence relation that governs our reasoning with prima facie oughts is itself nonmonotonic.

6 A nonmonotonic approach to conditional oughts

Because it seems to demand a nonmonotonic consequence relation, it is natural to hope that a useful theory of conditional oughts could be developed within the framework of nonmonotonic logic. In order to illustrate how such a development might proceed, this section first sketches a preliminary analysis of conditional oughts, which generalizes the theory of simple oughts set out earlier, and then explores some problems with the preliminary proposal.

6.1 Conditioned extensions

We focus on *ought contexts*: structures of the form $\langle \mathcal{W}, \Gamma \rangle$, like default theories, except that the set of defaults is replaced by a set Γ of conditional ought statements. The two components of an ought context are supposed to represent the background set of conditional oughts and the particular facts relevant to an agent's normative reasoning in that context.

Let us say that a conditional ought $\bigcirc(B/A)$ is *overridden* in the context $\langle \mathcal{W}, \Gamma \rangle$ just in case there is a statement $\bigcirc(D/C) \in \Gamma$ such that (i) $|\mathcal{W}| \subseteq |C|$, (ii) $|C| \subset |A|$ and (iii) $\mathcal{W} \cup \{D, B\}$ is inconsistent. The idea here is that a conditional ought should be overridden in some context whenever another ought is applicable, more specific than the original, and inconsistent with the original. In the definition, clause (i) tells us that us that $\bigcirc(D/C)$ is applicable in the context, in the sense that its antecedent condition is satisfied; clause (ii) tells us that $\bigcirc(D/C)$ is more specific than $\bigcirc(B/A)$, in the sense that the antecedent condition C is more restrictive than the antecedent condition A; and clause (iii) tells us $\bigcirc(D/C)$ and $\bigcirc(B/A)$ are inconsistent in the context, in the sense that their consequents B and D cannot both be realized along with the background information $\mathcal{W}$.

Using this characterization of the circumstances under which conditional oughts are overridden, we can now define a new kind of extension for ought contexts—known as a *conditioned extensions*—as follows.

Definition 7 *The set $\mathcal{E}$ is a conditioned extension of the ought context $\langle \mathcal{W}, \Gamma \rangle$ just in case there is a set $\mathcal{F}$ such that*

$$\begin{aligned} \mathcal{F} = \{B : \quad & \bigcirc(B/A) \in \Gamma, \\ & |\mathcal{W}| \subseteq |A|, \\ & \bigcirc(B/A) \textit{ is not overridden in } \langle \mathcal{W}, \Gamma \rangle, \\ & \neg B \notin \mathcal{E}\}, \end{aligned}$$

and $\mathcal{E} = Cn[\{\mathcal{W}\} \cup \mathcal{F}]$.

This is, of course, a fixed point definition; and so there is reason to suspect, just as certain default theories lack conventional extensions, that certain ought contexts might lack conditioned extensions. Fortunately, the suspicion turns out to be unfounded.

Theorem 6 *Every ought context* $\langle \mathcal{W}, \Gamma \rangle$ *has a conditioned extension* $\mathcal{E}$.

Proof Given $\langle \mathcal{W}, \Gamma \rangle$, first define

$$\mathcal{F}_1 = \{B : \quad \bigcirc(B/A) \in \Gamma, \quad |\mathcal{W}| \subseteq |A|, \quad \bigcirc(B/A) \text{ is not overridden in } \langle \mathcal{W}, \Gamma \rangle\},$$

and then let $\mathcal{F}_2$ be some maximal subset of $\mathcal{F}_1$ that is consistent with $\mathcal{W}$; these are guaranteed to exist. Let $\mathcal{E} = Cn[\{\mathcal{W}\} \cup \mathcal{F}_2]$. Evidently, $\mathcal{E}$ is a conditioned extension of $\langle \mathcal{W}, \Gamma \rangle$ if and only if $\mathcal{F}_2 = \mathcal{F}$ (where $\mathcal{F}$ is as defined in the text); and it is clear from the definition of $\mathcal{F}_2$ that $\mathcal{F}_2 = \mathcal{F}$ just in case $\mathcal{F}_2 = \mathcal{F}_1 \cap \{B : \neg B \notin \mathcal{E}\}$. So suppose first that $B \in \mathcal{F}_1$ and $\neg B \notin \mathcal{E}$. Then B is consistent with $\{\mathcal{W}\} \cup \mathcal{F}_2$, and so $B \in \mathcal{F}_2$, since $\mathcal{F}_2$ is maximal. Next, suppose $B \in \mathcal{F}_2$. Of course, $B \in \mathcal{F}_1$; and we must have $\neg B \notin \mathcal{E}$ as well, for otherwise we would have both B and $\neg B$ in $Cn[\{\mathcal{W}\} \cup \mathcal{F}_2]$, and so $\mathcal{F}_2$ would not be consistent with $\mathcal{W}$. ∎

Because conditioned extensions are guaranteed to exist, we can define a consequence relation $\vdash_{CF}$ between ought contexts and conditional ought statements in the following way.

Definition 8 $\langle \mathcal{W}, \Gamma \rangle \vdash_{CF} \bigcirc(B/A)$ *if and only if* $B \in \mathcal{E}$ *for some conditioned extension* $\mathcal{E}$ *of* $\langle \mathcal{W} \cup \{A\}, \Gamma \rangle$.

And it is then natural to define the consequences of a set of conditional oughts by reference to the special case in which the factual component of an ought context is empty.

Definition 9 $\Gamma \vdash_{CF} \bigcirc(B/A)$ *if and only if* $\langle \emptyset, \Gamma \rangle \vdash_{CF} \bigcirc(B/A)$.

Returning to our earlier asparagus example, we can now see that the present analysis yields the desired results. The unique conditioned extension of the ought context $\langle \{A\}, \Gamma_3 \rangle$ is $Cn[\{A, F, N\}]$; and so we have $\Gamma_3 \vdash_{CF} \bigcirc(N/A)$, as desired, but we do not have $\Gamma_3 \vdash_{CF} \bigcirc(\neg F/A)$. Just as in those theories based on the semantic framework of conditional modal logics, the present account of conditional oughts is nonmonotonic in the antecedent of the conditional. The ought context $\langle \{\top\}, \Gamma_3 \rangle$, for example, has as its unique conditioned extension the set $Cn[\{\neg F, N\}]$, and so we have $\Gamma_3 \vdash_{CF} \bigcirc(\neg F/\top)$; but, as mentioned, we do not have $\Gamma_3 \vdash_{CF} \bigcirc(\neg F/A)$. In addition, however, the consequence relation $\vdash_{CF}$ is itself nonmonotonic. The unique extension of $\langle \{A\}, \Gamma_4 \rangle$ is $Cn[\{A, \neg F, N\}]$, and so we have $\Gamma_4 \vdash_{CF} \bigcirc(\neg F/A)$; but monotonicity fails because, although Γ_4 is a subset of Γ_3, we do not have $\Gamma_3 \vdash_{CF} \bigcirc(\neg F/A)$.

The present account exhibits, also, several properties desirable in a conditional deontic logic. Since conditioned extensions are closed under logical consequence, the consequents of supported ought statements are closed under consequence as well: if $\Gamma \vdash_{CF} \bigcirc(B/A)$ and $B \vdash C$, then $\Gamma \vdash_{CF} \bigcirc(C/A)$. Again, by examining the definition of conditioned extensions we can see that conditional oughts are sensitive only to the propositions expressed by their antecedents, not to the particular sentences expressing those propositions: if $|A| = |B|$, then $\Gamma \vdash_{CF} \bigcirc(C/A)$ just in case $\Gamma \vdash_{CF} \bigcirc(C/B)$. And finally, an ought context $\langle \mathcal{W}, \Gamma \rangle$ will have an inconsistent extension if and only if the set $\mathcal{W}$ is itself inconsistent; and from this we conclude that $\Gamma \vdash_{CF} \bigcirc(\perp/A)$ if and only if $|A| = |\perp|$.[9]

It turns out, moreover, that the consequence relation $\vdash_{CF}$ is a conservative extension of the relation $\vdash_F$ described earlier, in the following sense:

Theorem 7 *Where Γ is a set of conditional oughts, let $\Gamma' = \{\bigcirc B : \bigcirc(B/A) \in \Gamma$ and $|A| = |\top|\}$. Then $\Gamma' \vdash_F \bigcirc C$ if and only if $\Gamma \vdash_{CF} \bigcirc(C/\top)$.*

Proof (sketch) We know by Theorem 1 that $\Gamma' \vdash_F \bigcirc C$ if and only if $C \in \mathcal{E}$ for some extension $\mathcal{E}$ of the default theory $\Delta_{\Gamma'}$. Reflection on the construction underlying Theorem 2.1 of Reiter [27] shows that $\mathcal{E}$ is an extension of $\Delta_{\Gamma'}$ just in case there is a set $\mathcal{F}$ such that

$$\begin{aligned}\mathcal{F} = \{B : \quad & \bigcirc B \in \Gamma', \\ & \neg B \notin \mathcal{E}\},\end{aligned}$$

and $\mathcal{E} = Cn[\mathcal{F}]$. It is easy to see that no conditional ought can be overridden in any context of the form $\langle\{\top\}, \Gamma\rangle$; and of course $|\top| \subseteq |A|$ if and only if $|A| = |\top|$. Therefore, we can conclude that $\mathcal{E}$ is an extension of $\Delta_{\Gamma'}$ just in case there is a set $\mathcal{F}$ such that

$$\begin{aligned}\mathcal{F} = \{B : \quad & \bigcirc(B/A) \in \Gamma, \\ & |\top| \subseteq |A|, \\ & \bigcirc(B/A) \text{ is not overridden in } \langle\{\top\}, \Gamma\rangle, \\ & \neg B \notin \mathcal{E}\},\end{aligned}$$

and $\mathcal{E} = Cn[\{\top\} \cup \mathcal{F}]$; that is, just in case $\mathcal{E}$ is a conditioned extension of $\langle\{\top\}, \Gamma\rangle$. The theorem then follows at once from the definition of the relation $\vdash_{CF}$. ∎

From this result and the discussion in Section 4, we can conclude that the consequence relation $\vdash_{CF}$, like $\vdash_F$, satisfies neither reflexivity nor cut.

6.2 Problems with the theory

This account of conditional deontic consequence exhibits a number of advantages not found in the usual modal approaches. The consequence relation $\vdash_{CF}$ is itself nonmonotonic, as is the antecedent place in derived conditional oughts; but unlike those accounts based on the the semantic framework of conditional logic, the current account does allow for a certain amount of strengthening, or monotonicity, in the antecedent of these derived oughts. And the theory generalizes the earlier treatment of reasoning in the presence of normative conflict, which already lies beyond the scope of modal approaches to deontic logic.

[9] The three properties described in this paragraph can be compared to the rules RCOEA, RCOM, and COD from Chellas[6, Section 10.2].

However, the current account of conditional deontic consequence is beset by several problems, and so can be taken, at best, only as a preliminary. I close simply by mentioning four issues that would have to be confronted in extending this account to a more complete theory.

First and most important, the current account does not allow any kind of transitivity, or chaining, across conditional oughts. We cannot derive $\bigcirc(C/A)$ from a premise set consisting of $\bigcirc(C/B)$ and $\bigcirc(B/A)$; and in particular, taking simple oughts as oughts conditional upon $\top$, we cannot derive $\bigcirc B$ from $\bigcirc(B/A)$ and $\bigcirc A$. Of course, this is similar to the situation in those accounts based on conditional logics, which also forbid transitivity of the conditional, and of the deontic conditional. However, the nonmonotonic framework allows for a new possibility that is not present in these standard logics—the possibility that transitivity should hold as a defeasible rule, subject to override. This is, in fact, exactly how transitivity is supposed to work in a number of application areas of nonmonotonic logics, such as the kind of reasoning supported by defeasible inheritance hierarchies. Here, we would want to conclude, for example, that Tweety flies given only the premises that Tweety is a bird and that birds fly; but we would allow this conclusion to be overridden by the additional information that Tweety is a penguin, where penguins are a particular class of birds that do not fly.

I think that it would be natural to incorporate this kind of defeasible transitivity also into an account of conditional deontic reasoning; but I have not attempted to do so here because the task of combining defeasible transitivity with a proper treatment of overriding (known in the inheritance literature as "preemption") would involve us in technical and conceptual issues intricate enough to obscure the main point of this paper. In spite of the efforts of a number of researchers—including Craig Boutilier [4], James Delgrande [7], Hector Geffner [11], and John Pollock [25]—I know of no solution to these problems for a language as expressive as propositional calculus that is generally accepted; and the matter is not settled even for the very simple language of inheritance hierarchies, as can be seen from Horty [17] and Touretzky et al. [30].

The second issue presented by the current account of conditional deontic consequence concerns the matter of reasoning with disjunctive antecedents. Where $\Gamma_5 = \{\bigcirc(C/A), \bigcirc(C/B)\}$, for example, it seems to some that we should be able to conclude from Γ_5 that $\bigcirc(C/A \vee B)$; but in fact, the only conditioned extension of $\langle\{A \vee B\}, \Gamma_5\rangle$ is $Cn[\{A \vee B\}]$, and so the present account does not yield this result. There seem to be two strategies available for handling this issue. First, we could try to modify the account so as to yield the kind of results that some view as desirable. Problems involving disjunctive reasoning are familiar in the context of default logic; and several proposals, such as that of Gelfond et al. [21], have been put forth for modifying standard default logic so that it allows defaults to be applied on the basis of disjunctive information. Given the similarity between conditional extensions of ought contexts and ordinary extensions of default theories, it should not be too difficult to adapt these proposals to the present context; but it is not simply an exercise, since the adaptation would have to involve extending the notion of overriding to apply properly to disjunctive antecedents. Alternatively, however, we might deny that inferences such as that from $\bigcirc(C/A)$ and $\bigcirc(C/B)$ to $\bigcirc(C/A \vee B)$ should be validated in a deontic setting; a discussion of the conditions under which this kind of inference seems to fail can be found in Horty [15].

The third issue presented by the current theory concerns a detail in the treatment of overridden oughts.

According to the this theory, an conditional ought can be overridden only by a single opposing statement, which is both applicable in the context and more specific. However, there are cases in which it is natural to suppose that an ought, although not overridden by a single opposing rule, might be overridden by a set of opposing rules. For example, let $\Gamma_6 = \{\bigcirc(Q/\top), \bigcirc(\neg(P \wedge Q)/A), \bigcirc(P/A)\}$. Here, it seems that in the context $\langle\{A\}, \Gamma_6\rangle$, the first rule should be overridden by the second two taken together, although it is not overridden by either individually.

The final problem concerns yet another detail in the present treatment of overriding. Suppose an ought statement is overridden by another which is itself overridden. What is the status of the original? According to the present treatment, it remains out of play; but it is also possible to imagine that the original rule should then be reinstated. As an example, let $\Gamma_7 = \{\bigcirc(Q/\top), \bigcirc(P/A), \bigcirc(\neg P/A \wedge B)\}$, and consider the context $\langle\{W\}, \Gamma_7\rangle$, where W is the formula $(A \wedge B) \wedge \neg(P \wedge Q)$. Of course, the first rule in Γ_7 is overridden in this context by the second, but the second is likewise overridden by the third. Since overridden rules remain out of play, according to the the present treatment, this context has $Cn[\{W, \neg P\}]$ as its only conditioned extension; and so we do not have $\bigcirc(Q/W)$. But it does not seem unreasonable to modify the present treatment so that the rule $\bigcirc(Q/\top)$ is reinstated in this context, since the rule that overrides it is itself overridden. In that case, we would have $Cn[\{W, \neg P, Q\}]$ as a conditioned extension; and so we would be able to derive $\bigcirc(Q/W)$ from Γ_7. The issue of reinstatement in inheritance hierarchies is explored in detail in Horty [17] and in Touretzky et al. [31].

The problems pointed out here with the current account of conditional deontic consequence are serious, but I do not feel that they should lead us to abandon the project of designing a conditional deontic logic using the techniques of nonmonotonic logic. In fact, none of these problems is unique to the deontic interpretation of the background nonmonotonic theory; instead, they reflect more general difficulties in nonmonotonic reasoning, which surface here just as they surface elsewhere. Of course, it is impossible to offer a final evaluation of the nonmonotonic approach to conditional deontic reasoning until these issues with the underlying logical framework are resolved. But the approach does seem to be promising; and it may be that, in bringing the techniques of nonmonotonic logic into contact with the new data provided by normative reasoning, we will not only discover new possibilities for the construction of deontic logics, but gain a deeper understanding of the underlying nonmonotonic logics as well.

Acknowledgments

This paper is an extensive rewrite of my earlier [14], incorporating also some material from my [16]. I am very grateful for discussions or correspondence concerning this material to Nuel Belnap, Craig Boutilier, Jon Doyle, David Etherington, Matt Ginsberg David Makinson, Judea Pearl, Michael Slote, Rich Thomason, David Touretzky, and Johan van Benthem. My work in this area has been supported by the National Science Foundation through Grants No. IRI-9003165 and IRI-8907122, by the Army Research Office through Grant No. DAAL-03-88-K0087, and by the National Endowment for Humanities through a Fellowship for University Teachers.

References

[1] Alan Anderson, Nuel Belnap, and J. Michael Dunn. *Entailment: The Logic of Relevance and Necessity, Volume 2.* Princeton University Press, 1992.

[2] Nuel Belnap. How a computer should think. In Gilbert Ryle, editor, *Contemporary Aspects of Philosophy*, pages 30–56. Oriel Press, 1977.

[3] Nuel Belnap. A useful four-valued logic. In J. Michael Dunn and G. Epstein, editors, *Modern Uses of Multiple-valued Logic*, pages 8–37. D. Reidel Publishing Company, 1977.

[4] Craig Boutilier. *Conditional Logics for Default Reasoning and Belief Revision.* PhD thesis, Computer Science Department, University of Toronto, 1992. Available as Technical Report KRR-TR-92-1.

[5] Brian Chellas. Conditional obligation. In Søren Stenlund, editor, *Logical Theory and Semantic Analysis*, pages 23–33. D. Reidel Publishing Company, 1974.

[6] Brian Chellas. *Modal Logic: An Introduction.* Cambridge University Press, 1980.

[7] James Delgrande. An approach to default reasoning based on a first-order conditional logic: revised report. *Artificial Intelligence*, 36:63–90, 1988.

[8] Daniel Dennett. The moral first aid manual. In *The Tanner Lectures on Human Values, Volume VIII*, pages 119–147. Cambridge University Press, 1988.

[9] Alan Donagan. Consistency in rationalist moral systems. *The Journal of Philosophy*, 81:291–309, 1984.

[10] Philippa Foot. Moral realism and moral dilemma. *Journal of Philosophy*, 80:379–398, 1983.

[11] Hector Geffner. *Default Reasoning: Causal and Conditional Theories.* PhD thesis, Computer Science Department, University of California at Los Angeles, 1989.

[12] Bengt Hansson. An analysis of some deontic logics. In *Deontic Logic: Introductory and Systematic Readings*, pages 121–147. D. Reidel Publishing Company, 1971.

[13] Jakko Hintikka. Deontic logic and its philosophical morals. In *Models for Modalities: Selected Essays*, pages 184–214. D. Reidel Publishing Company, 1969.

[14] John Horty. Deontic logic as founded on nonmonotonic logic. *Annals of Mathematics and Artificial Intelligence*, 9:69–91, 1993.

[15] John Horty. *Agency, Deontic Logic, and Utilitarianism.* Book Manuscript, Philosophy Department and Institute for Advanced Computer Studies, University of Maryland, 1994.

[16] John Horty. Moral dilemmas and nonmonotonic logic. *Journal of Philosophical Logic*, 23:35–65, 1994.

[17] John Horty. Some direct theories of nonmonotonic inheritance. In D. Gabbay, C. J. Hogger, and J. A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning*, pages 111–187. Oxford University Press, 1994.

[18] E. J. Lemmon. Moral dilemmas. *Philosophical Review*, 70:139–158, 1962.

[19] David Lewis. *Counterfactuals.* Oxford University Press, 1973.

[20] David Lewis. Semantic analyses for dyadic deontic logic. In Søren Stenlund, editor, *Logical Theory and Semantic Analysis*, pages 1–14. D. Reidel Publishing Company, 1974.

[21] H. Przymusinska M. Gelfond, V. Lifschitz and M. Truszczynski. Disjunctive defaults. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference (KR-91).* Morgan Kaufmann Publishers, 1991.

[22] Ruth Barcan Marcus. Moral dilemmas and consistency. *Journal of Philosophy*, 77:121–136, 1980.

[23] Drew McDermott. Non-monotonic logic II. *Journal of the Association for Computing Machinery*, 29:33–57, 1982.

[24] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann Publishers, 1988.

[25] John Pollock. How to reason defeasibly. *Artificial Intelligence*, 57:1–42, 1992.

[26] Arthur Prior. *Formal Logic.* Oxford University Press, 1962.

[27] Raymond Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[28] Jean Paul Sartre. *L'Existentialisme est un Humanisme.* Nagel, 1946. Translated as "Existentialism is a Humanism" in *Existentialism from Dostoevsky to Sartre*, W. Kaufmann (ed.), Meridian Press, 1975.

[29] Robert Stalnaker and Richmond Thomason. A semantic analysis of conditional logic. *Theoria*, 36:23–42, 1970.

[30] David Touretzky, John Horty, and Richmond Thomason. A clash of intuitions: the current state of nonmonotonic multiple inheritance systems. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence(IJCAI-87)*, pages 476–482. Morgan Kaufmann, 1987.

[31] David Touretzky, Richmond Thomason, and John Horty. A skeptic's menagerie: conflictors, preemptors, reinstaters, and zombies in nonmonotonic inheritance. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI-91)*, pages 478–483. Morgan Kaufmann Publishers, 1991.

[32] Edna Ullman-Margalit. On presumption. *The Journal of Philosophy*, 70:143–163, 1973.

[33] Bas van Fraassen. The logic of conditional obligation. *The Journal of Philosophical Logic*, 72:417–438, 1972.

[34] Bas van Fraassen. Values and the heart's command. *The Journal of Philosophy*, 70:5–19, 1973.

[35] G. H. von Wright. Deontic logic. *Mind*, 60:1–15, 1951.

[36] Bernard Williams. Ethical consistency. *Proceedings of the Aristotelian Society*, 39 (supplemental):103–124, 1965. A revised version appears in *Problems of the Self: Philosophical Papers 1956–1972*, Cambridge University Press, 1973, pages 166–186.