Philosophers' Imprint

volume 7, no. 3
april 2007

REASONS AS DEFAULTS

John F. Horty

Department of Philosophy & Institute for Advanced Computer Studies
University of Maryland, College Park

<*www.philosophersimprint.org/007003/*>

1. Introduction

Much of the recent literature on reasons is focused on a common range of issues, concerning, for example, the relation between reasons and motivation, desires, and values, the issue of internalism versus externalism in the theory of reasons, or the objectivity or reasons. This paper is concerned with a different, and orthogonal, set of questions: What are reasons, and how do they support actions or conclusions? Given a collection of individual reasons, possibly suggesting conflicting actions or conclusions, how can we determine which course of action, or which conclusion, is supported by the collection as a whole? What is the mechanism of support?

I begin by considering one possible line of response, which I refer to as the *weighing conception*, since it is based on the idea that reasons support actions or conclusions by contributing a kind of normative or epistemic weight, and that the goal is then to select those options whose overall weight is greatest. This general idea is almost certainly ancient, but we know that it goes back at least to 1772, where we can find a version of the weighing conception described with some elegance in a letter from Benjamin Franklin to his friend Joseph Priestley, the chemist. Priestley had written to Franklin for advice on a practical matter. In his reply, Franklin regrets that he has no help to offer on the specific matter at hand, since he is not sufficiently familiar with the facts, but recommends a general technique for reaching decisions in situations of the kind facing his friend:

> My way is to divide half a sheet of paper by a line into two columns; writing over the one Pro, and over the other Con. Then, during the three or four days consideration, I put down under the different heads short hints of the different motives, that at different times occur to me, for or against the measure. When I have thus got them all together in one view, I endeavor to estimate their respective weights; and where I find two, one on each side, that seem equal, I strike them both out. If I find a reason pro equal to some two reasons con, I strike out the three.

> If I judge some two reasons con, equal to three reasons pro, I strike out the five; and thus proceeding I find at length where the balance lies . . . (Franklin 1772, pp. 348–349)

I suspect that most of us would now regard Franklin's picture as quixotic, or at least extraordinarily optimistic, both in its assumption that the force of reasons can be captured through an assignment of numerical weights, and in the accompanying assumption that practical reasoning can then be reduced to an application of arithmetic operations (indeed, Franklin goes on to characterize his technique as a "moral or prudential algebra"). But neither of these assumptions is actually essential. A number of contemporary writers are still willing to endorse what they think of as a more general form of the weighing conception—like Franklin's, but without the commitment to a precise numerical representation, or to arithmetic operations. According to this more general view, reasons can still be thought of as supporting conclusions by contributing weights, of a sort; the weights contributed by different reasons can still be thought of as combined in some way, even if the combination function is not arithmetic; and these combined weights can still be balanced against each other, with the correct outcome defined as that whose weight is greatest. One example of this generalized weighing conception, distinguished by its exceptional clarity, can be found in a recent paper by John Broome, who summarizes his position as follows:

> Each reason is associated with a metaphorical weight. This weight need not be anything so precise as a number; it may be an entity of some vaguer sort. The reasons for you to ϕ and those for you not to ϕ are aggregated or weighed together in some way. The aggregate is some function of the weights of the individual reasons. The function may not be simply additive . . . It may be a complicated function, and the specific nature of the reasons may influence it. Finally, the aggregate comes out in favor of your ϕing, and that is why you ought to ϕ. (Broome 2004, p. 37)

My objection to this picture is not so much that it is a version of the weighing conception—although, in fact, the theory I present in this paper is set out explicitly as an alternative to this view. Instead, my objection is that the generalized weighing conception as described here is simply incomplete as an account of the way in which reasons support conclusions. Broome distances himself from the more objectionable features of the quantitative weighing conception—numbers, additive functions—but fails to tell us what should take their place. If the weights associated with reasons are not numbers, what are they; what are these entities of a vaguer sort? If these weights are not aggregated through simple addition, how are they aggregated; what is this more complicated function?

In raising this objection, I do not mean to criticize Broome, who surely does not intend to present anything like a complete account of his generalized weighing conception in the paper I have cited, but only to sketch the view before getting on with other work. Nevertheless, I do feel that the objection highlights a real problem for contemporary advocates of the generalized weighing conception, and one that I have not seen addressed. Once we move past the level of a rough outline, it will not do to say only that reasons lend some kind of weight to conclusions, and that these weights are assembled somehow. A theory of the relation between reasons and their outcomes should be subject to the same standards of rigor that Frege brought to the study of the relation between premises and their consequences.

Let us return to our initial questions: What are reasons, and how do they support conclusions? My answer is that reasons are provided by defaults, and that they support conclusions in accord with the logic of default reasoning (sometimes known as nonmonotonic logic, sometimes as defeasible logic). The goal of this paper is to articulate and begin to develop this answer.

Although there is no single theory that we can now point to as the correct logic for default reasoning, I begin by describing what seems to me to be one particularly useful way of developing such a logic. This logic is presented here only in as much detail as necessary to show that there really is a concrete theory at work, to provide some idea of the shape of that theory, and also of the issues it raises; the overall theory is set out more carefully, and explored from a technical perspective, in my (2007).[1] After presenting this default logic, I then show how it can be elaborated to deal with certain issues involved in developing a more robust theory of reasons. I focus on two such issues: first, situations in which the priority relations among reasons, or defaults, themselves seem to be established through default reasoning; second, the treatment of undercutting defeat and exclusionary reasons. Finally, and by way of application, I show how the resulting account can shed some light on a topic in the theory of reasons that has recently attracted a good deal of attention: Jonathan Dancy's interesting and influential argument from reason holism to a form of extreme particularism in moral theory.

2. A theory of default reasoning

2.1 Default theories and scenarios

We take as background an ordinary propositional language, with $\supset$, $\wedge$, $\vee$ and $\neg$ as the usual propositional connectives, and with $\top$ as a special constant representing truth. The turnstile $\vdash$ indicates ordinary logical consequence.

Now against this background, let us begin with a standard example, known as the Tweety Triangle.[2] If an agent is told only that Tweety is

a bird, it would be natural for that agent to conclude that Tweety is able to fly. Our everyday reasoning seems to be governed by a general default according to which birds, as a rule, are able to fly; and on the view recommended here, it is this default, instantiated for Tweety, that provides a reason for the conclusion. But suppose the agent is told, in addition, that Tweety is a penguin. There is also a default according to which penguins, as a rule, are not able to fly, which now provides a reason for a conflicting conclusion. Because the default about penguins is stronger than the default about birds, it is natural to suppose that the first of these reasons is defeated by the second, so that the agent should withdraw its initial judgment and conclude instead that Tweety cannot fly.[3]

Where A and B are formulas from the background language, we let $A \rightarrow B$ represent the *default rule* that allows us to conclude B, by default, whenever it has been established that A. To illustrate: if B stands for the statement that Tweety is a bird, and F for 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. This particular default can be thought of as an instance for Tweety of the general default

$$Bird(x) \rightarrow Fly(x),$$

telling us that, as a rule, birds are able to fly (to get from this general default to the particular instance $B \rightarrow F$, think of B as an abbreviation for the statement *Bird*(*Tweety*) and of F as an abbreviation for *Fly*(*Tweety*)). It is, in many ways, easier to understand general default rules like this—*defeasible generalizations*—than it is to understand their particular instances. However, in order to avoid the complexities involved in a treatment of variables and instantiation in default logic,

[1] The particular logic presented here has its roots in two earlier sources. The first is the original default logic of Reiter (1980), one of the most widely applied formalisms for nonmonotonic reasoning, and arguably the most successful; the second is a body of literature on the semantics of nonmonotonic inheritance networks, initiated by Touretzky (1986), developed by a number of authors, and reviewed in my (1994b). The current theory is mapped out along the general lines suggested by Reiter, but includes an account of priority relations among default rules based on the treatment of this topic found in the literature on inheritance reasoning.

[2] It is called this because of its triangular shape when depicted as an inheritance network, a graphical representation of default relations among classes of entities; see my (1994b) for an overview of nonmonotonic inheritance reasoning.

[3] In an effort to find language that is both gender neutral and unobtrusive, I often assume that the agent is an impersonal reasoning device, such as a computer, which can appropriately be referred to with the pronoun 'it'.

we restrict ourselves in this paper to a propositional language, and therefore focus only on particular defaults, rather than the defeasible generalizations they instantiate.

Throughout the discussion, we will be slipping back and forth, rather casually, between what might be called *practical* and *epistemic* reasons—reasons for actions, versus reasons for conclusions. The default that Tweety flies given that he is a bird might be classified as providing an epistemic reason, supporting the proposition that Tweety flies. By contrast, the default that I ought to have lunch with you given that I promised to do so (a particular instance of the defeasible generalization that I ought to do whatever I promise) is most naturally interpreted as providing a practical reason. It does not support the conclusion that I will have lunch with you, but provides me with a reason for doing so.

Various theses could be advanced concerning the relation between these two kinds of reasons. One such thesis is that epistemic reasons should be subsumed as a species under the genus of practical reasons. On this view, our reason for concluding that Tweety flies does not, if fact, support a proposition, but actually recommends an action: *concluding* that Tweety flies. Another thesis is that practical reasons should be subsumed as a species under the genus of epistemic reasons. On this view, my reason to have lunch with you does not recommend an action but actually supports a proposition: that I *ought* to have lunch with you. Yet a third thesis is that neither practical nor epistemic reasons can be assimilated to the other, but that they are simply distinct kinds of reasons, though sharing many important properties.

The account set out here is intended to be independent of these theses, or others, concerning the relation between practical and epistemic reasons; it can be adapted, I believe, to accommodate a variety of different positions on the topic. Although we will not address the relations between practical and epistemic reasons, or the complicated topic of their interactions, we will, at various points, be discussing each of these two kinds of reasons individually, and will then use the same notation in both cases—relying on context to indicate whether the conclusion B in a default of the form $A \rightarrow B$ is supposed to represent a supported proposition or a recommended action. For expository convenience, simply because the theory is more naturally motivated in this way, we will begin by focusing on epistemic reasons almost exclusively, and then turn to practical reasons later on.

We assume two functions—*Premise* and *Conclusion*—that pick out the premises and conclusions of default rules: if δ is the default $A \rightarrow B$, for example, then $\mathit{Premise}(\delta)$ is the statement A and $\mathit{Conclusion}(\delta)$ is the statement B. The second of these functions is lifted from individual defaults to sets of defaults in the obvious way, so that, where $\mathcal{D}$ is a set of defaults, we have

$$\mathit{Conclusion}(\mathcal{D}) = \{\mathit{Conclusion}(\delta) : \delta \in \mathcal{D}\}$$

as the set of their conclusions.

As we have seen, some defaults have greater strength, or higher priority, than others; some reasons are better than others. In order to represent this information, we introduce an ordering relation $<$ on the set of defaults, with $\delta < \delta'$ taken to mean that the default δ' has a higher priority than δ. We assume that this priority ordering on defaults is transitive, so that $\delta < \delta'$ and $\delta' < \delta''$ implies $\delta < \delta''$, and also irreflexive, so that $\delta < \delta$ always fails; such an ordering relation is referred to as a *strict partial ordering*.

The priority relations among defaults can have different sources. In the Tweety Triangle, for example, the priority of the default about penguins over the default about birds has to do with specificity: a penguin is a specific kind of bird, and so information about penguins in particular takes precedence over information about birds in general. But there are also priority relations that have nothing to do with specificity. Reliability is another source. Both the weather channel and the arthritis in my left knee provide reasonably reliable predictions of oncoming precipitation, but the weather channel is more reliable. And once we move from epistemic to practical reasons, then authority provides yet another source for priority relations. National laws typically override state or provincial laws, and more recent court decisions have more authority than older decisions. Direct orders override standing orders, and orders from the Colonel override orders from the Major.

We will begin, in this section, by considering the special case in which all priority relations among defaults are fixed in advance, so that there is no need to consider either the source of these priority relations or the way in which they are established, but only their effect on the conclusions reached through default reasoning. Formally, where $\mathcal{D}$ is a set of defaults subject to the strict partial ordering $<$, and where $\mathcal{W}$ is, in addition, some set of ordinary formulas, we define a *fixed priority default theory* as a structure of the form $\langle \mathcal{W}, \mathcal{D}, < \rangle$. Such a structure—a body of ordinary information together with an ordered set of defaults—represents the initial data provided to the agent as a basis for its reasoning.

Most research in nonmonotonic logic is motivated by the epistemic interpretation of defaults, and so concentrates on the problem of characterizing the belief sets supported by default theories. The defaults themselves are thought of as rules for extending the beliefs derivable from a set of formulas beyond their classical consequences, and for this reason, the belief sets they support are often referred to as *extensions*. We will concentrate here, however, not on extensions themselves, but on *scenarios*, where a scenario based on a default theory $\langle \mathcal{W}, \mathcal{D}, < \rangle$ is defined simply as a particular subset $\mathcal{S}$ of the set $\mathcal{D}$ of defaults contained in the theory. From an intuitive standpoint, a scenario is supposed to represent the set of defaults that have been accepted by the agent, at some stage of the reasoning process, as providing sufficient support for their conclusions.

The concept of a scenario has a natural interpretation under both the epistemic and the practical readings of default rules. Under the epistemic reading, the agent, in selecting a scenario, can be thought of as choosing the defaults it will use to extend its initial information to a full belief set: where $\mathcal{S}$ is a scenario based on $\langle \mathcal{W}, \mathcal{D}, < \rangle$, we can define the belief set that is *generated* by this scenario as the set of formulas derivable from its conclusions together with the agent's initial information—that is, the set of formulas derivable from $\mathcal{W} \cup Conclusion(\mathcal{S})$. Under the practical reading, with defaults understood as recommending actions rather than supporting propositions, the set of conclusions of the defaults belonging to a scenario $\mathcal{S}$—that is, the set $Conclusion(\mathcal{S})$—can be taken as the set of actions that the agent has settled upon.

Our initial task is to characterize, as we will say, the *proper scenarios*—those sets of defaults that might ultimately be accepted by an ideal reasoning agent on the basis of the information contained in some ordered default theory. With this notion in hand, the extensions of default theories, ideal belief sets, can be defined as those belief sets that are generated by proper scenarios; or from a practical perspective, a proper scenario can be taken to specify a set of actions that a rational individual might decide to perform.

2.2 *Binding defaults*

We begin with the concept of a binding default. If defaults provide reasons, then the binding defaults are supposed to represent those that provide *good* reasons, in the context of a particular scenario. This reference to a scenario is not accidental: according to the theory developed here, the set of defaults that an agent might take as providing good reasons depends on the set of defaults it already accepts, the agent's current scenario.

The concept of a binding default is defined in terms of three preliminary ideas, which we turn to first—triggering, conflict, and defeat.

Not every default, of course, is even applicable in every scenario. The default that birds fly, for example, provides no reason at all for an agent to conclude that Tweety flies unless the agent is already committed to the proposition that Tweety is a bird. The triggered defaults, those that are applicable in a particular scenario, are simply those whose premises are entailed under that scenario—those defaults, that is, whose premises follow from the agent's initial information together with the conclusions of the defaults that the agent has already accepted.

Definition 1 (Triggered defaults) Where $\mathcal{S}$ is a scenario based on the fixed priority default theory $\langle \mathcal{W}, \mathcal{D}, < \rangle$, the defaults from $\mathcal{D}$ that are *triggered* in $\mathcal{S}$ are those belonging to the set

$$\mathit{Triggered}_{\mathcal{W},\mathcal{D},<}(\mathcal{S}) = \{\delta \in \mathcal{D} : \mathcal{W} \cup \mathit{Conclusion}(\mathcal{S}) \vdash \mathit{Premise}(\delta)\}.$$

To illustrate, let B, F, and W stand, respectively, for the propositions that Tweety is a bird, that Tweety flies, and that Tweety has wings; and let δ_1 and δ_2 stand for the defaults $B \rightarrow F$ and $F \rightarrow W$, instances for Tweety of the general defaults that birds fly and that flying animals tend to have wings. Imagine that an agent is provided with the ordered default theory $\langle \mathcal{W}, \mathcal{D}, < \rangle$ as its initial information, where $\mathcal{W} = \{B\}$, $\mathcal{D} = \{\delta_1, \delta_2\}$, and the ordering $<$ is empty; and suppose the agent has not yet accepted any of the defaults from $\mathcal{D}$, so that its initial scenario is simply $\mathcal{S}_0 = \emptyset$. We then have $\mathit{Triggered}_{\mathcal{W},\mathcal{D},<}(\mathcal{S}_0) = \{\delta_1\}$ so that, in this initial scenario, only δ_1 is triggered, providing the agent with a reason for its conclusion, the proposition F. Now suppose the agent does in fact accept this default, and so moves to the new scenario $\mathcal{S}_1 = \{\delta_1\}$. Then since $\mathit{Triggered}_{\mathcal{W},\mathcal{D},<}(\mathcal{S}_1) = \{\delta_1, \delta_2\}$, the default δ_2 is now triggered as well, providing a reason for the new conclusion W.

This discussion of triggered defaults, those that can be thought of as providing the agent with reasons, leads to a terminological question: should these reasons then be identified with the defaults themselves, or with propositions? Suppose, as in our example, that the agent's background theory contains the default $B \rightarrow F$, an instance for Tweety of the general default that birds fly, together with B, the proposition that Tweety is a bird, so that the default is triggered. In this case, it seems plain that the agent has a reason to conclude that Tweety flies. But how, exactly, should this reason be reified? Should it be identified with the default $B \rightarrow F$ itself, or with the proposition B?

This question, like many questions concerning reification, is somewhat artificial. Evidently, both the default and the proposition are involved in providing the agent with a reason for concluding that Tweety flies. The default would have no bearing if it were not triggered by some fact, a true proposition; the fact would be nothing but an incidental feature of the situation if it did not trigger some default. When it comes to reification, then, the reason relation could be projected in either direction, toward defaults or propositions, and the choice is largely arbitrary.

Nevertheless, it seems to correspond most closely to our ordinary usage to reify reasons as propositions. The present paper will therefore be based on an analysis according to which *reasons are identified with the premises of triggered defaults*; and we will speak of these triggered defaults, not as reasons themselves, but as *providing* certain propositions—their premises—as reasons for their conclusions. To illustrate: in the case of our example, we will say that B, the fact the Tweety is a bird, is a reason for concluding that Tweety flies, and that this reason is provided by the default $B \rightarrow F$.

Triggering is a necessary condition that a default must satisfy in order to be classified as binding in a scenario, but it is not sufficient. Even if some default is triggered, it might not be binding, all things considered; two further aspects of the scenario could interfere.

The first is easy to describe. A default will not be classified as binding in a scenario, even if it happens to be triggered, if that default is conflicted—that is, if the agent is already committed to the negation of its conclusion.

Definition 2 (Conflicted defaults) Where $\mathcal{S}$ is a scenario based on the fixed priority default theory $\langle \mathcal{W}, \mathcal{D}, < \rangle$, the defaults from $\mathcal{D}$ that are *conflicted* in $\mathcal{S}$ are those belonging to the set

$$\mathit{Conflicted}_{\mathcal{W},\mathcal{D},<}(\mathcal{S}) = \{\delta \in \mathcal{D} : \mathcal{W} \cup \mathit{Conclusion}(\mathcal{S}) \vdash \neg \mathit{Conclusion}(\delta)\}.$$

The intuitive force of this restriction can be illustrated through another standard example, known as the Nixon Diamond.[4] Let Q, R, and

[4] Again, because its depiction as an inheritance network has the shape of a diamond.

P stand for the respective propositions that Nixon is a Quaker, that Nixon is a Republican, and that Nixon is a pacifist; and let δ_1 and δ_2 represent the defaults $Q \rightarrow P$ and $R \rightarrow \neg P$, instances of the generalizations that Quakers tend to be pacifists and that Republicans tend not to be pacifists. Imagine that the agent's initial information is provided by the theory $\langle \mathcal{W}, \mathcal{D}, < \rangle$, where $\mathcal{W} = \{Q, R\}$, $\mathcal{D} = \{\delta_1, \delta_2\}$, and the ordering $<$ is again empty; and suppose once more that the agent has not yet accepted either of these two defaults, so that its initial scenario is $\mathcal{S}_0 = \emptyset$.

In this situation, we have $\mathit{Triggered}_{\mathcal{W},\mathcal{D},<}(\mathcal{S}_0) = \{\delta_1, \delta_2\}$; the default δ_1 provides a reason for the conclusion P, and the default δ_2 provides a reason for the conclusion $\neg P$. Although these two defaults support conflicting conclusions, neither is conflicted in the initial scenario: $\mathit{Conflicted}_{\mathcal{W},\mathcal{D},<}(\mathcal{S}_0) = \emptyset$. The agent must therefore find some way of dealing with the conflicting reasons presented by its epistemic state. Now suppose that, on whatever grounds, the agent decides to favor one of these two defaults—say δ_1, with the conclusion P—and so moves to the new scenario $\mathcal{S}_1 = \{\delta_1\}$. In this new scenario, the other default will now be conflicted: $\mathit{Conflicted}_{\mathcal{W},\mathcal{D},<}(\mathcal{S}_1) = \{\delta_2\}$. From the standpoint of the new scenario, the reason provided by δ_2 can no longer be classified as a good reason, since the agent has already settled on a default that provides a reason for a conflicting conclusion.

The second restriction governing the notion of a binding default holds that, even if it is triggered, a default cannot be classified as binding if it happens to be defeated. Although the concept of a defeated default is considerably more difficult to define than that of a conflicted default, the basic idea is simple enough: an agent should not accept a default in the face of a stronger default supporting a conflicting conclusion.

This idea can be illustrated by returning to our original Tweety Triangle, with P, B, and F representing the propositions that Tweety is a penguin, that Tweety is a bird, and that Tweety flies. Let us take δ_1 and δ_2 as the defaults $B \rightarrow F$ and $P \rightarrow \neg F$, instances of the general rules that birds fly and that penguins do not. Imagine that the agent is provided with the theory $\langle \mathcal{W}, \mathcal{D}, < \rangle$ as its initial information, where $\mathcal{W} = \{P, B\}$, $\mathcal{D} = \{\delta_1, \delta_2\}$, and now $\delta_1 < \delta_2$; the default about penguins has higher priority than the default about birds. And suppose once again that the agent has not yet accepted either of these two defaults, so that its initial scenario is $\mathcal{S}_0 = \emptyset$.

In this situation, we again have $\mathit{Triggered}_{\mathcal{W},\mathcal{D},<}(\mathcal{S}_0) = \{\delta_1, \delta_2\}$; the default δ_1 provides a reason for concluding F, while the default δ_2 provides a reason for concluding $\neg F$. And we again have $\mathit{Conflicted}_{\mathcal{W},\mathcal{D},<}(\mathcal{S}_0) = \emptyset$; neither of these defaults is itself conflicted. Nevertheless, it does not seem that the agent should be free, as in the previous Nixon Diamond, to settle this conflict however it chooses. Here, it seems appropriate to say, on intuitive grounds, that the default δ_1, supporting the conclusion F, is defeated by the stronger default δ_2, since this default is also triggered, and since it supports the conflicting conclusion $\neg F$.

Motivated by this example, it is natural to propose a definition according to which a default is defeated in a scenario if that scenario triggers some stronger default with a conflicting conclusion.

Definition 3 (Defeated defaults: preliminary definition) Where $\mathcal{S}$ is a scenario based on the fixed priority default theory $\langle \mathcal{W}, \mathcal{D}, < \rangle$, the defaults from $\mathcal{D}$ that are *defeated* in $\mathcal{S}$ are those belonging to the set

$$\begin{aligned}\mathit{Defeated}_{\mathcal{W},\mathcal{D},<}(\mathcal{S}) = \{\delta \in \mathcal{D} :\ & \exists \delta' \in \mathit{Triggered}_{\mathcal{W},\mathcal{D},<}(\mathcal{S}) : \\ & (1)\ \delta < \delta', \\ & (2)\ \mathit{Conclusion}(\delta') \vdash \neg \mathit{Conclusion}(\delta)\}.\end{aligned}$$

This preliminary definition yields the correct results in the case of the Tweety Triangle: it follows from the definition that $\mathit{Defeated}_{\mathcal{W},\mathcal{D},<}(\mathcal{S}_0) = \{\delta_1\}$, since $\delta_2 \in \mathit{Triggered}_{\mathcal{W},\mathcal{D},<}(\mathcal{S}_0)$ and we have both (1) $\delta_1 < \delta_2$ and (2) $\mathit{Conclusion}(\delta_2) \vdash \neg \mathit{Conclusion}(\delta_1)$. Indeed, this preliminary definition yields correct results in all of the examples to be considered here, and we can safely rely on it as our official definition throughout this paper. In fact, however, the preliminary definition is not uniformly accurate, and leads to incorrect results in certain more complicated cases; those readers interested in the problems involved in formulating a proper definition can find a more extensive discussion in my (2007), as well as the papers cited there.

Once the concept of defeat is in place, we can define the set of defaults that are classified as binding in a particular scenario quite simply, as those that are triggered in that scenario, but neither conflicted nor defeated.

Definition 4 (Binding defaults) Where $\mathcal{S}$ is a scenario based on the fixed priority default theory $\langle \mathcal{W}, \mathcal{D}, < \rangle$, the defaults from $\mathcal{D}$ that are *binding* in $\mathcal{S}$ are those belonging to the set

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

Since the binding defaults are supposed to represent the good reasons, in the context of a particular scenario, it is natural to isolate the concept of a stable scenario as one containing all and only the defaults that are binding in that very context. Formally, where $\mathcal{S}$ is a scenario based on the default theory $\langle \mathcal{W}, \mathcal{D}, < \rangle$, we can say that $\mathcal{S}$ is a *stable scenario* just in case

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

An agent that has accepted a set of defaults that forms a stable scenario is in an enviable position. Such an agent has already accepted exactly those defaults that it recognizes as providing good reasons, in the context of the defaults it accepts; the agent, therefore, has no incentive either to abandon any of the defaults it has already accepted, or to accept any others.

2.3 *Reasoning with proper scenarios*

Our goal, we recall, is to characterize the proper scenarios—those sets of defaults that an ideally rational agent might come to accept based on the initial information contained in some default theory. Can we, then, simply identify the proper scenarios with the stable scenarios? Again, I offer two answers to this question, a preliminary answer and a final answer.

The preliminary answer is that, in the vast range of ordinary cases, including all of those to be considered in this paper, we can indeed identify the proper scenarios with the stable scenarios. This preliminary answer can be solidified into a preliminary definition.

Definition 5 (Proper scenarios: preliminary definition) Let $\mathcal{S}$ be a scenario based on the ordered default theory $\langle \mathcal{W}, \mathcal{D}, < \rangle$. Then $\mathcal{S}$ is a *proper scenario* based on $\langle \mathcal{W}, \mathcal{D}, < \rangle$ just in case $\mathcal{S} = \mathit{Binding}_{\mathcal{W},\mathcal{D},<}(\mathcal{S})$.

Unfortunately, however, the final answer is that there are also certain aberrant theories which allow stable scenarios that cannot really be classified as proper—that is, as scenarios that an ideal reasoner would accept. Since these aberrant cases do not concern us here, we will again rely on the preliminary definition as our official definition throughout this paper; those readers who are interested in a correct definition can turn to my (2007) for a discussion.

The concept of a proper scenario can be illustrated by returning to the Tweety Triangle. As the reader can verify, the proper scenario based on this default theory is $\mathcal{S}_1 = \{\delta_2\}$, where δ_2 is the default $P \to \neg F$, so that $\mathit{Conclusion}(\mathcal{S}_1) = \{\neg F\}$. In this case, then, the account proposed here associates with the default theory a unique proper scenario supporting the intuitively correct conclusion, that Tweety cannot fly. In other cases, however, this account—like many others in nonmonotonic reasoning—defines a relation between default theories and their proper scenarios that may seem anomalous from a more conventional logical perspective: certain default theories may be associated with multiple proper scenarios.[5]

The canonical example of a default theory with more than one proper scenario is the Nixon Diamond, which has two: both $\mathcal{S}_1 = \{\delta_1\}$ and $\mathcal{S}_2 = \{\delta_2\}$, where δ_1 is $Q \rightarrow P$ and δ_2 is $R \rightarrow \neg P$, so that $Conclusion(\mathcal{S}_1) = \{P\}$ and $Conclusion(\mathcal{S}_2) = \{\neg P\}$. In light of these two extensions, one of which contains P and the other $\neg P$, what is the agent supposed to conclude: is Nixon a pacifist or not? More generally, when an ordered default theory allows more than one proper scenario, how should we define its consequences?

The question is vexed, and has not been adequately addressed even in the literature on nonmonotonic reasoning. I do not have space to explore the matter in detail here, but will simply describe three options, in order to illustrate the range of possibilities.

One option is to interpret the different proper scenarios associated with a default theory simply as different equilibrium states that an ideal reasoner might arrive at on the basis of its initial information. The agent could then be expected to select, arbitrarily, a particular one of these scenarios and endorse the conclusions supported by it. In the case of the Nixon Diamond, for example, the agent could appropriately arrive either at the scenario $\mathcal{S}_1$ or at the scenario $\mathcal{S}_2$, appropriately endorsing either the conclusion that Nixon is a pacifist, or else the conclusion that he is not.

This option—now generally described as the *credulous*, or *choice*, option—is highly nonstandard from a theoretical perspective, but not, I think, incoherent.[6] It involves viewing the task of a default logic, not as guiding the reasoning agent to a unique set of appropriate conclusions, but as characterizing different, possibly conflicting conclusion sets as rational outcomes based on the initial information; default logic could then be seen as analogous to other fields, such as game theory, for example, that appeal to multiple equilibrium states in their characterization of rationality. And regardless of its theoretical pedigree, it seems clear that this credulous option is frequently employed in our everyday reasoning. Given conflicting defeasible rules, we often simply do adopt some internally coherent point of view in which these conflicts are resolved in some particular way, regardless of the fact that there are other coherent points of view in which the conflicts are resolved in different ways.

A second option is to suppose that each formula that is supported by some proper scenario must be given some weight, at least. We might, for example, take $\mathcal{B}(A)$ to mean that there is good reason to believe the statement A; and we might suppose that a default theory provides good reason to believe a statement whenever that statement is included in some extension of the theory, some internally coherent point of view. In the case of the Nixon Diamond, the agent could then be expected to endorse both $\mathcal{B}(P)$ and $\mathcal{B}(\neg P)$—since each of P and $\neg P$ is supported by some proper scenario—thus concluding that there is good reason to believe that Nixon is a pacifist, and also good reason to believe that he is not.

This general approach is particularly attractive when defaults are provided with a practical, rather than an epistemic, interpretation, so that the default $A \rightarrow B$ is taken to mean that A provides a reason for performing the action indicated by B. In that case, the modal operator wrapped around the conclusions supported by the various proper scenarios associated with a default theory could naturally be read as the deontic operator $\bigcirc$, representing what the agent ought to do. And when different proper scenarios support conflicting conclusions, say A and $\neg A$, we could then expect the reasoning agent to endorse both $\bigcirc(A)$ and $\bigcirc(\neg A)$, thereby facing a normative, but not a logical, conflict. This approach, as it turns out, leads to an attractive deontic logic.[7]

[5] And others may be associated with no proper scenarios at all, a matter that need not concern us here.

[6] This reasoning strategy was first labelled as "credulous" by Touretzky et al. (1987), and as the "choice" option by Makinson (1994); it had earlier been characterized as "brave" by McDermott (1982).

[7] The resulting logic generalizes that of van Fraassen (1973). The interpretation of van Fraassen's account within default logic was first established in my (1994a); a defense of the overall approach can be found in my (2003).

A third option is to suppose that the agent should endorse a conclusion just in case it is supported by every proper scenario based on the original default theory; in the Nixon Diamond, for example, the agent would then conclude neither that Nixon is a pacifist nor that he is not, since neither P nor $\neg P$ is supported by both proper scenarios. This option is now generally described as *skeptical*.[8] It is by far the most popular option, and is sometimes considered to be the only coherent form of reasoning in the presence of multiple proper scenarios, though I have recently argued that the issue is more complex.[9]

3. Elaborating the theory

The central thesis of this paper is that reasons can usefully be thought of as provided by defaults, but so far I have done little more than provide an account of default reasoning. I now want to support my central thesis by showing how this account can be elaborated to deal with two issues involved in developing a more robust theory of reasons. First, the priorities among defaults have, so far, been taken as fixed in advance, but there are cases in which these priorities must themselves be established through defeasible reasoning. And second, the notion of defeat defined here captures only one form, generally called "rebutting" defeat, in which a stronger default defeats a weaker default by contradicting its conclusion. There is at least one other form, generally called "undercutting" defeat—and related to the discussion of "exclusionary" reasons from the literature on practical reasoning—in which one default defeats another, not by contradicting its conclusion, but by undermining its capacity to provide a reason.

3.1 Variable priority default theories

The definition

We have concentrated on fixed priority default theories, in which priority relations among default rules are fixed in advance; but in fact, some of the most important things we reason about, and reason about defeasibly, are the priorities among the very defaults that guide our defeasible reasoning. This is particularly true in well-structured normative domains, such as the law, where the resolution of a dispute often involves an explicit decision concerning the priority relations among different rules bearing on some issue.

Our first task, then, is to show how this kind of reasoning can be accommodated within the general framework presented here. What we want is an account in which, just as before, our reasoning is guided by a set of defaults subject to a priority ordering, but in which it is now possible for the priorities among defaults to be established through the same process of reasoning they serve to guide. Although this may sound complicated—perhaps forbiddingly so, perhaps circular—it turns out that the present theory can be extended to provide such an account in four simple steps, through the adaptation of known techniques.[10]

The first step is to enrich our object 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 d_X, for some subscript X, and that each such constant refers to the default δ_X. And we will assume also that the object language now contains the relation symbol $\prec$, representing priority among defaults.

To illustrate this notation, suppose that δ_1 is the default $A \to B$, that δ_2 is the default $C \to \neg B$, and that δ_3 is the default $\top \to d_1 \prec d_2$,

[8] The label is again due to Touretzky et al. (1987); the same reasoning strategy had earlier been described as "cautions" by McDermott (1982).

[9] An argument that the skeptical approach, as defined here, presents the only coherent option for epistemic default reasoning is presented by Pollock (1995, pp. 62–63); some of my doubts can be found in my (2002).

[10] The basic idea underlying the techniques to be described here was first introduced by Gordon (1993); the idea was then refined and developed by a number of people, most notably Brewka (1994, 1996) as well as Prakken and Sartor (1995, 1996).

where, in keeping with our convention, d_1 and d_2 refer to the defaults δ_1 and δ_2. Then what δ_3 says is that, by default, δ_2 has a higher priority than δ_1. As a result, we would expect that, when both of these defaults are triggered—that is, when both A and C hold—the default δ_1 will generally be defeated by δ_2, since the two defaults have conflicting conclusions. Of course, since δ_3 is itself a default, the information it provides concerning the priority between δ_1 and δ_2 is defeasible as well, and can likewise be overridden.

The second step is to shift our attention from structures of the form $\langle \mathcal{W}, \mathcal{D}, < \rangle$—that is, from fixed priority default theories—to structures of the form $\langle \mathcal{W}, \mathcal{D} \rangle$, containing a set $\mathcal{W}$ of ordinary formulas 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 defeasible reasoning. Because conclusions about the priorities among defaults might themselves vary depending on which defaults the agent accepts, these new structures are known as *variable priority default theories*. We stipulate as part of this definition that the set $\mathcal{W}$ of ordinary formulas from a variable priority default theory must contain each possible instance of the irreflexivity and transitivity schemata

$$\neg(d \prec d),$$
$$(d \prec d' \wedge d' \prec d'') \supset d \prec d'',$$

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

Now suppose the agent accepts some scenario containing these new priority statements; the third step, then, is to lift the priority ordering that is implicit in the agent's scenario to an explicit ordering that can be used in our metalinguistic reasoning. This is done in the simplest possible way. If $\mathcal{S}$ is some scenario based on the default theory $\langle \mathcal{W}, \mathcal{D} \rangle$, we now take the statement $\delta <_{\mathcal{S}} \delta'$ to mean that the default δ' *has a higher priority than* δ *according to the scenario* $\mathcal{S}$, where this notion is defined as follows:

$$\delta <_{\mathcal{S}} \delta' \text{ if and only if } \mathcal{W} \cup \mathit{Conclusion}(\mathcal{S}) \vdash d \prec d'.$$

What this means is that: δ' has a higher priority than δ according to the scenario $\mathcal{S}$ just in case the conclusions of the defaults belonging to this scenario, when taken together with the ordinary information from the background theory, entail the formula $d \prec d'$, telling us that δ' has a higher priority than δ. Because the ordinary information from $\mathcal{W}$ 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 variable priority default theories. This is accomplished by leveraging our previous definition, which sets out the conditions under which $\mathcal{S}$ counts as a proper scenario for a fixed priority theory $\langle \mathcal{W}, \mathcal{D}, < \rangle$, where $<$ can be any strict partial ordering over the defaults. Using this previous definition, we can now stipulate that $\mathcal{S}$ is a proper scenario for the variable priority theory $\langle \mathcal{W}, \mathcal{D} \rangle$ just in case $\mathcal{S}$ is a proper scenario for the particular fixed priority theory $\langle \mathcal{W}, \mathcal{D}, <_{\mathcal{S}} \rangle$, where $\mathcal{W}$ and $\mathcal{D}$ are carried over from the variable priority theory, and $<_{\mathcal{S}}$ is the priority relation derived from the scenario $\mathcal{S}$ itself.

Definition 6 (Proper scenarios: variable priority default theories) Let $\langle \mathcal{W}, \mathcal{D} \rangle$ be a variable priority default theory and $\mathcal{S}$ a scenario. Then $\mathcal{S}$ is a *proper scenario* based on $\langle \mathcal{W}, \mathcal{D} \rangle$ if and only if $\mathcal{S}$ is a proper scenario based on the fixed priority default theory $\langle \mathcal{W}, \mathcal{D}, <_{\mathcal{S}} \rangle$.

The intuitive picture is this. In searching for a proper scenario, the agent arrives at some scenario $\mathcal{S}$, 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 original scenario $\mathcal{S}$, then the scenario is proper.

Some examples

The approach described here can be illustrated through a variant of the Nixon Diamond, 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. Suppose, then, that Nixon is confronted with the default theory $\langle \mathcal{W}, \mathcal{D} \rangle$, with $\mathcal{W}$ containing the formulas Q and R, reminding Nixon that he is both a Quaker and a Republican, and with $\mathcal{D}$ containing only δ_1 and δ_2, where δ_1 is the default $Q \to P$ and δ_2 is $R \to \neg P$. Given our current perspective, these two defaults should now be interpreted as providing practical reasons: δ_1 tells Nixon that, as a Quaker, he ought to become a pacifist, while δ_2 tells him that, as a Republican, he ought not to become a pacifist. Nothing in his initial theory tells Nixon how to resolve the conflict between these two defaults, and so he is faced with a practical dilemma: his initial theory yields two proper scenarios, the familiar $\mathcal{S}_1 = \{\delta_1\}$ and $\mathcal{S}_2 = \{\delta_2\}$, supporting the conflicting conclusions P and $\neg P$.

Now imagine that Nixon decides to consult with certain authorities to help him resolve his dilemma. Let us suppose that he discusses the problem first with a respected member of his Friends Meeting House, who tells him that δ_1 should take priority over δ_2, but that he also talks with a Republican party official, who tells him just the opposite. The advice of these two authorities can be encoded by supplementing the set $\mathcal{D}$ with the new defaults δ_3 and δ_4, where δ_3 is $\top \to \ d_2 \prec d_1$ and δ_4 is $\top \to \ d_1 \prec d_2$. Given our practical perspective, these two defaults should be interpreted, not as evidence, but as advice; the default δ_3, for example, should be interpreted, not as providing Nixon with evidence that δ_1 actually *has* more weight than δ_2, but as suggesting that he should *place* more weight on δ_1 in his deliberations.[11]

Since his chosen authorities disagree, Nixon has not yet resolved his practical dilemma, now represented by the two proper scenarios $\mathcal{S}_3 = \{\delta_1, \delta_3\}$ and $\mathcal{S}_4 = \{\delta_2, \delta_4\}$, which again favor conflicting courses of action. According to the scenario $\mathcal{S}_1$, supporting the statements $d_2 \prec d_1$ and P, Nixon should place more weight on δ_1 than on the conflicting δ_2, and so become a pacifist; according to the scenario $\mathcal{S}_2$, supporting the statements $d_1 \prec d_2$ and $\neg P$, Nixon should instead place more weight on δ_2 and not become a pacifist. What is especially interesting about this theory, however, is not that it yields two proper scenarios, favoring two courses of action, but that it yields *only* two proper scenarios, favoring only two courses of action. After all, the default δ_1 conflicts with δ_2, while the default δ_3 conflicts with δ_4. Since there are two conflicts, each of which can go either way, why are there not four proper scenarios, favoring four courses of action?

The answer is that the two conflicts are not independent. Any resolution of the conflict between δ_3 and δ_4 commits Nixon to a particular priority ordering between δ_1 and δ_2, which then determines the resolution of that conflict. From an intuitive standpoint, it would be incorrect for Nixon to accept δ_3, for example, according to which more weight is to be given to δ_1 than to δ_2, but then to accept δ_2 anyway, and choose not to become a pacifist. This intuition is captured formally because $\mathcal{S}_5 = \{\delta_2, \delta_3\}$—the scenario containing the combination of δ_3 and δ_2—leads to a derived priority ordering according to which $\delta_2 <_{\mathcal{S}_5} \delta_1$. If we supplement our current variable priority theory with this derived priority ordering to get the fixed priority theory $\langle \mathcal{W}, \mathcal{D}, <_{\mathcal{S}_5} \rangle$, we can now see that the default δ_2 is defeated in the context of $\mathcal{S}_5$ by the default δ_1, a stronger default with a conflicting conclusion. The scenario $\mathcal{S}_5$ is not, therefore, a proper scenario based on the fixed priority theory $\langle \mathcal{W}, \mathcal{D}, <_{\mathcal{S}_5} \rangle$, and so it cannot, according to our definition, be a proper scenario based on our original variable priority theory either.

Finally, let us imagine that Nixon, still faced with the conflict, continues to seek further counsel. Perhaps he now goes to his father, who tells him that the church elder's advice is to be preferred to that of the party official; this information can be represented by adding the rule δ_5 to the set $\mathcal{D}$ of defaults, where δ_5 is $\top \to d_4 \prec d_3$. With this new de-

[11] The idea that our reasoning itself determines what reasons we ought to place more weight on is discussed and defended in Schroeder (2007).

fault, Nixon has at last resolved his conflict. As the reader can verify, the theory now yields the single proper scenario $\mathcal{S}_3 = \{\delta_1, \delta_3, \delta_5\}$—supporting the conclusions $d_4 \prec d_3$, $d_2 \prec d_1$, and P—according to which Nixon should favor δ_3 over δ_4, and so favor δ_1 over δ_2, and so become pacifist.

This modification of the Nixon Diamond should, I hope, serve to illustrate the workings of variable priority default theories, but perhaps not their usefulness, due to the whimsical nature of the example. For a more realistic example, we consider a situation from commercial law, originally described by Thomas Gordon (1993), but simplified and adapted for present purposes.[12]

We are to imagine that both Smith and Jones have individually lent money to Miller for the purchase of an oil tanker, which serves as collateral for both loans, so that both lenders have a security interest in the ship—a right to recoup the loan value from the sale of the ship in case of default. Miller has, we imagine, defaulted on both the loans; the ship will be sold, and the practical question is which of the two lenders has an initial claim on the proceeds. The specific legal issue that arises is whether Smith's security interest in the ship has been *perfected*—roughly, whether it can be protected against security interests that might be held by others, such as that of Jones.

As it happens, there are two relevant bodies of regulation governing the situation: the Uniform Commercial Code (UCC), according to which a security interest can be perfected by taking possession of the collateral, and the Ship Mortgage Act (SMA), according to which a security interest in a ship can be perfected only by filing certain financial documents. In this case, we are to imagine that Smith is in possession of the ship but has failed to file the necessary documents, so that the two statutes yield conflicting results: according to UCC, Smith's security interest in the ship is perfected, but according to SMA, it is not.

There are, of course, various legal principles for resolving conflicts of this kind. One is the principle of *Lex Posterior*, which gives precedence to the more recent of two regulations. Another is the principle of *Lex Superior*, which gives precedence to the regulation supported by the higher authority. Here, UCC supplies the more recent of the two regulations, having been drafted and then enacted by all the various states (except Louisiana) in the period between 1940 and 1964, while SMA dates from 1920. However, SMA derives from a higher authority, since it is federal law, rather than state law. Given only this information, then, the conflict remains: according to *Lex Posterior*, UCC should take precedence over SMA, while according to *Lex Superior*, SMA should take precedence over UCC.

But let us suppose that, for whatever reason—custom, legislation, a court decision—one of these two principles for conflict resolution has gained favor over the other: perhaps *Lex Posterior* is now favored over *Lex Superior*. In that case, the current situation is analogous in structure to the previous Nixon example, and can be represented in the same way.

To aid comprehension, we use mnemonic symbols in our formalization. Let *Perfected*, *Possession*, and *Documents* represent the respective propositions that Smith's security interest in the ship is perfected, that Smith possesses the ship, and that Smith has filed the appropriate financial documents. Then the relevant portions of UCC and SMA can be represented as the defaults δ_{UCC} and δ_{SMA}, where δ_{UCC} is $\mathit{Possession} \rightarrow \mathit{Perfected}$ and δ_{SMA} is $\neg \mathit{Documents} \rightarrow \neg \mathit{Perfected}$. The principles of *Lex Posterior* and *Lex Superior* can be captured by the defeasible generalizations

$$\begin{array}{l} \mathit{Later}(d, d') \rightarrow\ d < d' \\ \mathit{Federal}(d) \wedge \mathit{State}(d') \rightarrow\ d' < d, \end{array}$$

telling us, quite generally, that later regulations are to be preferred

[12] Other realistic examples are developed by Prakken and Satror (1996), who consider the issues surrounding a conflict between European Community and Italian law concerning the marketing of a particular product under the label of "pasta," and also a conflict between separate Italian laws concerning the renovation of historic buildings.

over earlier regulations, and that federal regulations are to be preferred over those issued by states; the particular instances of these two principles of concern to us here can be represented as δ_{LP} and δ_{LS}, where δ_{LP} is $\mathit{Later}(d_{SMA}, d_{UCC}) \rightarrow d_{SMA} < d_{UCC}$ and δ_{LS} is $\mathit{Federal}(d_{SMA}) \wedge \mathit{State}(d_{UCC}) \rightarrow d_{UCC} < d_{SMA}$. Finally, we can take δ_{LSLP} as the default $\top \rightarrow d_{LS} < d_{LP}$, again an instance of a general principle telling us that *Lex Posterior* is to be favored over *Lex Superior*.

Now let $\langle \mathcal{W}, \mathcal{D} \rangle$ be the variable priority default theory in which $\mathcal{D}$ contains these five defaults—δ_{UCC}, δ_{SMA}, δ_{LP}, δ_{LS}, and δ_{LSLP}—and in which $\mathcal{W}$ contains the facts of the situation—*Possession*, $\neg$*Documents*, $\mathit{Later}(d_{SMA}, d_{UCC})$, $\mathit{Federal}(d_{SMA})$, and $\mathit{State}(d_{UCC})$—telling us, again, that Smith has possession of the ship but did not file documents, that UCC is later than SMA, and that SMA is federal law while UCC is state law. This default theory then yields the set $\mathcal{S} = \{\delta_{UCC}, \delta_{LP}, \delta_{LSLP}\}$ as its unique proper scenario—supporting the conclusions $d_{LS} < d_{LP}$, $d_{SMA} < d_{UCC}$, and *Perfected*—and so recommending a course of action according to which δ_{LP} is to be favored over δ_{LS}, so that δ_{UCC} is then favored over δ_{SMA}, and we should therefore judge that Smith's security interest in the oil tanker is perfected.

3.2 *Threshold default theories*

The definition

We have considered, thus far, only one form of defeat—generally called "rebutting" defeat—according to which a default supporting a conclusion is said to be defeated by a stronger default supporting a conflicting conclusion. There is also a second form of defeat, according to which one default supporting a conclusion is thought to be defeated by another, not because it supports a conflicting conclusion, but because it challenges the connection between the premise and the conclusion of the original default. In the literature on epistemic reasons, this second form of defeat is generally referred to as "undercutting" defeat, and was first pointed out, I believe, by John Pollock in (1970).[13]

The distinction between these two forms of defeat can be illustrated by a standard example. Suppose an object in front of me looks red. Then it is reasonable for me to conclude that it is red, through an application of a general default according to which things that look red tend to be red. But let us imagine two drugs. The effect of Drug #1 is to make red things look blue and blue things look red; the effect of Drug #2, by contrast, is to make everything look red. Now, if the object looks red but I have taken Drug #1, then it is natural to appeal to another default, stronger than the original, according to which things that look red once I have taken Drug #1 tend to be blue, and so not red. This new default would then defeat the original in the sense we have considered so far, by providing a stronger reason for a conflicting conclusion. If the object looks red but I have taken Drug #2, on the other hand, then it seems again that I am no longer entitled to the conclusion that the object is red. But in this case, the original default is not defeated in the same way. There is no stronger reason for concluding that the object is not red; instead, it is as if the original default is itself undercut, and no longer provides any reason for its conclusion.

This second form of defeat, or something very close to it, is discussed also in the literature on practical reasoning, where it is considered as part of the general topic of "exclusionary" reasons, first introduced by Joseph Raz in (1975). Raz provides a number of examples to motivate the concept, but we consider here only the representative case of Colin, who must decide whether to send his son to a private school. We are to imagine that there are various reasons pro and con. On one hand, the school will provide an excellent education for Colin's son, as well as an opportunity to meet a more varied group of friends; on the other hand, the tuition is high, and Colin is concerned that a deci-

[13] Some of the early formalisms for knowledge representation in artificial intelligence allowed for a form of undercutting defeat, such as the NETL system described in Fahlman (1979); but the idea quickly evaporated in the artificial intelligence literature, and did not appear in this field again until it was reintroduced by writers explicitly reflecting on Pollock's work.

sion to send his own son to a private school might serve to undermine support for public education more generally.

However, Raz asks us to imagine also that, in addition to these ordinary reasons pro and con, Colin has promised his wife that, in all decisions regarding the education of his son, he will consider only those reasons that bear directly on his son's interests. And this promise, Raz believes, cannot properly be viewed as just another one of the ordinary reasons for sending his son to the private school, like the fact that the school provides a good education. It must be viewed, instead, as a reason of an entirely different sort—a "second-order" reason for excluding from consideration all those ordinary, or "first-order," reasons that do not bear on the interests of Colin's son. Just as, once I have taken Drug #2, I should disregard the default according to which things that look red tend to be red, Colin's promise should lead him, likewise, to disregard those defaults that do not bear on the interests of his son. An exclusionary reason, on this interpretation, is nothing but an undercutting defeater in the practical domain.

Now, how can this phenomenon of undercutting, or exclusionary, defeat be accounted for? The standard practice is to postulate undercutting defeat as a separate, and primitive, form of defeat, to be analyzed alongside the concept of rebutting defeat; this practice is followed, most notably, by Pollock.[14] What I would like to suggest, however, is that, once we have introduced the ability to reason about priorities among defaults, the phenomenon of undercutting defeat can then be analyzed, more naturally, simply as a special case of priority adjustment. The basic idea is straightforward. In our priority ordering, we posit some particular value—say, τ, for the *threshold* value—low enough that we feel safe in considering only those defaults whose priority lies above this threshold. A default is then undercut when our reasoning forces us to conclude that its priority falls below threshold.[15]

In order to implement this idea, we revise our earlier definition of triggering to include the additional requirement that a default cannot be triggered unless it lies above the threshold value—that is, we will now require that a default δ cannot belong to $\mathit{Triggered}_{\mathcal{W},\mathcal{D},<}(\mathcal{S})$ unless $\mathcal{W} \cup \mathit{Conclusion}(\mathcal{S}) \vdash \mathit{Premise}(\delta)$ and, in addition, $\tau < \delta$. But of course, this single revision is not enough. With this revision alone, no defaults at all would now be triggered, since it has not yet been established that any defaults actually lie above threshold. In order to guarantee that the right defaults, and only the right defaults, lie above the threshold value, we must supplement our variable priority default theories with some additional information, and then modify our definition of a triggered default even further.

We begin by introducing the concept of a *threshold default theory* as a variable priority default theory $\langle \mathcal{W}, \mathcal{D} \rangle$ in which the set $\mathcal{D}$ of defaults and the set $\mathcal{W}$ of background facts are subject to two further constraints. In formulating these constraints, we refer to defaults of the sort considered thus far as *ordinary defaults*, and we take t as the linguistic constant corresponding to τ, the threshold weight.

The first constraint on threshold default theories is that, for each ordinary default δ_X belonging to $\mathcal{D}$, there is also a special *threshold default* δ^*_X, of the form $\top \to t \prec d_X$. The role of the threshold default δ^*_X is to tell us that, by default, the priority value assigned to the ordinary default δ_X lies above threshold. Just as each ordinary default δ_X is represented by a term d_X from the object language, we suppose that each threshold default δ^*_X is represented by a term d^*_X; and we let $\mathcal{D}^*$ represent the entire set of threshold defaults from $\mathcal{D}$. The second constraint

[14] See Pollock (1995) and the papers cited there; a survey of the work on this topic in nonmonotonic reasoning, which largely follows the same approach, can be found in Prakken and Vreeswijk in (2002).

[15] The general idea developed here conforms to a proposal put forth by Dancy (2004), who introduces the concepts of "intensifiers" and "attenuators" as considerations that strengthen or weaken the force of reasons, and who then thinks of a "disabler" as a consideration that attenuates the strength of a reason more or less completely—or in our current vocabulary, one that attenuates the default providing the reason so thoroughly that it falls below threshold. The idea that undercutting comes in degrees, and that what is typically referred to as "undercutting" is best analyzed as an extreme case of attenuation in the strength of reasons, is likewise noted by Schroeder (2005), who refers to this idea as the "undercutting hypothesis."

is that the set $\mathcal{W}$ of ordinary information must contain, in addition to the formulas mentioned earlier, guaranteeing a strict partial ordering, each instance of the schema $d^* \prec d$, in which d^* ranges over arbitrary threshold defaults and d ranges over ordinary defaults. The point of these formulas is to guarantee that threshold defaults are uniformly lower in priority than ordinary defaults.

The first of these two constraints can be provided with a sort of pragmatic justification. The ordinary defaults belonging to $\mathcal{D}$ are there to guide our reasoning. Since our plan is to modify the definition of triggering so that only defaults lying above threshold can be triggered, there would be little point in including an ordinary default within $\mathcal{D}$ unless we could assume, at least by default, that it lies above threshold. On the other hand, we cannot simply postulate, as a hard fact, that the ordinary defaults lie above threshold, since we need to allow for the possibility that they might be undercut—that our reasoning might lead us to conclude that some particular default should fall below threshold. We want to be able to assume that the ordinary defaults lie above threshold, then, but we cannot require that they do; therefore, we rely on threshold defaults to place ordinary defaults above threshold by default.

There is, however, a complication. Only defaults that are triggered can affect our reasoning. If we are to rely on threshold defaults to place the ordinary defaults above threshold, so that these ordinary defaults can be triggered, we must first guarantee that the threshold defaults themselves are triggered. This cannot be done, of course, by appealing to a further set of defaults whose role is to place the threshold defaults above threshold, since that idea leads to a regress; we would then have to guarantee that the defaults belonging to this further set lie above threshold, and so on. Instead, we halt the regress at the first step simply by *stipulating* that each threshold default is triggered. According to our revised treatment, then, the defaults from $\mathcal{D}$ that are triggered in the context of a scenario $\mathcal{S}$ are defined as including the set entire $\mathcal{D}^*$ of threshold defaults as well as those ordinary defaults lying above threshold whose premises follow from the information contained in that scenario.

Definition 7 (Triggered defaults: revised definition) Where $\mathcal{S}$ is a scenario based on the fixed priority default theory $\langle \mathcal{W}, \mathcal{D}, < \rangle$, the defaults from $\mathcal{D}$ that are *triggered* in $\mathcal{S}$ are those belonging to the set

$$\begin{aligned} &\mathit{Triggered}_{\mathcal{W},\mathcal{D},<}(\mathcal{S}) = \\ &\qquad \mathcal{D}^* \cup \{\delta \in \mathcal{D} : \tau < \delta \;\&\; \mathcal{W} \cup \mathit{Conclusion}(\mathcal{S}) \vdash \mathit{Premise}(\delta)\}. \end{aligned}$$

Since each threshold default has as its premise the trivial statement $\top$, which follows from the information contained in any scenario whatsoever, the net effect of our stipulation is that, for threshold defaults, the requirement that a triggered default must lie above threshold is suspended. What this means is that a threshold default can never be undercut—we can never conclude that such a default cannot be triggered because it falls below threshold. Since threshold defaults provide reasons for placing ordinary defaults above threshold, it follows that there will always be some reason for concluding that any ordinary default should lie above threshold. This result is, in a sense, simply a restatement of our pragmatic idea that there is no point even in registering a default unless we have reason to believe that it lies above threshold.

The second constraint governing threshold default theories—that the set $\mathcal{W}$ of ordinary formulas must contain each instance of the schema $d^* \prec d$—can now be understood against the background of this pragmatic idea. We do not want threshold defaults to be undercut; there should always be some reason for taking any ordinary default seriously, placing it above threshold. However, we do want to allow for the possibility that threshold defaults might be defeated by stronger reasons for placing ordinary defaults below threshold, so that these ordinary defaults can themselves be undercut. We therefore stipulate that each ordinary default has a higher priority than any threshold default.

Some examples

To illustrate these ideas, we begin by describing a particular threshold default theory $\langle \mathcal{W}, \mathcal{D} \rangle$ representing the epistemic example sketched earlier, involving the two drugs. Let L, R, $D1$, and $D2$ stand for the respective propositions that the object before me looks red, that it is red, that I have taken Drug #1, and that I have taken Drug #2. We can then take δ_1, δ_2, δ_3, and δ_4 as the ordinary defaults belonging to $\mathcal{D}$, where δ_1 is $L \to R$, δ_2 is $L \wedge D1 \to \neg R$, δ_3 is $\top \to d_1 \prec d_2$, and δ_4 is $D2 \to d_1 \prec t$. According to the first of these defaults, it is reasonable to conclude that the object is red if it looks red; according to the second, that the object is not red if it looks red once I have taken Drug #1; according to the third, that the second default is stronger than the first; and according to the fourth, that the strength of the first default falls below threshold if I have taken Drug #2.

Since $\langle \mathcal{W}, \mathcal{D} \rangle$ is a threshold default theory, it is subject to our two additional constraints. By the first constraint, the set $\mathcal{D}$ must also contain four threshold defaults, one corresponding to each of the ordinary defaults—that is, a default δ_i^* of the form $\top \to t \prec d_i$, for each i from 1 through 4. And by the second constraint, the set W of ordinary formulas must contain sixteen statements of the form $d_i^* \prec d_j$, where i and j range independently from 1 through 4.

Now let us first suppose that $\mathcal{W}$ contains, in addition, the formulas L and $D1$, representing the situation in which the object looks red but I have taken Drug #1. In this case, the threshold default theory yields $\mathcal{S}_1 = \{\delta_1^*, \delta_2^*, \delta_3^*, \delta_4^*, \delta_2, \delta_3\}$ as its unique proper scenario, supporting the four statements of the form $t \prec d_i$, for i from 1 through 4, as well as $\neg R$ and $d_1 \prec d_2$. The scenario $\mathcal{S}_1$ thus allows us to conclude that each of the ordinary defaults lies above threshold, that δ_2 has a higher priority than δ_1, and that the object is not red. The default δ_1 is defeated in the scenario, since its conclusion conflicts with that of δ_2, whose greater strength is established by δ_3. The default δ_4 is not triggered, since its premise is not entailed in the context of this scenario.

Next, let us suppose instead that $\mathcal{W}$ contains the formulas L and $D2$, representing the situation in which the object looks red but I have taken Drug #2. The theory now yields $\mathcal{S}_2 = \{\delta_2^*, \delta_3^*, \delta_4^*, \delta_3, \delta_4\}$ as its unique proper scenario, supporting, in this case, only three statements of the form $t \prec d_i$, for i from 2 through 4, along with $d_1 \prec d_2$, and now $d_1 \prec t$ as well. The scenario $\mathcal{S}_2$ allows us to conclude, then, that the three ordinary defaults δ_2, δ_3, and δ_4 lie above threshold, that δ_2 has a higher priority than δ_1, and that δ_1 in fact falls below threshold; no conclusions can be reached about the actual color of the object. Here, the threshold default δ_1^*, which would otherwise have placed δ_1 above threshold, is defeated by δ_4, a stronger default—its greater strength is established by the formula $d_1^* \prec d_4$ from $\mathcal{W}$—supporting the conflicting conclusion that δ_1 should fall below threshold. Neither δ_1 nor δ_2 is triggered, δ_2 because its premise is not entailed in the context of the scenario, and δ_1 because it fails to satisfy our new requirement that triggered defaults must lie above threshold.

It is useful to consider this situation from the standpoint of our earlier analysis of a reason as the premise of a triggered default. Once I have taken Drug #2, so that δ_1 falls below threshold, then according to our analysis, this default no longer provides any reason for concluding that the object is red. It is not as if δ_1 is defeated, in our standard sense of rebutting defeat. There is no stronger triggered default supporting the contrary conclusion—no reason at all, in fact, to conclude that the object is not red. Instead, we are forced to conclude that δ_1 must lie below threshold, so that it cannot itself be triggered, and therefore, provides no reason of its own.

It is possible, of course, for situations to be considerably more complicated than this: ordinary defeaters and undercutters can themselves be defeated or undercut, both defeaters and undercutters of defeaters and undercutters can likewise be defeated or undercut, and so on. We cannot explore the ramifications among these possibilities in any detail here, but it is worth considering a situation that is just one degree more complex. Suppose that, as before, the object looks red and I have taken Drug #2, which makes everything look red, but that I have also taken Drug #3, an antidote to Drug #2 that neutralizes its effects. How should this situation be represented?

From an intuitive standpoint, what Drug #3 gives me is, not any positive reason for concluding that the object is red, but instead, a reason for disregarding the reason provided by Drug #2 for disregarding the reason provided by my senses for concluding that the object is red. Its associated default therefore provides a reason for disregarding a reason for disregarding a reason—an undercutter undercutter. Formally, then, letting $D3$ stand for the proposition that I have taken Drug #3, the effect can be captured through δ_5, the default $D3 \rightarrow d_4 \prec t$, according to which, once I have taken this new drug, the previous default δ_4 should fall below threshold. Suppose, now, that the ordinary default δ_5 is added to $\mathcal{D}$. By our threshold constraints, $\mathcal{D}$ must contain δ_5^*, the corresponding threshold default $\top \rightarrow t \prec d_5^*$, and $\mathcal{W}$ must contain the formulas $d_i^* \prec d_j$ for i and j from 1 through 5. If $\mathcal{W}$ also contains L, $D2$, and $D3$—representing the situation in which the object looks red, and I have taken both Drugs #2 and #3—then the unique proper scenario for the corresponding threshold default theory is $\mathcal{S}_3 = \{\delta_1^*, \delta_2^*, \delta_3^*, \delta_5^*, \delta_1, \delta_3, \delta_5\}$, supporting statements of the form $t \prec d_i$ where i is 1, 2, 3, or 5, along with the statements R, $d_1 \prec d_2$, and $d_4 \prec t$. The scenario $\mathcal{S}_3$ therefore allows us to conclude that all the ordinary defaults except δ_4 lie above threshold, that δ_2 has a higher priority than δ_1, that δ_4 lies below threshold, and that the object is red. By forcing δ_4 below threshold, the new δ_5, undercuts any reason for disregarding δ_1, which can now emerge from below threshold to support the conclusion that the object is red.

Turning to the practical domain, we can illustrate the use of undercutting defeat—or exclusionary reasons—by reconsidering the case of Colin, who is deliberating about sending his son to a private school. Let S, E, and H represent the propositions that Colin's son is sent to the school, that the school provides an excellent education, but that its tuition is high; and take δ_1 as the default $E \rightarrow S$ and δ_2 as $H \rightarrow \neg S$. These two defaults should be interpreted as telling Colin that the excellent education favors sending his son to the private school, while the high tuition favors not doing so. Simplifying somewhat, let us suppose that these are the only two reasons bearing directly on the issue. But there is also Colin's promise to his wife, which we can represent through the generalization

$$\neg \mathit{Welfare}(d) \rightarrow d \prec t,$$

telling Colin that, in this decision, he should disregard any considerations that do not center around his son's welfare; and suppose δ_3 is the default $\neg \mathit{Welfare}(d_2) \rightarrow d_2 \prec t$, the particular instance of this general default that is of concern here.

Let $\langle \mathcal{W}, \mathcal{D} \rangle$ be the threshold default theory in which $\mathcal{D}$ contains δ_1, δ_2, and δ_3 as its ordinary defaults, while $\mathcal{W}$ contains E, H, and $\neg \mathit{Welfare}(d_2)$, according to which: the education is excellent, the tuition is high, but δ_2, the consideration provided by the high tuition, does not concern the welfare of Colin's son. Of course, since this is a threshold default theory, we must suppose also that $\mathcal{D}$ contains the threshold defaults associated with each of the ordinary defaults—a default δ_i^* of the form $\top \rightarrow t \prec d_i$, for i from 1 through 3. And we must suppose that $\mathcal{W}$ contains statements of the form $d_i^* \prec d_j$ for i and j from 1 through 3.

Now if Colin were to consider only the defaults δ_1 and δ_2 in this situation, it is easy to see that he would be faced with a conflict—incomparable reasons recommending different actions. Because of his promise to his wife, however, Colin's deliberation is also constrained by the default δ_3, an exclusionary reason requiring him to remove δ_2 from consideration, placing it below threshold, and so allowing δ_1 to stand unopposed. The theory thus yields $\mathcal{S}_1 = \{\delta_1^*, \delta_3^*, \delta_1, \delta_3\}$ as its unique proper scenario, supporting the statements $t \prec d_1$, $t \prec d_3$, S, and $d_2 \prec t$. According to $\mathcal{S}_1$, what Colin should conclude is that both δ_1 and δ_3 lie above threshold, that δ_2 does not, and that he should send his son to the private school; since δ_2 falls below threshold, it does not provide any reason to the contrary.[16]

[16] Just as in the epistemic case, where undercutters can be undercut, exclu-

4. Applying the theory

In this final section, I now want to apply the account of reasons developed in this paper to an argument recently advanced by Jonathan Dancy in support of particularism in moral theory. We begin with some general definitions.

It is often thought that our ability to settle on appropriate actions, or at least to justify these actions as appropriate, involves, at some level, an appeal to general principles. Let us refer to any view along these lines as a form of *generalism*. Certainly the view presented here qualifies, since it is based, ultimately, on a system of principles intended to capture defeasible generalizations.

Standing in contrast to generalism is the position known as *particularism*, which tends to downplay the importance of general principles, and to emphasize instead a kind of receptivity to the features of particular situations. It is useful, however, to distinguish different versions of this particularist perspective. We can imagine, first, a *moderate* particularism, which holds only that a significant part of our practical evaluation, at least, is not grounded in an appeal to general principles.

It is hard to argue with moderate particularism, phrased in this way. It frequently happens, for example, that the application of a set of principles in some particular situation yields what appears to be the wrong result from an intuitive standpoint, and we therefore feel that these principles must themselves be revised. But how can we conclude, in a case like this, that the original result was wrong? It cannot be through an application of our principles, since it is these principles that generated the result we now disagree with. And what guides us as we revise our principles? Some writers have suggested that we must appeal to a more fundamental set of principles, somehow lying in the background. Although this suggestion is useful in many cases—as when we rely on moral principles to correct errors in a legal system, for example—there is a limit to its applicability. As long as we admit that any finite set of principles can lead, at times, to errors in practical evaluation, the appeal to still further principles for the diagnosis and repair of these errors must eventually result in either regress or circularity.

Moderate particularism is an irenic doctrine, which is compatible with generalism. The two ideas can be combined in a view according to which our everyday evaluative reasoning is typically based on principles, but which also admits the possibility of situations in which the accepted stock of principles yields incorrect results, and must then be emended by a process of reasoning that does not itself involve appeal to further principles. This is, I believe, a sensible view, and one that suggests a promising research agenda centered around questions concerning the update and maintenance of complex systems of principles. Many of these questions would have analogs in legal theory, and to the study of normative systems more generally.

In addition to moderate particularism, however, there is also a more radical position that might be called *extreme* particularism. While the moderate view allows for an appeal to principles in the course of our everyday practical evaluation, insisting only that there may be special circumstances in which a straightforward application of these rules yields incorrect results, extreme particularism holds that principles have no role to play in practical evaluation at all.

Since it denies the legitimacy of any appeal to principles whatsoever, extreme particularism is flatly inconsistent with generalism. Nevertheless, it is exactly this radical position that has been advanced by Dancy, who argues that extreme particularism follows from a more

sionary reasons can themselves be excluded: perhaps Colin has promised his mistress to disregard any promises made to his wife. What I disagree with is the suggestion—occasionally found in the literature on practical reasoning, and even in the epistemic literature—that reasons form a kind of hierarchy, so that, just as undercutters are "second-order" reasons, undercutter undercutters are "third-order" reasons, and so on. Perhaps there are some domains, such as the law, where this kind of stratification is the ideal, but even there I suspect the ideal is seldom realized; and it is hard to see why we could assume any sort of neat stratification in less regimented areas. In addition to promising his mistress to disregard any promises made to his wife, Colin might easily, and at the same time, have promised his wife to disregard any promises made to his mistress. His entire life, and the reasons governing it, could be a tangled mess, but the theory would apply all the same.

general *holism* about reasons—the idea that the force of reasons is variable, so that what counts as a reason for an action or conclusion in one setting need not support the same action or conclusion in another.[17]

4.1 The argument

In Dancy's view, holism is a general phenomenon that applies to both practical and epistemic reasons. Both, as he writes, are capable of shifting polarity: a consideration that functions as a reason for some action or conclusion in one context need not serve as a reason for the same action or conclusion in another, and indeed, might even be a reason against it. Dancy presents a variety of cases intended to establish this possibility, in both the practical and theoretical domains. Since these examples follow a common pattern, we consider only two representatives.

Beginning with the practical domain, imagine that I have borrowed a book from you. In most situations, the fact that I have borrowed a book from you would give me a reason to return it to you. But suppose I discover that the book I borrowed is one you had previously stolen from the library. In that context, according to Dancy, the fact that I borrowed the book from you no longer functions as a reason to return it to you; in fact, I no longer have any reason to return it to you at all.[18] In order to illustrate the same phenomenon in the epistemic domain, Dancy turns to a standard example, and one that we have already considered. In most situations, the fact that an object looks red gives me a reason for thinking that it is red. But suppose I know that I have taken Drug #1, which, as we recall, makes red things look blue and blue things look red. In this new context, according to Dancy, the fact that an object looks red no longer functions as a reason for thinking that it is red; it is, instead, a reason for thinking that the object is blue, and so not red.[19]

Let us grant, for the moment, that examples like these are sufficient to establish a general holism of reasons. How is this holistic view supposed to lead to extreme particularism, a thoroughgoing rejection of any role for general principles in practical evaluation? To answer this question, we must consider the nature of the generalizations involved in these principles, and it is useful to focus on a concrete example. Consider, then, the principle that lying is wrong. What could this mean?

We might understand this principle, first of all, as a universally quantified material conditional according to which any action that involves lying is thereby wrong, regardless of the circumstances. Although some writers have endorsed a view along these lines, very few people today would find such an unyielding conception to be tenable. It is, of course, possible to weaken the proposal by viewing the simple principle that lying is wrong as a sort of abbreviation for a much more complicated rule, still a material conditional, but one laden with exception clauses covering all the various circumstances in which it may be acceptable to lie—saving a life, avoiding a pointless insult, and so on. The problem with this suggestion, however, is that no satisfactory rule of this form has ever been displayed, and it is legitimate to doubt our ability even to formulate such fully-qualified rules with any degree of confidence, let alone learn these rules or reason with them.

Alternatively, we might take the principle that lying is wrong to express the idea, not that all acts that involve lying are wrong, or even all acts that involve lying and also satisfy some extensive list of qualifications, but simply that lying is always a feature that counts against an action, a "wrong-making" feature. On this view, the fact that an action involves lying would always count as some reason for taking it to be wrong, even though that action might actually be judged as the right thing to do overall, when various other reasons are taken into account.

[17] The argument is set out with minor variations in a number of publications beginning with Dancy's (1983), but I focus here on the versions found in his (1993), (2000), (2001), and particularly the canonical (2004). Similar arguments have been advanced by others; a particularly clear presentation can be found in Little (2000).

[18] This example can be found in Dancy (1993, p. 60); it is followed by a number of similar examples.

[19] Dancy (2004, p. 74); the example is discussed also in (2000, p. 132) and in Section 3 of his (2001).

The function of principles, then, would be to articulate general reasons for or against actions or conclusions, which may not be decisive, but which at least play an invariant role in our deliberation, always favoring one particular side. This is the view suggested by some of W. D. Ross's remarks about prima facie duties, and it is, in addition, the view of principles that is endorsed by Dancy as the most attractive option available:

> Moral principles, however we conceive of them, seem to be all in the business of specifying features as *general* reasons. The principle that it is wrong to lie, for instance, presumably claims that mendacity is always a wrong-making feature wherever it occurs (that is, it always makes the same negative contribution, though it often does not succeed in making the action wrong overall). (Dancy 2004, p. 76)

But now, suppose reason holism is correct, so that any consideration favoring an outcome in one situation might not favor the same outcome in another. In that case, there would be no general reasons, no considerations to play an invariant role in our deliberation, carrying the same force regardless of context. If the function of principles is to specify general reasons like this, then, there is simply nothing for them to specify; any principle telling us that a reason plays some particular role in our deliberation would have to be incorrect, since there would always be some context in which that reason plays a different role. This is Dancy's conclusion—that, as he says, "a principle-based approach to ethics is inconsistent with the holism of reasons."[20]

4.2 *Evaluating the argument*

Let us start with the epistemic example that Dancy offers to establish reason holism. This example was represented earlier as a threshold default theory, which we now reformulate here, somewhat simplified for convenience.[21] Suppose again that L, R, and $D1$ stand for the respective propositions that the object before me looks red, that it is red, and that I have taken Drug #1; and suppose that the default δ_1 is $L \rightarrow R$, that δ_2 is $L \wedge D1 \rightarrow \neg R$, and that δ_3 is $\top \rightarrow d_1 \prec d_2$. Let $\langle \mathcal{W}, \mathcal{D} \rangle$ be the theory in which $\mathcal{D}$ contains δ_1, δ_2, and δ_3, and in which $\mathcal{W}$ contains L and $D1$, according to which the object looks red but I have taken the drug. Since this is a threshold default theory, $\mathcal{D}$ must again contain, in addition to these three ordinary defaults, the corresponding threshold defaults δ_i^*, of the form $\top \rightarrow t \prec d_i$, for each i from 1 through 3; and $\mathcal{W}$ must contain statements of the form $d_i^* \prec d_j$ for i and j from 1 through 3.

It is easy to see that this default theory yields $\mathcal{S}_1 = \{\delta_1^*, \delta_2^*, \delta_3^*, \delta_2, \delta_3\}$ as its unique proper scenario, supporting the three statements of the form $t \prec d_i$, for i from 1 through 3, as well as $\neg R$ and $d_1 \prec d_2$. The theory thus allows us to conclude that each of the ordinary defaults δ_1, δ_2, and δ_3 lies above threshold, that δ_2 has a higher priority than δ_1, and that the object is not red.

Dancy's argument depends, however, not so much on the conclusions supported by particular situations, but instead, on the statements that are or are not to be classified as reasons in those situations. This is an issue that can usefully be explored here from the standpoint of our analysis of a reason as the premise of a triggered default. And when we do explore the issue from this standpoint, we find that—with the situation represented as above—the proposition L, that the object looks red, is indeed classified as a reason. Why? Because a default is triggered in the context of a scenario just in case its premise is entailed by that scenario and the default lies above threshold. The

[20] Dancy (2004, p. 77). A couple of paragraphs later, Dancy admits that there may actually be a few factors whose role in normative evaluation is not sensitive to context, such as "causing of gratuitous pain on unwilling victims," for example, but he tends to downplay the theoretical significance of allowing isolated exceptions like these, and I agree; a robust generalism would require a wide range of normative principles, not just an occasional principle here and there.

[21] The simplification is that we now ignore Drug #2, which no longer concerns us.

premise of the default δ_1, the statement L, is already included with the ordinary information provided by $\mathcal{W}$, and so entailed by any scenario; and in the context of the scenario $\mathcal{S}_1$, the default lies above threshold. This default is therefore triggered; and so, according to our analysis, its premise is classified as a reason for its conclusion, the statement R, that the object is red.

If L is a reason for R, then why does the scenario not support this conclusion? Well, it is natural to say that a reason, some proposition favoring a conclusion, is *defeated* whenever the default that provides that proposition as a reason is itself defeated. And in the particular case at hand, the default δ_1, which provides L as a reason for the conclusion R, is defeated by the stronger default δ_2, which provides $L \wedge D1$, the proposition that the object looks red but I have taken the drug, as a reason for the conflicting conclusion $\neg R$.

Our initial representation of this situation, then, illustrates the possibility of an alternative interpretation of one of the key examples of the sort that Dancy relies on to establish reason holism. On Dancy's view, the situation in which L and $D1$ both hold—the object looks red but I have taken the drug—provides a context in which, although L is normally a reason for R, it now loses its status as a reason entirely; what is a reason in the normal run of cases is not a reason in this case, and so we are driven to reason holism. On our current representation, by contrast, L is still classified as a reason for R, but simply as a defeated reason.

I mention this possibility here only to show that there are different ways of interpreting those situations in which some familiar consideration appears not to play its usual role as a forceful or persuasive reason. In each case, we must ask: does the consideration fail to function as a reason at all, or does it indeed function as a reason, but simply as one that is defeated? The answer often requires a delicate judgment, and at times different interpretations of the same situation are possible—this is a point we will return to when we consider Dancy's practical example.

In the particular case at hand, as it happens, the idea that looking red still functions as a reason, but simply a defeated reason, is one that Dancy entertains but immediately rejects:

> It is not as if it is some reason for me to believe that there is something red before me, though that reason is overwhelmed by contrary reasons. It is no longer *any reason at all* to believe that there is something red before me; indeed it is a reason for believing the opposite. (Dancy 2004, p. 74)

And, in this particular case, I would have to agree. Once I have taken the drug, it does not seem as if looking red still provides some reason for concluding that the object is red, which is then defeated by a stronger reason to the contrary; in this case, it just does seem that the status of looking red as a reason is itself undermined.

What is crucial to see, however, is that this interpretation of the situation, Dancy's preferred interpretation, can likewise be accommodated within the framework set out here, by appeal to our treatment of undercutting defeat. Let δ_4 be the new default $D1 \rightarrow d_1 \prec t$, according to which the previous default δ_1—which established looking red as a reason for being red—now falls below threshold in any situation in which I have taken the drug. Then the appropriate results can be reached by supplementing the previous theory with this new default, and then satisfying the necessary threshold constraints: formally, let $\langle \mathcal{W}, \mathcal{D} \rangle$ be the threshold default theory in which $\mathcal{D}$ contains the ordinary defaults δ_1, δ_2, δ_3, and now δ_4, as well as the corresponding threshold defaults δ_i^*, of the form $\top \rightarrow t \prec d_i$, for i from 1 through 4, and in which $\mathcal{W}$ contains L and $D1$, together with statements of the form $d_i^* \prec d_j$ for i and j from 1 through 4.

This new default theory now yields $\mathcal{S}_2 = \{\delta_2^*, \delta_3^*, \delta_4^*, \delta_2, \delta_3, \delta_4\}$ as its unique proper scenario, supporting three statements of the form $t \prec d_i$ for i from 2 through 4, as well as $d_1 \prec d_2$, $\neg R$, and $d_1 \prec t$. The theory thus allows us to conclude that the ordinary defaults δ_2, δ_3, and δ_4 lie above threshold, that δ_2 has a higher priority than δ_1, that δ_1 in fact falls below threshold, and of course, that the object is not red.

Now a reason, once again, is the premise of a triggered default, and given this new representation of the situation, the default δ_1, which had previously provided L as a reason for R, is no longer triggered. Why not? Because a default is triggered in the context of a scenario if its premise is entailed by that scenario and, in addition, it lies above threshold; and while the premise of δ_1 is indeed entailed, the default now falls below threshold. Since L, looking red, is no longer the premise of a triggered default, it is not classified as a reason. What is a reason in the usual range of cases is in this situation, therefore, not just a defeated reason, but no longer any reason at all, exactly as Dancy claims.

With this observation in place, we can now turn to an evaluation of Dancy's argument. The argument hinges on the idea that extreme particularism, the rejection of general principles, follows from reason holism. The framework developed here, however, supports reason holism, allowing for the possibility that a consideration that counts as a reason in one situation might not be classified as a reason in another. Yet this framework is itself based on a system of principles, defaults that can be thought of as instances of defeasible generalizations; indeed, what is and what is not a reason in any particular situation is determined by these defaults. The framework thus provides a counterinstance to the idea that reason holism entails extreme particularism, that holism is inconsistent with any form of generalism. Reason holism is consistent with this form of generalism, at least, and so Dancy's argument is, strictly speaking, invalid.

Clearly, there is a disagreement. How should it be diagnosed? Why is it that Dancy considers holism to be inconsistent with any appeal to general principles, while in the framework developed here, it seems that the two ideas, holism and generalism, can be accommodated together?

The disagreement has its roots, I believe, in our different views concerning the meaning of general principles. We both acknowledge that the principles guiding practical reasoning cannot usefully be taken to express universally generalized material conditionals; the practical principle that lying is wrong cannot mean that every action that involves lying is wrong. What Dancy suggests instead, as we have seen, is that these principles should be taken to specify considerations that play an invariant role as reasons. The principle that lying is wrong is supposed to mean that lying always provides some reason for evaluating an action less favorably, even in those cases in which our overall evaluation is favorable. And presumably, the epistemic principle according to which things that look red tend to be red must likewise be taken to mean that looking red always provides some reason for concluding that an object is red, even in those cases in which our overall conclusion is that the object is not red.

Now, given this understanding of general principles, it follows at once—it is *obvious*—that reason holism must lead to their rejection. If holism is correct, so that what counts as a reason in one situation need not be a reason in another, then, of course, any principle that identifies some consideration as playing an invariant role as a reason has to be mistaken. If what it means to say that lying is wrong is that lying always favors a negative evaluation of an action, and there are certain situations in which it does not, then the practical principle itself is faulty, and cannot properly guide our actions; if what it means to say that looking red indicates being red is that looking red always provides some reason for concluding that an object is red, and there are certain situations in which it does not; then again, the epistemic principle is faulty.

I agree, then, that reason holism must entail the rejection of general principles, given Dancy's understanding of these principles. In developing a framework within which holism is consistent with general principles, therefore, I mean to rely on a different understanding of these principles, not as specifying invariant reasons, but instead, as codifying the defaults that guide our reasoning. On this view, the general principle that lying is wrong should be taken to mean simply that lying is wrong by default—that is, to a first approximation, that once we learn that an action involves lying, we ought to judge that it is wrong, unless certain complicating factors interfere. And the principle that looking red indicates being red should likewise be taken to mean that this relation holds by default—that, once we see that an object looks red, we ought to conclude that it is red, again in the absence of other complicating factors.

This explication of general principles as statements codifying defaults involves a frank and explicit appeal to ceteris paribus restrictions; the principles tell us what to conclude in the absence of external complications. Ceteris paribus statements like these are sometimes criticized as vacuous (the joke is that an explication of this kind reduces the substantive claim that lying is wrong to a less substantive claim along the lines of "Lying is wrong except when it isn't"). It is also argued that these statements, which specify the appropriate conclusions in the absence of complicating factors, tell us nothing about the more usual case in which complicating factors are present.

Both of these criticisms have, I believe, some merit when lodged against many of the usual appeals to ceteris paribus generalizations, since these appeals, often, do not move beyond the level of a first approximation. The criticisms have no merit in the present case, however. Here, our first approximation to the meaning of a general principle is nothing but a high level summary of the workings of this principle in the underlying default theory, which specifies in detail, not only what the complicating factors might be—when a default is conflicted, defeated, undercut—but also how the various issues introduced by these complicating factors are to be resolved, and the appropriate set of conclusions to be arrived at in each case. What the present account has to contribute, then, is a concrete theory of default reasoning to support the explication of general principles as defaults, a theory that is precise, supported by our intuitions, and consistent with reason holism.

4.3 Loose ends

I want to conclude by considering two remaining issues presented by Dancy's examples.

First, we have seen how the present account allows us to understand the idea that L, looking red, which normally counts as a reason for R, being red, fails to count as a reason in the situation in which I have taken the drug; but in fact, Dancy claims more than this. What he claims is that L not only fails to count as a reason for R but is actually, in this situation, a reason for $\neg R$. This further claim is not yet supported by the present framework. Can it be? Should it be? I believe that it should be, and it can be, but that it requires a different sort of explanation—roughly, pragmatic rather than semantic.

To see why it sometimes seems natural to think of L as a reason for $\neg R$, consider a slight variation on Dancy's example. Suppose you and I both know that I have taken the drug, which makes red things look blue and blue things look red—that is, we both know $D1$—but that my eyes are closed, so that I cannot see the object before me. Now imagine that I open my eyes, see that the object looks red, and so announce, since I know I have taken the drug, that it must be blue, and therefore not red. And suppose you ask me *why* I concluded that the object is not red. It would then be appropriate for me to respond, "Because I realized that it looks red," apparently citing the fact that the object looks red as a reason for concluding that it is not red. In formulating an account of reasons, we cannot just ignore how we speak, and the ease of my response in this situation certainly suggests that there is a sense in which looking red can, in some cases, be taken as a reason for not being red.

What the current theory tells us, by contrast, is that my real reason for concluding that the object is not red is, not just that it looks red, but that it looks red and I have taken the drug—not just L, but $L \wedge D1$. This is the antecedent of the triggered default δ_2, which supports the conclusion $\neg R$. In the present situation, however, I would argue that there is a pragmatic principle at work that allows me to cite this conjunctive reason by mentioning only a salient part, the simple statement L. Different propositions can be salient at different times and in different ways. In this case, what makes L salient is simply that it provides *new* information. The statement $D1$ was already part of our background knowledge. Although I actually mean to cite the proposition $L \wedge D1$ as my reason for concluding that the object is not red, then, a principle of conversational economy allows me to do so simply by mentioning the new information that, taken together with our shared background information, entails this proposition.

This pragmatic analysis can be supported by considering another variant of the situation. Suppose, this time, that my eyes are open and I can see that the object looks red, but I am not yet aware that I have taken the drug. Since the object looks red, my initial judgment would be that it is red. But imagine that I am now told about the drug, and so revise my judgment to conclude that the object is blue, and so not red. In this case, if I were again asked why I changed my mind, it would no longer be appropriate to cite the fact that the object looks red. After all, I knew that it looked red before, when I judged it to be red. Instead, I would now be more likely to say, "Because I learned that I took the drug."

This response conforms to the pattern of the previous variant, and so suggests a common pragmatic principle at work. In both cases, my final state of information can be represented by the set $\{L, D1\}$, which entails $L \wedge D1$, the actual reason for my conclusion that the object is not red. In the first case, I arrive at this final state of information by adding L to the set $\{D1\}$ of background information; in the second, I arrive at this final state by adding $D1$ to the background set $\{L\}$. In both cases, then, I am able to cite $L \wedge D1$ itself as a reason for my conclusion in exactly the same way, by mentioning only the new information that, taken together with our shared background knowledge, entails this proposition—L in the first case, and $D1$ in the second.

The second, and concluding, matter that I want to discuss is the practical example offered by Dancy to illustrate reason holism, which we have yet to consider in any detail: I borrow a book from you, but then learn that this book is one you have previously stolen from the library. According to Dancy, this situation is one in which my borrowing a book from you, which generally functions as a reason for returning it to you, no longer counts as such a reason. What is generally a reason is not a reason in this particular situation, and so again, we have holism.

In his discussion of this example, just as in his discussion of the epistemic case, Dancy explicitly considers and rejects the possibility that the consideration, having borrowed a book, is still functioning as a reason, but simply as a defeated reason:

> It isn't that I have *some* reason to return it to you and more reason to put it back in the library. I have no reason at all to return it to you. (Dancy 1993, p. 60)

But here, in contrast to the epistemic case, I do not think the matter is so straightforward; I cannot agree that Dancy's reading of the situation provides the unique interpretation, or even the most natural. I would, in fact, be inclined toward a very different interpretation myself. In the situation as described, I would tend to feel that my having borrowed the book from you gives me a personal obligation to return it to you, and that it is simply not my business to supervise your relations with the library; that is someone else's job.

This autobiographical detail—how I personally would view the matter—carries little importance except that it suggests a different and, I hope, coherent interpretation of the situation Dancy describes. The situation is especially interesting, in fact, precisely because it does serve so naturally to illustrate what I consider to be a pervasive phenomenon: situations described at this level of generality often allow for a number of different, equally coherent interpretations. In order to establish this point, I will simply list five different interpretations of the situation as Dancy describes it, arranged in a sort of spectrum depending on the relative strength given to my reasons for returning the book to you compared to my reasons for returning it to the library.

First, there is my interpretation: borrowing the book from you gives me a reason for returning it to you, but your having stolen it from the library gives me no particular reason to do anything at all (though it might well count as a reason supporting certain actions by the library police), so that what I ought to do is return the book to you. Second, I can imagine someone who agrees that my borrowing the book gives me a reason for returning it to you, but who also feels that your having stolen it gives me some reason for returning it to the library, though the reason for returning it to you is stronger, so that, on balance, I ought to return it to you. Third, I can imagine someone who feels that my borrowing the book gives me a reason for returning it to you, that your having stolen it also gives me a reason for returning it to the library, and that these two reasons are, in fact, incomparable in priority, so that I am now faced with a dilemma, incomparable reasons supporting conflicting actions; I would then have to resolve the matter in whatever way I resolve dilemmas, perhaps flipping a coin or seeking further advice. Fourth, I can imagine someone who feels that my borrowing the book gives me a reason for returning it to you, that your having stolen it gives me a reason for returning it to the library, but in this case, that the reason for returning it to the library is stronger, so that I ought to return it to the library. And, fifth, we have Dancy's preferred interpretation: your having stolen the book both gives me a reason for returning the book to the library and undercuts any reason I might otherwise have had for returning it to you, so that I ought to return it to the library.

I hope the reader will agree that each of these interpretations is indeed coherent; a rational person could adopt any of them. Moreover, each of these interpretations can be given a precise representation, and distinguished from the others, within the framework developed here. I will forbear from actually displaying these five different representations in the text of this paper—the interested reader can find them in the appendix—but not from drawing the moral that the possibility of these different formalizations suggests. This is, I believe, that the formal study of reasons presented here carries benefits analogous to the benefits of formal work in other areas. By providing a precise representation of reasons and their interactions, it allows us to tease apart different interpretations of particular situations that might otherwise escape notice, to suggest new possibilities, and where disagreement occurs, to localize the source of that disagreement.

Appendix A. The library book example

This appendix presents threshold default theories representing, in order, each of the five interpretations mentioned in the text of Dancy's library book example. Let B, S, Y, and L represent the respective propositions that I borrowed the book from you, that you stole it from the library, that I return the book to you, and that I return it to the library; and suppose that Y and L are inconsistent—I cannot return the book both to you and to the library. Suppose that δ_1 is the default $B \rightarrow Y$, that δ_2 is $S \rightarrow L$, that δ_3 is $\top \rightarrow d_2 \prec d_1$, that δ_4 is $\top \rightarrow d_1 \prec d_2$, and that δ_5 is $S \rightarrow d_1 \prec t$.

Interpretation #1: Let $\langle \mathcal{W}, \mathcal{D} \rangle$ be the threshold default theory in which the set $\mathcal{W}$ of ordinary information contains B and S, the set $\mathcal{D}$ of defaults contains δ_1, and both $\mathcal{W}$ and $\mathcal{D}$ contain, in addition, the necessary threshold information. The unique proper scenario based on this theory is $\mathcal{S}_1 = \{\delta_1^*, \delta_1\}$, supporting the statements $t \prec d_1$ and Y. Since δ_1 lies above threshold and its premise is entailed, this default is triggered, providing B as a reason for Y, which is what I ought to do.

Interpretation #2: Let $\langle \mathcal{W}, \mathcal{D} \rangle$ be the threshold default theory in which the set $\mathcal{W}$ of ordinary information contains B and S, the set $\mathcal{D}$ of defaults contains δ_1, δ_2, and δ_3, and both $\mathcal{W}$ and $\mathcal{D}$ contain, in addition, the necessary threshold information. The unique proper scenario based on this theory is $\mathcal{S}_2 = \{\delta_1^*, \delta_2^*, \delta_3^*, \delta_1, \delta_3\}$, supporting the statements $t \prec d_i$ for i from 1 to 3, $d_2 \prec d_1$, and Y. Since both δ_1 and δ_2 lie above threshold and their premises are entailed, both defaults are triggered, providing B as a reason for Y and S as a reason for L. However, since δ_1 is accorded a higher priority than δ_2, the reason S for L is defeated by the reason B for Y, so that what I ought to do is Y.

Interpretation #3: Let $\langle \mathcal{W}, \mathcal{D} \rangle$ be the threshold default theory in which the set $\mathcal{W}$ of ordinary information contains B and S, the set $\mathcal{D}$ of defaults contains δ_1 and δ_2 without any specification of priority, and both $\mathcal{W}$ and $\mathcal{D}$ contain, in addition, the necessary threshold information. This theory allows two proper scenarios, both $\mathcal{S}_3 = \{\delta_1^*, \delta_2^*, \delta_1\}$ and $\mathcal{S}_3' = \{\delta_1^*, \delta_2^*, \delta_2\}$, where both $\mathcal{S}_3$ and $\mathcal{S}_3'$ support $t \prec d_i$ where i is 1 or 2, but $\mathcal{S}_3$ supports Y while $\mathcal{S}_3'$ supports L. Since δ_1 and δ_2

lie above threshold in both scenarios, and their premises are entailed, both defaults are triggered in both scenarios, providing B as a reason for Y and S as a reason for L. Since there is no priority information, neither default is defeated and so there is a conflict; the two scenarios represent different ways of resolving the conflict.

Interpretation #4: Let $\langle \mathcal{W}, \mathcal{D} \rangle$ be the threshold default theory in which the set $\mathcal{W}$ of ordinary information contains B and S, the set $\mathcal{D}$ of defaults contains δ_1, δ_2, and δ_4, and both $\mathcal{W}$ and $\mathcal{D}$ contain, in addition, the necessary threshold information. The unique proper scenario based on this theory is $\mathcal{S}_2 = \{\delta_1^*, \delta_2^*, \delta_4^*, \delta_2, \delta_4\}$, supporting the statements $t \prec d_i$ where i is 1, 2, or 4, as well as $d_1 \prec d_2$, and L. Since both δ_1 and δ_2 lie above threshold and their premises are entailed, both defaults are triggered, providing B as a reason for Y and S as a reason for L. However, since δ_2 is now accorded a higher priority than δ_1, the reason B for Y is now defeated by the reason S for L, so that what I ought to do is L.

Interpretation #5: Let $\langle \mathcal{W}, \mathcal{D} \rangle$ be the threshold default theory in which the set $\mathcal{W}$ of ordinary information contains B and S, the set $\mathcal{D}$ of defaults contains δ_1, δ_2, and δ_5, and both $\mathcal{W}$ and $\mathcal{D}$ contain, in addition, the necessary threshold information. The unique proper scenario based on this theory is $\mathcal{S}_5 = \{\delta_2^*, \delta_5^*, \delta_2, \delta_5\}$, supporting the statements $t \prec d_i$ where i is 2 or 5, as well as $d_1 \prec t$, and L. In this case, since δ_2 lies above threshold and its premise is entailed, the default is triggered, providing S is a reason for L; but since δ_1 now falls below threshold, it is no longer triggered, and so B no longer counts as a reason for Y. Since there is a reason for L and no conflicting reason at all, L is what I ought to do.

Acknowledgements

Three people were extraordinarily generous to me during the preparation of this paper. Henry Richardson read an early draft and discussed it with me extensively; his insight and enthusiasm kept me working on a project I would not have finished without him. Jonathan Dancy read the same draft (a long paper by an unknown author that arrived on his desk out of the blue), sent me a remarkably perceptive set of comments, and then invited me to present this material at the Texas particularism conference, where I was able to discuss it with an audience of experts. Finally, Mark Schroeder, who read the early draft as well as the current version, has been helpful beyond all measure during a series of conversations spanning a period of two years; many of the key ideas in this paper were formed during those conversations.

References

[Brewka, 1994] Gerhard Brewka. Reasoning about priorities in default logic. In *Proceedings of the Twelveth National Conference on Artificial Intelligence (AAAI-94)*, pages 940–945. AAAI/MIT Press, 1994.

[Brewka, 1996] Gerhard Brewka. Well-founded semantics for extended logic programs with dynamic preferences. *Journal of Artificial Intelligence Research*, 4:19–36, 1996.

[Broome, 2004] John Broome. Reasons. In R. Jay Wallace, Michael Smith, Samuel Scheffler, and Philip Pettit, editors, *Reason and Value: Essays on the Moral Philosophy of Joseph Raz*, pages 28–55. Oxford University Press, 2004.

[Dancy, 1983] Jonathan Dancy. Ethical particularism and morally relevant properties. *Mind*, 92:530–547, 1983.

[Dancy, 1993] Jonathan Dancy. *Moral Reasons*. Basil Blackwell Publisher, 1993.

[Dancy, 2000] Jonathan Dancy. The particularist's progress. In Brad Hooker and Margaret Little, editors, *Moral Particularism*. Oxford University Press, 2000.

[Dancy, 2001] Jonathan Dancy. Moral particularism. In Edward Zalta, editor, *The Stanford Encyclopedia of Philosophy (Summer 2001 Edition)*. Stanford University, 2001. Available at http://plato.stanford.edu/-archives/sum2001/entries/moral-particularism/.

[Dancy, 2004] Jonathan Dancy. *Ethics Without Principles*. Oxford University Press, 2004.

[Fahlman, 1979] Scott Fahlman. *NETL: A System for Representing and Using Real-world Knowledge*. The MIT Press, 1979.

[Franklin, 1772] Benjamin Franklin. Letter to Joseph Priestly, 1772. Reprinted in Frank Mott and Chester Jorgenson, editors, *Benjamin Franklin: Representative Selections*, pages 348–349, American Book Company, 1936; pagination refers to this version.

[Gordon, 1993] Thomas Gordon. *The Pleadings Game: An Artificial-Intelligence Model of Procedural Justice*. PhD thesis, Technische Hochschule Darmstadt, 1993.

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

[Horty, 1994b] John Horty. Some direct theories of nonmonotonic inheritance. In D. Gabbay, C. Hogger, and J. 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.

[Horty, 2002] John Horty. Skepticism and floating conclusions. *Artificial Intelligence*, 135:55–72, 2002.

[Horty, 2003] John Horty. Reasoning with moral conflicts. *Nous*, 37:557–605, 2003.

[Horty, 2007] John Horty. Defaults with priorities. Forthcoming in *Journal of Philosophical Logic*, 2007+.

[Little, 2000] Margaret Little. Moral generalities revisited. In Brad Hooker and Margaret Little, editors, *Moral Particularism*. Oxford University Press, 2000.

[Makinson, 1994] David Makinson. General patterns in nonmonotonic reasoning. In D. Gabbay, C.Hogger, and J. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning*, pages 35–110. Oxford University Press, 1994.

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

[Pollock, 1970] John Pollock. The structure of epistemic justification. In *Studies in the Theory of Knowledge*, American Philosophical Quarterly Monograph Series, number 4, pages 62–78. Basil Blackwell Publisher, Inc., 1970.

[Pollock, 1995] John Pollock. *Cognitive Carpentry: A Blueprint for How to Build a Person*. The MIT Press, 1995.

[Prakken and Sartor, 1995] Henry Prakken and Giovanni Sartor. On the relation between legal language and legal argument: assumptions, applicability, and dynamic priorities. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law (ICAIL-95)*. The ACM Press, 1995.

[Prakken and Sartor, 1996] Henry Prakken and Giovanni Sartor. A dialectical model of assessing conflicting arguments in legal reasoning. *Artificial Intelligence and Law*, 4:331–368, 1996.

[Prakken and Vreeswijk, 2002] Henry Prakken and Gerard Vreeswijk. Logical systems for defeasible argumentation. In Dov Gabbay and F. Guethner, editors, *Handbook of Philosophical Logic (Second Edition)*, pages 219–318. Kluwer Academic Publishers, 2002.

[Raz, 1975] Joseph Raz. *Practical Reasoning and Norms*. Hutchinson and Company, 1975. Reprinted by Oxford University Press, 2002.

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

[Schroeder, 2005] Mark Schroeder. Holism, weight, and undercutting. Unpublished manuscript, 2005.

[Schroeder, 2007] Mark Schroeder. Weighting for a plausible Humean theory of reasons. *Nous*, 41:138–160, 2007.

[Touretzky *et al.*, 1987] 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.

[Touretzky, 1986] David Touretzky. *The Mathematics of Inheritance Systems*. Morgan Kaufmann, 1986.

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