JOHN HORTY

DEFAULTS WITH PRIORITIES

1. INTRODUCTION

If we are told only that Tweety is a bird, it is natural to conclude that Tweety is able to fly; our everyday reasoning seems to be governed by a default according to which birds, as a rule, can fly. But if we are then told that Tweety is actually unable to fly – information that it is, after all, consistent with our initial premise – we would withdraw our original conclusion. Any logic put forth to capture this form of reasoning must therefore exhibit a *nonmonotonic* consequence relation, allowing for the possibility that the conclusion set might shrink as the premise set grows.

The study of nonmonotonic logics began in earnest about twenty-five years ago, with simultaneous exploration along several different avenues. This paper is concerned with the consistency-based approach, exemplified by Reiter's default logic [23], one of the most widely applied nonmonotonic logics, and arguably the most successful.

The paper focuses on priority relations among default rules, a matter that was not treated in Reiter's original theory. To illustrate, suppose we are told that Tweety is a penguin, and therefore a bird. Then it seems that two conflicting defaults come into play. There is the default according to which birds can fly, but there is also a default according to which penguins, as a rule, cannot fly. Still, in spite of these two conflicting defaults, we have no difficulty arriving at a definite conclusion: since the second of these two defaults is naturally thought to carry a higher priority than the first, we favor this second default over the first, and conclude that Tweety is unable to fly.

Priority relations among defaults can have different sources. In this particular case, the priority of the second default over the first has to do with specificity: a penguin is a specific kind of bird, and so information about penguins in particular should take precedence over information about birds in general. But there are other 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 about oncoming precipitation. But the weather channel is more reliable, so that I favor its predictions in case of conflict. And if we move to the normative interpretation of defaults, as explored in my [11] and [15], for example, 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.

My concern here, however, is not with the source of priority relations among default rules, but instead, with the way in which these priority relations – which I will simply take as given – are to be accommodated within a logic for default reasoning. This is not a new topic; there are already several proposals addressing the problem in the literature, some of which I will return to later on.

The present paper explores the problem from a new angle. One area in which the analysis of priority relations among default rules has met with considerable success has been in the theory of nonmonotonic inheritance reasoning, initiated by Touretzky in [26], developed by Thomason, Touretzky, and myself in a series of papers that includes [17], [27], and [28], and then systematized, to some extent, in my [12]. From the perspective of nonmonotonic reasoning more generally, the ideas and techniques introduced in this work have often seemed to be rather narrow and specialized, and perhaps applicable only to the very restricted language of inheritance networks. My aim in this paper is to show, to the contrary, that these ideas can be generalized to richer languages, and to explore one way of doing so; the result is a promising account of prioritized default reasoning that compares favorably to other work in the area.

2. BASIC CONCEPTS

2.1. *Default Theories and Scenarios*

We assume as background an ordinary propositional language, with $\supset$, $\neg$, $\wedge$, and $\vee$ as the usual connectives, and with $\top$ as a constant representing truth. The turnstile $\vdash$ indicates standard logical consequence, and where $\mathcal{E}$ is a set of formulas, we define $Th(\mathcal{E}) = \{A : \mathcal{E} \vdash A\}$ as its logical closure, the set of formulas derivable from $\mathcal{E}$.

Where A and B are formulas from the background language, we then let $A \rightarrow B$ represent the *default rule* that allows us to conclude B, by default, whenever it has been established that A. It is most useful, I believe, to think of default rules as providing *reasons* for conclusions. If

B stands for the statement that Tweety is a bird, and F for the statement that Tweety can fly, then the particular default $B \rightarrow F$ tells us that Tweety's being a bird functions as a reason for concluding that he is able to fly; this particular default can be viewed as an instance for Tweety of a general default according to which birds, as a rule, are able to fly.

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 *Premise*(δ) is the statement A and *Conclusion*(δ) 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

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

as the set of their conclusions.

As we have seen, some defaults have 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. Among the various possible ordering constraints, it is most natural to require only that this priority relation should be transitive and irreflexive – that is, a strict partial ordering. We suppose that this priority relation is likewise lifted from individual defaults to sets, so that $\mathcal{D} < \mathcal{D}'$ means that $\delta < \delta'$ for each δ in $\mathcal{D}$ and δ' in $\mathcal{D}'$; and for convenience, we abbreviate $\{\delta\} < \mathcal{D}'$ as $\delta < \mathcal{D}'$.

Where $\mathcal{D}$ is a set of defaults, and $<$ is a strict partial ordering on $\mathcal{D}$, we let $\mathcal{D}_<$ stand for the pair $\langle \mathcal{D}, < \rangle$, an ordered set of defaults. And finally, where $\mathcal{W}$ is some set of formulas from our background language and $\mathcal{D}_<$ is an ordered set of defaults, we define an *ordered 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 an agent as a basis for its reasoning.[1]

The goal of a default logic is to specify the *belief sets* supported by default theories, where we adopt the common idealization of a belief set as a logically closed set of formulas. Defaults are generally though of as rules for extending the conclusions derivable from a set of ordinary formulas beyond its classical consequences, and for this reason, the belief sets associated with default theories are often referred to as *extensions*. Throughout this paper, however, we will concentrate in the first instance, not on belief sets 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; we can then take

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

as the belief set that is *generated* by such a scenario.

Where $\mathcal{S}$ is a scenario based on the default theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$, we will say, for convenience, that a statement A is consistent with, or entailed by, $\mathcal{S}$ just in case A is consistent with, or entailed by, the set $\mathcal{W} \cup Conclusion(\mathcal{S})$; and likewise, that a default $\mathcal{D}$ is consistent with or entailed by $\mathcal{S}$ just in case the statement $Conclusion(\mathcal{D})$ is consistent with or entailed by $\mathcal{S}$.

From an intuitive standpoint, a scenario is supposed to represent the set of defaults that have been accepted by an agent, at some stage of its reasoning process, as providing sufficient support for their conclusions. Our central task in this paper is to characterize, as we will say, the *proper scenarios* – those scenarios that might ultimately be accepted by an ideal reasoning agent on the basis of the information contained in an ordered default theory. With this notion in hand, we can then define the extensions of ordered default theories quite simply, as the belief sets that are generated by their proper scenarios.

2.2. *Binding Defaults*

We begin with the concept of a binding default. If defaults provide reasons, then the binding defaults 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 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 consider first – triggering, conflict, and defeat.

Defaults provide reasons, but of course, not every reason is applicable in every context, every scenario: the default that birds fly, for example, provides no support at all for the conclusion that Tweety flies unless the agent is already committed, by its current scenario, to the proposition that Tweety is a bird. The defaults that are triggered in a particular scenario, representing the applicable reasons, are simply those whose premises are entailed by that scenario.

DEFINITION 1 (TRIGGERED DEFAULTS). Where S is a scenario based on the ordered default theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$, the defaults from $\mathcal{D}$ that are *triggered* in S are those belonging to the set

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

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 $B \to F$ and $F \to W$, instances for Tweety of the general defaults that birds fly and that flying animals have wings. Imagine that a reasoning agent is provided with the ordered default theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$ as 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 $S_0 = \emptyset$. We then have $Triggered_{\mathcal{W}, \mathcal{D}_<}(S_0) = \{\delta_1\}$ so that, in this initial scenario, the default δ_1 provides the agent with a reason for its conclusion, the proposition F. Now suppose the agent does in fact accept this reason, and so moves to the new scenario $S_1 = \{\delta_1\}$. Then since $Triggered_{\mathcal{W}, \mathcal{D}_<}(S_1) = \{\delta_1, \delta_2\}$, the default δ_2 now provides the agent, in this new scenario, with a reason for the conclusion W.

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 scenario already entails the negation of its conclusion.

DEFINITION 2 (CONFLICTED DEFAULTS). Where S is a scenario based on the ordered default theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$, the defaults from $\mathcal{D}$ that are *conflicted* in S are those belonging to the set

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

The intuitive force of this restriction can be illustrated through a standard example, known as the Nixon Diamond (because its depiction as an inheritance network has the shape of a diamond). Let Q, R, and 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 $Q \to P$ and $R \to \neg P$, instances of the general rules that

Quakers tend to be pacifists and that Republicans tend not to be pacifists. Imagine that the theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$ is provided to the agent as initial information, where $\mathcal{W} = \{Q, R\}$, $\mathcal{D} = \{\delta_1, \delta_2\}$, and the ordering is again empty; and suppose 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 have $Triggered_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S}_0) = \{\delta_1, \delta_2\}$; 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: $Conflicted_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S}_0) = \emptyset$.

The agent must 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 – so that it moves to the new scenario $\mathcal{S}_1 = \{\delta_1\}$. In this new scenario, the other default will now be classified as conflicted: $Conflicted_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S}_1) = \{\delta_2\}$. The reason provided by δ_2 loses its force, since the agent has already settled on a contrary 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, as we will see, this notion is considerably more complicated 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 initial example, which is known as the Tweety Triangle (because of its triangular shape when depicted as an inheritance network). Again, we let P, B, and F stand for the propositions that Tweety is a penguin, that Tweety is a bird, and that Tweety flies; and 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. Suppose 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 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 $Triggered_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S}_0) = \{\delta_1, \delta_2\}$; 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 $Conflicted_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S}_0) = \emptyset$; neither of these defaults is itself conflicted. Nevertheless, it does not seem on intuitive grounds that the agent should be free, as before, to settle this conflict however it chooses. Here, the default δ_1,

supporting the conclusion F, seems to be, in some sense, defeated by δ_2, since this default is stronger and supports the conflicting conclusion $\neg F$.

Our challenge is to provide a general definition of the concept of defeat at work in cases like this. Motivated by the Tweety Triangle, it is natural to begin with the proposal that a default should be defeated in a scenario if that scenario triggers some stronger default with a conflicting conclusion – or put formally: that the default δ should be defeated in the scenario $\mathcal{S}$ if there is some default $\delta' \in Triggered_{\mathcal{W,D_<}}(\mathcal{S})$ such that (1) $\delta < \delta'$ and (2) $Conclusion(\delta') \vdash \neg Conclusion(\delta)$. This simple proposal is nothing but a straightforward adaptation of the notion of preemption developed for the restricted language of inheritance hierarchies; but unfortunately, it is too simple in the present, more general setting, and for two reasons. First, it seems possible for a default to be defeated, not just by a single stronger default, but by a set of stronger defaults – a *defeating set*, rather than a single defeater – each of which may be individually consistent with the original default, but which are inconsistent with this default when taken together. And second, in determining whether one default, or set of defaults, conflicts with another, it seems that we can legitimately appeal to certain facts to which the agent is already committed, through either its initial information or its current scenario.

Both of these difficulties can be illustrated by an abstract example in which the default δ_1 is $\top \rightarrow (A \supset B)$, δ_2 is $\top \rightarrow (B \supset C)$, and δ_3 is $\top \rightarrow \neg C$. Consider the theory $\langle \mathcal{W, D_<} \rangle$ where $\mathcal{W} = \{A\}$, $\mathcal{D} = \{\delta_1, \delta_2, \delta_3\}$, and we have both $\delta_3 < \delta_1$ and $\delta_3 < \delta_2$; and suppose the agent has not yet accepted any of the three defaults, so that its current scenario is $\mathcal{S}_0 = \emptyset$. Here, it seems reasonable to say that the single default δ_3 is defeated by the set $\mathcal{S}_1 = \{\delta_1, \delta_2\}$. Why? Because both defaults belonging to $\mathcal{S}_1$ are triggered in the current scenario; because both of these defaults have a higher priority than δ_3; and because, when taken together with the statement A, to which the agent is already committed, the conclusions of these defaults conflict with the conclusion of δ_3.

Generalizing from this example, it may now appear that we reach a proper analysis of defeat by stipulating that: the default δ is defeated in the scenario $\mathcal{S}$ just in case there is a defeating set $\mathcal{D}' \subseteq Triggered_{\mathcal{W,D_<}}(\mathcal{S})$ such that (1) $\delta < \mathcal{D}'$ and (2) $\mathcal{W} \cup Conclusion(\mathcal{S} \cup \mathcal{D}') \vdash \neg Conclusion(\delta)$. Let us refer to this proposal as the *candidate definition*. In fact, this candidate definition is nearly correct, but requires further refinement in order to handle certain problems arising when a potential defeating set is inconsistent with the agent's current scenario.

The problems can be illustrated by an example that extends the earlier Nixon Diamond with a weaker but irrelevant default. As before, let Q, R, and P represent the propositions that Nixon is a Quaker, a Republican, and a pacifist; let δ_1 be $Q \rightarrow P$ and δ_2 be $R \rightarrow \neg P$. But this time, let S represent some proposition that is entirely irrelevant to Nixon's pacifism, perhaps the proposition that Nixon enjoys the seashore; and let δ_3 be the default $\top \rightarrow S$, an instance for Nixon of the rule that people in general tend to enjoy the seashore. Suppose the agent is provided with $\langle \mathcal{W}, \mathcal{D}_< \rangle$ as initial information, where $\mathcal{W} = \{Q, R\}$, $\mathcal{D} = \{\delta_1, \delta_2, \delta_3\}$, and the ordering tells us that the new default has a lower priority than the previous two: $\delta_3 < \delta_1$ and $\delta_3 < \delta_2$. And imagine that, as before, the agent has selected the default δ_1 over the conflicting default δ_2, so that its current scenario is $\mathcal{S}_1 = \{\delta_1\}$.

Now, once the conflict concerning Nixon's pacifism has been settled, can the agent then simply accept the additional default δ_3 and so conclude S, that Nixon likes the seashore? The intuitive answer is Yes. The new default provides a reason for this conclusion, and there is apparently nothing in the vicinity to oppose this reason. Unfortunately, however, the candidate definition tells us otherwise – that, in the agent's current scenario, the new default δ_3 is actually defeated. How can this be? Well, taking $\mathcal{D}' = \{\delta_2\}$ as a potential defeating set, it is clear to begin with that $\mathcal{D}' \subseteq Triggered_{\mathcal{W},\mathcal{D}_<}(\mathcal{S}_2)$. Furthermore, we have (1) $\delta_3 < \mathcal{D}'$, and since the set $\mathcal{W} \cup Conclusion(\mathcal{S}_1 \cup \mathcal{D}')$ – that is, $\mathcal{W} \cup \{P, \neg P\}$ – is inconsistent, entailing anything at all, we also have (2) $\mathcal{W} \cup Conclusion(\mathcal{S}_1 \cup \mathcal{D}') \vdash \neg Conclusion(\delta_3)$.

This example might seem to suggest that the candidate definition should be supplemented with a restriction according to which the defeating set $\mathcal{D}'$ should be consistent with the current scenario $\mathcal{S}$. Perhaps the original clause (2) should be replaced with a pair of clauses requiring both (2a) that $\mathcal{W} \cup Conclusion(\mathcal{S} \cup \mathcal{D}')$ is consistent, and (2b) that $\mathcal{W} \cup Conclusion(\mathcal{S} \cup \mathcal{D}') \vdash \neg Conclusion(\delta)$. However, this suggestion will not work either, as we can see by returning to the Tweety Triangle. Suppose, in this example, that the reasoning agent has mistakenly come to accept δ_1 – that is, the default $B \rightarrow F$, according to which Tweety flies because he is a bird – so that its current scenario is $\mathcal{S}_1 = \{\delta_1\}$. From an intuitive standpoint, we would nevertheless like δ_1 to be defeated by δ_2 – the stronger default $P \rightarrow \neg F$, according to which Tweety does not fly because he is a penguin. But this defeat relation would no longer hold, since the new clause (2a) requires that a default can be defeated only by another that is consistent with the agent's current scenario, and δ_2 is not consistent with $\mathcal{S}_1$.

What I would like to suggest, instead, is that the defeating set $\mathcal{D}'$ should be consistent, not necessarily with the agent's current scenario $\mathcal{S}$ as it stands, but with the scenario that results when a certain subset $\mathcal{S}'$ is retracted from this current scenario, so that the defeating set can then be consistently accommodated. For convenience, we let

$$\mathcal{S}^{\mathcal{D}'/\mathcal{S}'} = (\mathcal{S} - \mathcal{S}') \cup \mathcal{D}'$$

indicate the result of retracting the defaults belonging to $\mathcal{S}'$ from the scenario $\mathcal{S}$, and then supplementing what remains with the defaults from $\mathcal{D}'$ – or more simply, as the notation suggests, replacing $\mathcal{S}'$ by $\mathcal{D}'$ in $\mathcal{S}$. The suggestion, then, is to require, not that the defeating set $\mathcal{D}'$ must be consistent with the scenario $\mathcal{S}$, but simply that there should be some appropriate set $\mathcal{S}'$ such that $\mathcal{S}^{\mathcal{D}'/\mathcal{S}'}$ is consistent. Returning to our variant of the Tweety Triangle, again taking $\mathcal{S}_1 = \{\delta_1\}$ as the agent's current scenario, if we now suppose that $\mathcal{D}' = \{\delta_2\}$ and $\mathcal{S}' = \{\delta_1\}$, then it turns out that $\mathcal{S}_1^{\mathcal{D}'/\mathcal{S}'} = \{\delta_2\}$ is consistent; and since this set entails $\neg Conclusion(\delta_1)$, the desired defeat relation is restored.

The key to this proposal is that, in order to accommodate a defeating set, we are free to retract certain defaults to which the agent is already committed. But are there any constraints on this process of accommodation; can we retract just anything at all from the agent's current scenario? No, there are limits. The definition to be presented here is based on the idea that the set $\mathcal{S}'$ of retracted defaults and the defeating set $\mathcal{D}'$ are subject to the constraint that $\mathcal{S}' < \mathcal{D}'$ – the defaults belonging to $\mathcal{S}'$ must be uniformly weaker than those belonging to $\mathcal{D}'$. We can retract as many defaults from the agent's current scenario as necessary in order to accommodate a defeating set, as long as the defaults we retract are themselves lower in priority than those we are attempting to accommodate.

DEFINITION 3 (DEFEATED DEFAULTS). Where $\mathcal{S}$ is a scenario based on the ordered 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

$Defeated_{\mathcal{W},\mathcal{D}_<}(\mathcal{S}) =$

$\{\delta \in \mathcal{D} :$ there is a set $\mathcal{D}' \subseteq Triggered_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$ such that

(1) $\delta < \mathcal{D}'$,

(2) there is a set $\mathcal{S}' \subseteq \mathcal{S}$ with $\mathcal{S}' < \mathcal{D}'$ such that

(a) $\mathcal{W} \cup Conclusion(\mathcal{S}^{\mathcal{D}'/\mathcal{S}'})$ is consistent,

(b) $\mathcal{W} \cup Conclusion(\mathcal{S}^{\mathcal{D}'/\mathcal{S}'}) \vdash \neg Conclusion(\delta)\}$.

When a default δ is defeated in accord with this definition, with $\mathcal{D}'$ as its defeating set, we say that $\mathcal{S}'$ is an *accommodating set* for $\mathcal{D}'$, a set of defaults whose retraction from the current scenario $\mathcal{S}$ allows the defeating set to be accommodated.

Evidently, this definition of defeat allows an accommodating set to be larger than necessary, in the sense that it might contain defaults that do not actually need to be retracted from the current scenario in order to accommodate the defeating set. We can, however, define the stricter notion of a minimal accommodating set, as follows: where some default is defeated in the scenario $\mathcal{S}$, with $\mathcal{D}'$ as a defeating set, $\mathcal{S}^*$ is a *minimal accommodating set* for $\mathcal{D}'$ just in case $\mathcal{S}^*$ is an accommodating set for $\mathcal{D}'$ and, for any proper subset $\mathcal{S}'$ of $\mathcal{S}^*$, the set $\mathcal{W} \cup Conclusion(\mathcal{S}^{\mathcal{D}'/\mathcal{S}'})$ is inconsistent. A minimal accommodating set, then, is some minimal set of defaults that must be retracted from the current scenario in order to accommodate a defeating set. And it is easy to see, first of all, that the concept of defeat remains unchanged if we restrict our attention to minimal accommodating sets, and second, that any defeating set which is already consistent with the current scenario has the empty set as its unique minimal accommodating set.

OBSERVATION 1. *Where $\mathcal{S}$ is a scenario based on the ordered default theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$, suppose δ is defeated in $\mathcal{S}$, with $\mathcal{D}'$ as a defeating set and $\mathcal{S}'$ as an accommodating set for $\mathcal{D}'$. Then there is some $\mathcal{S}^* \subseteq \mathcal{S}'$ such that δ is likewise defeated in $\mathcal{S}$ with $\mathcal{D}'$ as a defeating set and $\mathcal{S}^*$ as a minimal accommodating set for $\mathcal{D}'$.*

OBSERVATION 2. *Where $\mathcal{S}$ is a scenario based on the ordered default theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$, suppose δ is defeated in $\mathcal{S}$, with $\mathcal{D}'$ as a defeating set. Then $\mathcal{S}^* = \emptyset$ is a minimal accommodating set for $\mathcal{D}'$ if and only if $\mathcal{W} \cup Conclusion(\mathcal{S} \cup \mathcal{D}')$ is consistent.*

The reader is invited to verify that our definition of a defeated default yields the correct defeat relations in the various examples considered here, as well as others of his or her own devising. Any definition this complicated, however, needs a justification apart from its application to particular examples, and I offer two.

We have, in the first place, a clear rationale for preferring conclusions based on $\mathcal{S}^{\mathcal{D}'/\mathcal{S}'}$ – the new scenario, which results from the original by retracting the accommodating set and adding the defeating set – to conclusions based on $\mathcal{S}$, the agent's original scenario. For there is a precise sense in which the new scenario represents a *stronger* set of

reasons than the original: setting aside those defaults shared by the two scenarios, it follows from our definition that each default belonging to the new but not to the original scenario will have a higher priority than any default belonging to the original scenario but not to the new one. This observation depends, of course, on our requirement that the defaults belonging to the defeating set must be uniformly stronger than those belonging to the accommodating set. Without this requirement, it would be hard to draw any meaningful strength comparisons between the new scenario and the original, and so hard to see why conclusions based on the new scenario should be preferred.

And second, since what is most distinctive about our definition of defeat is its appeal to an accommodating set, to be retracted from the agent's current scenario, it is worth focusing on the defaults belonging to this set; how can we justify retracting defaults to which the agent is already committed? As we have already seen, there is no need to justify the retraction of defaults belonging to arbitrary accommodating sets, possibly containing defaults that do not actually need to be retracted in order to accommodate some defeating set. It is enough to limit our attention to defaults from minimal accommodating sets, those whose retraction is necessary; and in this case, there is no real difficulty justifying the retraction of these defaults at all, since it turns out that any default belonging to such a set must itself be defeated.

OBSERVATION 3. *Where S is a scenario based on the ordered default theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$, suppose δ is defeated in S, with $\mathcal{D}'$ as a defeating set and S^* as a minimal accommodating set for $\mathcal{D}'$. Then each default belonging to S^* is likewise defeated in S, with $\mathcal{D}'$ as a defeating set and S^* as a minimal accommodating set for $\mathcal{D}'$.*

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

DEFINITION 4 (BINDING DEFAULTS). Where S is a scenario based on the ordered default theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$, the defaults from $\mathcal{D}$ that are *binding* in S are those belonging to the set

$$Binding_{\mathcal{W}, \mathcal{D}_<}(S) = \{\delta \in \mathcal{D} : \delta \in Triggered_{\mathcal{W}, \mathcal{D}_<}(S),$$

$$\delta \notin Conflicted_{\mathcal{W}, \mathcal{D}_<}(S),$$

$$\delta \notin Defeated_{\mathcal{W}, \mathcal{D}_<}(S)\}.$$

This definition of a binding default is modeled on, and can usefully be compared with, the definition from [12] of an inheritable argument path as one that is constructible, but neither conflicted nor preempted.

3. PROPER SCENARIOS AND EXTENSIONS

3.1. *Definitions*

Since the binding defaults represent those that provide 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.

DEFINITION 5 (STABLE SCENARIOS). Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory and $\mathcal{S}$ a scenario. Then $\mathcal{S}$ is a *stable scenario* based on $\langle \mathcal{W}, \mathcal{D}_< \rangle$ just in case $\mathcal{S} = Binding_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$.

An agent who 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.

Our goal, we recall, is to characterize the proper scenarios – those that an ideal reasoner could come to accept as an appropriate basis for its beliefs, when provided with some default theory as initial information. Can we, then, simply identify the proper scenarios with the stable scenarios? The answer is No, as we can see from the following example. Let δ_1 be the default $A \rightarrow A$, and consider the theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$ in which $\mathcal{W} = \emptyset$, $\mathcal{D} = \{\delta_1\}$, and $<$ is empty. Here, the set $\mathcal{S}_1 = \{\delta_1\}$ is a stable scenario based on this theory, since the single default δ_1 is triggered in the context of this scenario, but neither conflicted nor defeated. But $\mathcal{S}_1$ should not be classified as a proper scenario. The best way to see this is to note that $Th(\{A\})$, the belief set generated by this scenario, contains the formula A. But we would not want the agent to accept this formula, since it is not, in an intuitive sense, *grounded* in the agent's initial information. We will return shortly to consider this concept of groundedness in more detail.

As this example shows, a stable scenario can generate too much information, but perhaps there is a simple solution to the problem. Even though, in the example, $\mathcal{S}_1$ is a stable scenario, it is not a *minimal* stable scenario. The only minimal stable scenario based on the agent's initial

information is $\mathcal{S}_0 = \emptyset$, generating the belief set $Th(\emptyset)$, which does seem to be appropriate. Is it possible, then, to identify the proper scenarios with the minimal stable scenarios?

No again. Let δ_1 be the default $A \to A$, let δ_2 be $\top \to \neg A$, and consider the theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$ in which $\mathcal{W} = \emptyset$, $\mathcal{D} = \{\delta_1, \delta_2\}$, and $<$ is empty. Here, $\mathcal{S}_1 = \{\delta_1\}$ is again a stable scenario, containing exactly the defaults that are binding in this scenario; the default δ_2 is not binding, since it is conflicted. In this case, however, the scenario $\mathcal{S}_0 = \emptyset$ is not stable, since the default δ_2 is binding in the context of this scenario, but not included. It follows that $\mathcal{S}_1$ is not only a stable scenario, but a minimal stable scenario. But again, we would not want to classify $\mathcal{S}_1$ as proper; the only proper scenario, in this case, is $\mathcal{S}_3 = \{\delta_2\}$, which generates the belief set $Th(\{\neg A\})$.

Rather than attempting to define the proper scenarios in terms of the notion of stability, then, we will adapt a quasi-inductive construction of the kind employed by Reiter. We begin by introducing the notion of an approximating sequence.

DEFINITION 6 (APPROXIMATING SEQUENCES). Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory and $\mathcal{S}$ a scenario. Then $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \ldots$ is an *approximating sequence* that is *based on* the theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$ and *constrained by* the scenario $\mathcal{S}$ just in case

$$\mathcal{S}_0 = \emptyset,$$

$$\mathcal{S}_{i+1} = \{\delta : \delta \in \mathit{Triggered}_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S}_i),$$

$$\delta \notin \mathit{Conflicted}_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S}),$$

$$\delta \notin \mathit{Defeated}_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})\}.$$

An approximating sequence is supposed to provide an abstract representation of the reasoning process carried out by an ideal agent in arriving at some scenario, a set of acceptable defaults. The sequence depends on two parameters: a base default theory representing the agent's initial information, and a constraining scenario against which it checks defaults for conflict or defeat. The agent begins its reasoning process, at the initial stage $\mathcal{S}_0$, without having accepted any defaults; and then, at each successive stage $\mathcal{S}_{i+1}$, it supplements its current stock of defaults with those that have been triggered at the previous stage $\mathcal{S}_i$, as long as they are neither conflicted nor defeated in the constraining set $\mathcal{S}$. It is easy to see that the scenarios belonging to an approximating

sequence are nested, each a subset of the next, so that the sequence really can be thought of as providing better and better approximations of some end result. The *limit* of an approximating sequence – defined as $\bigcup_{i \geq 0} S_i$ – represents this end result, the scenario that the agent will arrive at after carrying out the reasoning process indefinitely.

We are particularly interested in the special case of an approximating sequence that is *constrained by its own limit* – a sequence, that is, representing a reasoning process in which defaults are evaluated for conflict or defeat with respect to the scenario that the agent will eventually arrive at after carrying out that very process. A proper scenario can be defined as the limit of an approximating sequence like this.

DEFINITION 7 (PROPER SCENARIOS). Where $\langle \mathcal{W}, \mathcal{D}_< \rangle$ is an ordered default theory and $\mathcal{S}$ is a scenario, let $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \ldots$ be an approximating sequence based on $\langle \mathcal{W}, \mathcal{D}_< \rangle$ and constrained by $\mathcal{S}$. Then $\mathcal{S}$ is a *proper scenario* based on $\langle \mathcal{W}, \mathcal{D}_< \rangle$ just in case $\mathcal{S} = \bigcup_{i \geq 0} \mathcal{S}_i$.

Having introduced this notion, we can now, as suggested earlier, define an extension of a default theory as a belief set that is generated by a proper scenario.

DEFINITIONS 8 (EXTENSIONS). Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory and $\mathcal{E}$ a set of formulas. Then $\mathcal{E}$ is an *extension* of $\langle \mathcal{W}, \mathcal{D}_< \rangle$ just in case $\mathcal{E} = Th(\mathcal{W} \cup Conclusion(\mathcal{S}))$ where $\mathcal{S}$ is a proper scenario based on this default theory.

The concept can be illustrated by considering the Tweety Triangle from the previous section. As the reader can verify, the unique proper scenario based on this default theory is $\mathcal{S}_2 = \{\delta_2\}$, where δ_2 is $P \rightarrow \neg F$, so that $Conclusion(\mathcal{S}_2) = \{\neg F\}$. The ordinary information contained in the theory is $\mathcal{W} = \{P, B\}$. The extension of the theory, therefore, is $\mathcal{E} = Th(\{P, B, \neg F\})$.

Let us return, now, to the concept of groundedness. As we have seen, there are stable scenarios, and even minimal stable scenarios, that are not proper. But it is easy to verify that each proper scenario is, in fact, stable.

THEOREM 1. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory and $\mathcal{S}$ a proper scenario based on this theory. Then $\mathcal{S}$ is also stable.*

What the concept of a proper scenario adds to that of a stable scenario, from an intuitive standpoint, is simply the requirement that the set of defaults accepted by an agent must be properly grounded in the

agent's initial information. A default can belong to a proper scenario only if it belongs to some scenario from the approximating sequence, and it can belong to such a scenario only if it is triggered in the empty scenario, or in some other scenario that occurs earlier in the sequence.

Membership in an approximating sequence guarantees groundedness by ensuring that the conclusion of a default rule cannot be appealed to until its premise is actually established – by, in effect, treating a default as a rule of inference. One way of arriving at a firm understanding of this concept of groundedness, therefore, is to make the identification between defaults and rules of inference explicit. We begin by extending the ordinary notion of a propositional proof to take account of these new rules.

DEFINITIONS 9 ($\mathcal{S}$-PROOFS). Where $\mathcal{S}$ is a set of defaults and $\mathcal{W}$ is a set of formulas, an $\mathcal{S}$-proof of A *from* $\mathcal{W}$ is a sequence $A_1, A_2, \ldots, A_n$ such that A_n is A and, for $j \leq n$, each A_j satisfies one of the following conditions: (1) A_j is an axiom of propositional logic; (2) A_j belongs to $\mathcal{W}$; (3) A_j follows from previous members of the sequence by modus ponens; (4) there is some $\delta \in \mathcal{S}$ such that *Conclusion*(δ) is A_j and *Premise*(δ) is a previous member of the sequence.

Evidently, an $\mathcal{S}$-proof is just like an ordinary propositional proof, except that it allows each default belonging to the set $\mathcal{S}$ to function as an additional rule of inference, justifying the placement of its conclusion in a proof sequence once its premise has been established. We let $Th^{\mathcal{S}}(\mathcal{W})$ denote the set of formulas that have $\mathcal{S}$-proofs from $\mathcal{W}$.[2]

Using this, notion, we can now explicate the concept of groundedness by stipulating that a scenario $\mathcal{S}$ is grounded in the agent's ordinary information $\mathcal{W}$ just in case the belief set generated by $\mathcal{S}$ contains only statements that have $\mathcal{S}$-proofs from $\mathcal{W}$.

DEFINITION 10 (GROUNDED SCENARIOS). Let $\mathcal{W}$ be a set of formulas and $\mathcal{S}$ a scenario. Then $\mathcal{S}$ is *grounded* in $\mathcal{W}$ just in case $Th(\mathcal{W} \cup Conclusion(\mathcal{S})) \subseteq Th^{\mathcal{S}}(\mathcal{W})$.

The concept can be illustrated by returning to the theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$ described at the beginning of this section, with $\mathcal{D} = \{\delta_1\}$ where δ_1 is the default $A \to A$, and with $\mathcal{W}$ and $<$ empty. We noted earlier that the scenario $\mathcal{S}_1 = \{\delta_1\}$ is not, in an intuitive sense, grounded in the agent's initial information; and this intuition can now be confirmed by appeal to our formal definition, since $Th(\mathcal{W} \cup Conclusion(\mathcal{S}_1)) = Th(\{A\})$ but $Th^{\mathcal{S}_1}(\mathcal{W}) = \emptyset$.

With this definition of groundedness in place, it can now be verified that the proper scenarios are grounded.

THEOREM 2. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory and $\mathcal{S}$ a proper scenario based on this theory. Then $\mathcal{S}$ is also grounded in $\mathcal{W}$.*

Together with the preceding Theorem, this result tells us that the proper scenarios are both stable and grounded. And indeed, the other direction can be established as well, leading to an alternative characterization of the proper scenarios as the stable, grounded scenarios.

THEOREM 3. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory. Then $\mathcal{S}$ is a proper scenario based on this theory if and only if $\mathcal{S}$ is both a stable scenario based on $\langle \mathcal{W}, \mathcal{D}_< \rangle$ and also grounded in $\mathcal{W}$.*

3.2. *Remarks*

It is easy to see that the extension of an ordered default theory must be consistent as long as the set of ordinary formulas from that theory is consistent: defaults alone cannot introduce inconsistency.

THEOREM 4. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory with $\mathcal{E}$ as an extension. Then $\mathcal{E}$ is consistent if and only if $\mathcal{W}$ is consistent.*

As with Reiter's default logic, however, the account presented here defines a relation between ordered default theories and their extensions that may seem anomalous from a more conventional logical perspective. Certain default theories may have multiple extensions, and others may have no extensions at all.

The canonical example of a default theory with multiple extensions is the earlier Nixon Diamond. This extension supports two proper scenarios, both the scenario $\mathcal{S}_1 = \{\delta_1\}$, considered earlier, and $\mathcal{S}_2 = \{\delta_2\}$, where δ_1 is $Q \to P$ and δ_2 is $R \to \neg P$. Since the ordinary information contained in this default theory is $\mathcal{W} = \{Q, R\}$, these two scenarios generate the two extensions $\mathcal{E}_1 = Th(\{Q, R, P\})$ and $\mathcal{E}_2 = Th(\{Q, R, \neg P\})$. In light of these two extensions, one of which contains P and the other $\neg P$, what is the agent supposed to conclude from the original theory: is Nixon a pacifist or not? More generally, when an ordered default theory leads to more than one extension, how should we define its consequences?

The question is vexed, and several proposals have been discussed in the literature. 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 $\mathcal{S}_2$, 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.[3] 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.[4]

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 generally described as *skeptical*.[5] 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.[6]

For an example of an ordered default theory with no extensions at all, let δ_1 be the default $\top \to A$ and δ_2 the default $A \to \neg A$, and consider the theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$ in which $\mathcal{W} = \emptyset$, $\mathcal{D} = \{\delta_1, \delta_2\}$, and $<$ orders these two defaults so that $\delta_1 < \delta_2$. By our definition, any extension of this theory would have to be generated by some proper scenario. But we can verify by enumeration that no subset of $\mathcal{D}$ is even a stable scenario, let alone proper: $\mathcal{S}_0 = \emptyset$ is not stable, since δ_1 is binding in the context of this scenario, but not included; $\mathcal{S}_1 = \{\delta_1\}$ is not stable, since it includes δ_1, which is defeated in this context; $\mathcal{S}_2 = \{\delta_2\}$ is not stable, since it contains δ_2, which is not triggered; and $\mathcal{S}_3 = \{\delta_1, \delta_2\}$ is not stable, since both of the defaults it includes are conflicted in the context. Since there is no stable scenario based on this default theory, there can be no proper scenario either, and so the theory has no extension.

There are several ways of responding to the possibility of default theories without extensions, which I will simply mention. One option is to observe that the problem springs, quite generally, from the presence of "vicious cycles" among defaults (compressed, in our simple example, into the single default δ_2), and to argue that such vicious cycles renders a default theory incoherent. It is then natural to attempt to formulate syntactic conditions ruling out vicious cycles, which would guarantee coherence and so the existence of extensions. This line of exploration has a long history in nonmonotonic reasoning, going back to Reiter's proof in [23] that extensions are guaranteed for normal default theories, to Touretzky's proof in [26] that acyclic inheritance networks must have extensions, and to the initial work on stratification in logic programs, such as that of Apt et al. [2]. In the present setting, the goal would be to find the weakest and most plausible syntactic restrictions necessary to guarantee the existence of proper scenarios, and so of extensions, for the ordered default theories defined here.

From a more general perspective, the strategy behind this first option is similar to Tarski's idea of responding to the semantic paradoxes by postulating a stratification of the language, to rule out vicious cycles. It is also possible, by contrast, to explore the idea of allowing vicious cycles among defaults, not imposing any syntactic restrictions at all, and then attempting to modify the present theory so as to allow for extensions even when these cycles are present. An approach along these lines would be similar to more recent work on the semantic paradoxes, and might well use tools developed in this work.[7]

Finally, again returning to the view that theories without extensions may be incoherent, it may be possible simply to live with these theories if one happens to favor the notion of skeptical consequence. According to this view, as we have seen, the consequences of a theory are the formulas that belong to each of its extensions. Since an incoherent theory has no extensions, any formula at all would lie in their intersection, and so an incoherent theory would have exactly the same set of consequences as an inconsistent theory.

4. Discussion

The problem of reasoning with prioritized defaults, as mentioned earlier, is not a new topic in nonmonotonic reasoning. The previous work in the area has followed several paths. Brewka [4], as well as Baader and Hollunder [3] and Marek and Truszczyński [20], have all explored the possibility of using priorities as control information to guide the process of reasoning with defaults, so that applicable defaults with higher priority must be satisfied before those of lower priority can be considered; this general idea has recently been developed in a more sophisticated form by Brewka and Eiter [6]. Delgrande and Schaub, in [7] and [8], have explored techniques for compiling priority information into ordinary default rules, revitalizing in a much more general and systematic way an idea that was first hinted at by Reiter and Criscuolo [24], and developed in a different direction by Etherington and Reiter [10]. Rintanen [25] explores the idea of ordering extensions on the basis of the defaults generating them, with better defaults leading to better outcomes, and then defining the preferred extensions as those that are maximal in the ordering.

I do not intend to discuss any of this work in detail, particularly since a useful survey and taxonomy of the different theories in the area has recently appeared in Delgrande et al. [9], but I would like to make a few comparative remarks. I begin by considering the theory of normal defaults

from Reiter's original logic, turn next to the idea of using priorities to control the order of application of defaults, and finally, to certain cases in which the current account may appear to yield questionable results. The examples considered in this section will serve, in addition, to distinguish this account from previous accounts in the literature.

4.1. *Normal Default Theories*

The defaults in Reiter's original theory are rules of the form $(A : C \ / \ B)$, with the rough interpretation: if A belongs to the agent's stock of beliefs, and C is consistent with these beliefs, then the agent should believe B as well. A *normal default* is a default rule in which the second and third of these elements match – that is, a rule of the form $(A : B \ / \ B)$, which we can write as $A \rightarrow B$, thus identifying Reiter's normal defaults with the default rules presented here. A *normal default theory*, in Reiter's sense, is a pair of the form $\langle \mathcal{W}, \mathcal{D} \rangle$ in which $\mathcal{W}$ is a set of ordinary formulas and $\mathcal{D}$ is a set of normal default rules. Using the notation of the current paper, the extensions defined by Reiter for these normal default theories – which I will refer to here as Reiter extensions – can be characterized as follows.

DEFINITION 11 (REITER EXTENSIONS). Let $\langle \mathcal{W}, \mathcal{D} \rangle$ be a normal default theory. Then $\mathcal{E}$ is a *Reiter extension* of $\langle \mathcal{W}, \mathcal{D} \rangle$ just in case $\mathcal{E} = \bigcup_{i \geq 0} \mathcal{E}_i$, where the sequence $\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}_2, \dots$ is defined as

$$\mathcal{E}_0 = \mathcal{W},$$

$$\mathcal{E}_{i+1} = Th(\mathcal{E}_i) \cup Conclusion(\{\delta \in \mathcal{D} : \mathcal{E}_i \vdash Premise(\delta),$$
$$\mathcal{E} \nvdash \neg Conclusion(\delta)\}).$$

Let us say that the normal default theory $\langle \mathcal{W}, \mathcal{D} \rangle$ *corresponds* to any ordered default theory of the form $\langle \mathcal{W}, \mathcal{D}_< \rangle$, sharing the same set $\mathcal{W}$ of ordinary formulas and the same set $\mathcal{D}$ of defaults. The normal default theory corresponding to an ordered default theory is arrived at, then, simply by removing all priority information from the ordered theory. What is the relation between the extensions of an ordered default theory, as defined here, and the Reiter extensions of its corresponding normal default theory?

The first thing to note is that the current account is a conservative generalization of Reiter's account, in the sense that the extensions of an ordered default theory without any real ordering information coincide with those of the corresponding normal default theory.

THEOREM 5. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory in which the ordering $<$ is empty. Then $\mathcal{E}$ is an extension of $\langle \mathcal{W}, \mathcal{D}_< \rangle$ if and only if $\mathcal{E}$ is a Reiter extension of $\langle \mathcal{W}, \mathcal{D} \rangle$, the corresponding normal default theory.*

But what about the more general case, when the ordering information from an ordered default theory is not empty? It is often assumed that the extensions of ordered, or prioritized, default theories should form a subset of the Reiter extensions of the corresponding normal default theories.[8] In fact, this relation does not hold in general for the current account, as we can see by considering the ordered theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$ defined as follows: $\mathcal{W}$ is empty; $\mathcal{D}$ contains an infinite number of defaults, where each default δ_i has the form $\top \rightarrow A$ when i is an odd integer and the form $\top \rightarrow \neg A$ when i is an even integer; and the defaults are ordered so that $\delta_i < \delta_j$ whenever $i < j$. The normal default theory $\langle \mathcal{W}, \mathcal{D} \rangle$ corresponding to this ordered theory allows just two Reiter extensions: $\mathcal{E}_1 = Th(\{A\})$ and $\mathcal{E}_2 = Th(\{\neg A\})$. But there are three proper scenarios based on the ordered theory itself: both the scenarios $\mathcal{S}_1 = \{\delta_i : i \text{ is odd}\}$ and $\mathcal{S}_2 = \{\delta_i : i \text{ is even}\}$, which generate the extensions $\mathcal{E}_1$ and $\mathcal{E}_2$ above, but also the scenario $\mathcal{S}_0 = \emptyset$, generating the extension $\mathcal{E}_0 = Th(\emptyset)$, which is not a Reiter extension of the corresponding normal theory.

Still, even though it does not hold in general that the extensions of ordered default theories form a subset of the Reiter extensions of the corresponding normal default theories, this relation can be established for certain well-behaved ordered default theories, and particularly, for those that contain only a finite set of defaults. The verification of this result relies on three initial observations, which have some interest on their own. The first, which holds of ordered default theories in general, not just finite theories, is that, whenever a default is defeated in the context of a stable scenario, the defeating set for that default must be consistent with the scenario. The second is that, in the special case of finite theories, any set that defeats a default in the context of a stable scenario must already be contained within that scenario. And the third, also restricted to finite theories, is that any default that is defeated in the context of a stable scenario must be conflicted in that context as well.

OBSERVATION 4. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory, and suppose $\mathcal{S}$ is a stable scenario based on this theory. Then if some default δ is defeated in $\mathcal{S}$, with $\mathcal{D}'$ as a defeating set, it follows that $Conclusion(\mathcal{S} \cup \mathcal{D}')$ is consistent.*

OBSERVATION 5. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory in which the set $\mathcal{D}$ of defaults is finite, and suppose $\mathcal{S}$ is a stable scenario based on this theory. Then if some default δ is defeated in $\mathcal{S}$, with $\mathcal{D}'$ as a defeating set, it follows that $\mathcal{D}' \subseteq \mathcal{S}$.*

OBSERVATION 6. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory in which the set $\mathcal{D}$ of defaults is finite, and suppose $\mathcal{S}$ is a stable scenario based on this theory. Then any default that is defeated in $\mathcal{S}$ must also be conflicted in $\mathcal{S}$.*

With these observations in place, we can now establish that, at least in the case of ordered default theories containing only a finite number of defaults, each extension must also be a Reiter extension of the corresponding normal default theory.

THEOREM 6. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory in which the set $\mathcal{D}$ of defaults is finite. Then if $\mathcal{E}$ is an extension of $\langle \mathcal{W}, \mathcal{D}_< \rangle$, it follows that $\mathcal{E}$ is also a Reiter extension of $\langle \mathcal{W}, \mathcal{D} \rangle$, the corresponding normal default theory.*

The question remains, however, whether this is a desirable result: do we want the extensions of default theories with priorities to be limited to a subset of the corresponding Reiter extensions?

There are certain examples suggesting that this result might be problematic. To understand a simple one, imagine that a naturalist studying the distribution of birds among a remote chain of islands has identified two new kinds of finches.[9] There is, first of all, the species of Ruffed Finches, whose nests are largely though not entirely confined to Green Island; and second, there is a particular subspecies of the Ruffed Finches, known as the Least Ruffed Finches, whose nests are distributed almost evenly between Green Island and Sand Island, with only a few strays found elsewhere. Now consider a particular individual, Frank, who happens to be a Least Ruffed Finch. What should the naturalist conclude, by default, about the location of Frank's nest?

The example can be coded formally by letting R, L, G, and S represent the respective propositions that Frank is a Ruffed Finch, that he is a Least Ruffed Finch, that his nest is on Green Island, and that his nest is on Sand Island. If we then suppose that δ_1 is the default $R \to G$ and δ_2 the default $L \to G \vee S$, instances of the generalizations that Ruffed Finches live on Green Island while Least Ruffed Finches are distributed between Green and Sand Islands, the relevant information can then be captured by the ordered default theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$ in which $\mathcal{W} = \{L, L \supset R\}$, $\mathcal{D} = \{\delta_1, \delta_2\}$, and $\delta_1 < \delta_2$. The unique proper scenario

based on this theory is $\mathcal{S}_1 = \{\delta_1, \delta_2\}$, generating the extension $\mathcal{E}_1 = Th(\mathcal{W} \cup \{G, G \vee S\})$, which is also the unique Reiter extension of the corresponding normal default theory. This extension supports the conclusion G, that Frank's nest is on Green Island, since he is a Ruffed Finch. But that does not seem like the right conclusion at all. From an intuitive standpoint, it seems that the naturalist should conclude only $G \vee S$, that Frank's nest is on either Green or Sand Island, because he is a Least Ruffed Finch; the more desirable extension therefore seems to be $\mathcal{E}_2 = Th(\mathcal{W} \cup \{G \vee S\})$, which is not an extension of the corresponding normal default theory.

I had previously thought that examples like this showed that any theory of prioritized default reasoning, such as the present theory, that returns only a subset of the corresponding Reiter extensions must be in error – since in this case, for instance, the intuitively correct $\mathcal{E}_2$ is not a Reiter extension of the corresponding normal default theory.[10] I now believe, however, that these examples do not suggest that the present theory needs to be modified, but only that it needs to be supplemented: we need a way of saying that certain defaults, even though they may be triggered, are not actually applicable to certain individuals, or classes of individuals. In this particular case, for example, it is not as if δ_1 is defeated by δ_2 in the sense of defeat defined here, since there is no conflict between their conclusions: living on Green Island is certainly consistent with living on either Green Island or Sand Island. Instead, it seems that the default δ_1 is simply not applicable to Least Ruffed Finches. This idea – that the applicability of defaults must be suspended in certain cases – has been studied extensively by Pollock, particularly in his [22], as a special kind of defeat, which he describes as "undercutting" defeat. In my own [16], I show how the idea can instead be incorporated into the present framework once it has been supplemented with the capability of reasoning about rule priorities.

4.2. *Controlling Order of Application*

A number of researchers, as mentioned, have explored the idea of using information about the relative priority of defaults to control their order of application. These various approaches differ in detail, but they fit a common pattern. Extensions are viewed as being constructed in a series of stages, with the defaults that are *active* at a stage defined as those that are triggered at that stage, whose conclusions have not yet been accepted, but which are not yet conflicted either.[11] At each stage, then, one of the most important active defaults is selected, and its conclusion is added to the agent's belief set. The set is then closed under

consequence, and the agent moves on to the next stage, continuing in this way until a fixed point is reached, possibly in the limit.

The idea behind this construction is that any conflict among a group of defaults is always resolved in favor of the more important members of the group, since these defaults are applied first. Once the conclusions of the more important defaults are placed in the agent's belief set, the less important defaults are then conflicted; they are therefore no longer active, and cannot be applied.

This basic idea is simple and attractive, and provides the correct results in several central cases, but it has been called into question because of its results in a series of abstract examples. A representative example, which I refer to as the Order Puzzle, is the theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$ in which $\mathcal{W} = \{W\}$ and $\mathcal{D} = \{\delta_1, \delta_3, \delta_3\}$, where δ_1 is $W \rightarrow H$, δ_2 is $W \rightarrow \neg O$, and δ_3 is $H \rightarrow O$, and in which the ordering places $\delta_1 < \delta_2$ and $\delta_2 < \delta_3$. It is easy to see that the order of application approach assigns to this theory the unique extension $\mathcal{E}_1 = Th(\mathcal{W} \cup \{H, \neg O\})$. At the first stage, only δ_1 and δ_2 are active; δ_3 is not yet triggered. Since δ_2 has higher priority, it must be applied, yielding $\neg O$. At the second stage, δ_1 alone is active, since δ_2 has already been applied and δ_3 is not yet triggered. It is therefore applied, yielding H. By the third stage, then, there are no longer any active defaults, since δ_1 and δ_2 have already been applied, and δ_3, although now triggered, is conflicted by the previous application of δ_2.

This particular example has a curious history. It was first noted by Brewka [4], who argued that the extension $\mathcal{E}_1$ is correct. Later, Brewka and Eiter [6] rejected $\mathcal{E}_1$ in favor of the extension $\mathcal{E}_2 = Th(\mathcal{W} \cup \{H, O\})$. This is also, as it happens, the unique extension generated by the present approach: the scenario $\mathcal{S}_2 = \{\delta_1, \delta_3\}$, which generates $\mathcal{E}_2$, is proper; but $\mathcal{S}_1 = \{\delta_1, \delta_2\}$, which generates $\mathcal{E}_1$, is not even stable, since δ_2 is defeated in that context by δ_3.

The argument advanced by Brewka and Eiter against $\mathcal{E}_1$ as an extension runs, roughly, along the following lines: since the original theory assigns δ_3 a higher priority than δ_2, any approach that takes priority seriously must prefer an extension generated by a scenario containing δ_3 to one generated by a scenario that is otherwise identical except that it contains δ_2 instead.[12] The belief sets $\mathcal{E}_1$ and $\mathcal{E}_2$ are generated by the scenarios $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively, which are otherwise identical except that the first contains δ_3 while the second contains δ_2. Therefore, $\mathcal{E}_2$ must be preferred over $\mathcal{E}_1$, so that $\mathcal{E}_1$ cannot lie among the most preferred belief sets.

Even if one accepts this argument, however, all it actually shows is that $\mathcal{E}_1$ should not be classified as an extension of the original theory, not

that $\mathcal{E}_2$ should be, which leaves open a third possibility: perhaps the theory is incoherent, and has no extensions at all. This possibility is embraced by Delgrande and Schaub, who argue in [7] that the Order Puzzle itself is meaningless, since the priority ranking of its defaults does not correspond to the order in which the propositions at work in this example would naturally be established, and who claim, therefore, that this theory is incoherent, and should have no extensions.[13]

So what is the correct result in the case of the Order Puzzle? Is it $\mathcal{E}_1$, as the order of application approach would suggest? Is it $\mathcal{E}_2$, as suggested by Brewka and Eiter, and by the present approach? Or is it better to conclude with Delgrande and Schaub that the original theory is incoherent, and that it has no extensions at all? The problems presented by the Order Puzzle are problems of coherence and interpretation. To establish that this theory is even coherent, we need to find a sensible interpretation, suggesting that the theory should actually have an extension; the interpretation will then support the present approach only if the extension it suggests is our $\mathcal{E}_2$.

How could we construct such an interpretation? We cannot appeal to the idea that default priority tracks specificity, as Delgrande and Schaub note; on any view of specificity, the default δ_1 would provide more specific information than δ_3, yet in this case, δ_3 is assigned the higher priority. A reliability interpretation is possible, with each default indicating something like a high conditional probability that its conclusion is satisfied, given that its premise is satisfied, and with the priority ordering measuring relative strength of these conditional probabilities. But notice that the extension naturally suggested by such an interpretation is actually $\mathcal{E}_1$, rather than $\mathcal{E}_2$. For the default δ_2 then tells us that $\neg O$ follows with high probability, given that W holds. And the potential competing argument has no force, since δ_1 already supports H given W less strongly than δ_2 supports $\neg O$ given W. As a result, even if the conditional support provided by δ_3 for O given H is arbitrarily strong, it still follows that the conditional probability of O given W will be less than that of $\neg O$ given W.

Is there, then, an interpretation of the Order Puzzle that is intuitively coherent and also supports $\mathcal{E}_2$ as an extension? I believe there is, but the interpretation I supply is not another epistemic interpretation, in which defaults are taken to extend belief sets. It is, instead, a normative interpretation of the kind explored in my [11] and [15]. Each default of the form $A \rightarrow B$ is taken to represent a *conditional command*, or imperative, enjoining the agent to guarantee the truth of B in any situation in which A holds, and the priorities among defaults are taken to represent the levels of authority associated with these various commands.

In a situation in which A holds, the conditional command $A \rightarrow B$ is said to be *obeyed* if the truth of B is guaranteed, and *disobeyed* otherwise. And in selecting a proper scenario, the agent should now be viewed as choosing an appropriate set of commands to obey, rather than an appropriate basis for its belief set; an extension represents the result of obeying these commands.

Suppose, then, that the agent is Corporal O'Reilly, and that he is subject to the commands of three superior officers: a Captain, a Major, and a Colonel. The Captain, who does not like to be cold, issues a standing order that, during the winter, the heat should be turned on. The Major, who is concerned about energy conservation, issues an order that, during the winter, the window should not be opened. And the Colonel, who does not like to be too warm and does not care about energy conservation, issues an order that, whenever the heat is on, the window should be opened. If we let W, H, and O stand, respectively, for the propositions that it is winter, the heat is turned on, and the window is open, then the defaults δ_1, δ_2, and δ_3 can be taken to represent the respective commands issued by the Captain, the Major, and the Colonel. And since the Colonel outranks the Major, and the Major outranks the Captain, we have the desired priority ordering: $\delta_1 < \delta_2$ and $\delta_2 < \delta_3$. Finally, suppose it is winter. The situation is then exactly as depicted in the Order Puzzle.

Although there are many things wrong with this set of commands (the Colonel's order is especially odd), I hope we can agree that it is at least coherent, in the sense that O'Reilly might, in fact, be subject to a set of commands like these. A thinking soldier could perhaps grasp the intentions behind the various imperatives and arrive at a plan of action that would satisfy all three officers. But it is not O'Reilly's job to think, or to help the officers express their intentions more effectively by issuing more subtle or carefully qualified commands. O'Reilly's job is to obey his orders exactly as they have been issued. If he fails to obey an order issued by an officer without an acceptable excuse, he will be court-martialed. And, let us suppose, there is only one *acceptable excuse* for disobeying such an order: that, under the circumstances, he is prevented from obeying the order issued by this officer by having chosen to obey another order or set of orders issued by other officers of equal or higher rank. Again, some of us may feel that there must be more to the concept of an acceptable excuse than this, but I hope we can agree that the present notion is at least coherent, in the sense that this narrow concept of an excuse may actually be the one at work in some normative system.

Under the current interpretation, a scenario is supposed to represent an appropriate selection of commands to obey, a way of responding

appropriately to the imperatives contained in that theory, where, in this particular case, an appropriate response is one that allows the agent to avoid court martial. So given the set of commands that O'Reilly has been issued, can he, in fact, avoid court martial? Yes, he can, by choosing the scenario $\mathcal{S}_2 = \{\delta_1, \delta_3\}$, obeying the orders issued by the Captain and the Colonel, thus guaranteeing H and O, and so the extension $\mathcal{E}_2$. In this scenario, O'Reilly fails to obey the Major's order, the default δ_2, but he has an excuse: he was prevented from doing so by obeying an order issued by the Colonel, an officer of higher rank. What if O'Reilly were instead to select the scenario $\mathcal{S}_1 = \{\delta_1, \delta_2\}$, guaranteeing H and $\neg O$, and so the extension $\mathcal{E}_1$? In that case, he would obey the Captain and the Major, but fail to obey the Colonel, and he would do so, furthermore, without an acceptable excuse: although O'Reilly is prevented from obeying the Colonel by complying with an order issued by the Major, that is no excuse, since the Colonel outranks the Major.

What this normative interpretation offers, then, is an intuitive way of understanding why $\mathcal{E}_2$, but not $\mathcal{E}_1$, should be classified as an extension of the Order Puzzle: the scenario $\mathcal{S}_2$, which generates the extension $\mathcal{E}_2$, allows O'Reilly to avoid court martial, while the scenario $\mathcal{S}_1$, which generates $\mathcal{E}_1$, does not.

Are there, however, any other options to consider, apart from the two scenarios $\mathcal{S}_1$ and $\mathcal{S}_2$? Well, it may seem that O'Reilly could reason in the following way.[14] If he obeys the Captain's command δ_1 to turn the heater on, then he will find himself in a situation in which he has no choice but to disobey either the Colonel's command δ_3 to open the window or the Major's command δ_2 to keep the window closed. Both the Colonel and the Major outrank the Captain. Therefore, it is best to disobey the Captain's command in order to avoid to being placed in a situation in which he is then forced to disobey one or the other of two higher-ranking officers. But of course, if he does disobey the Captain's command δ_1, and the heater is left off, there can then be no possible justification for failing to obey the Major's command δ_2, to keep the window closed.

This line of reasoning seems to suggest the scenario $\mathcal{S}_3 = \{\delta_2\}$, generating the extension $\mathcal{E}_3 = Th(\mathcal{W} \cup \{\neg O\})$. Is $\mathcal{S}_3$, then, a desirable scenario; is $\mathcal{E}_3$ a desirable extension? Not according to the current theory, since this scenario fails to contain the default δ_1, representing the Captain's command, which is triggered in the context but neither conflicted nor defeated; the scenario is, therefore, not stable. Nor is this scenario one that allows O'Reilly to avoid court martial. According to the notion at work here, we recall, an agent has an acceptable excuse for

disobeying some command issued by an officer only if, under the circumstances, the agent is prevented from doing so by obeying other commands issued by officers of equal or higher rank. And in this sense, O'Reilly has no excuse at all for failing to obey the Captain's command; the Captain has ordered him to turn on the heater, and he is not prevented from doing so by obeying the commands of any other officers at all, let alone officers of equal or higher rank.

Of course, in an effort to justify his actions, O'Reilly might advance an argument along the lines set out above, explaining how obeying the Captain would inevitably have led to disobeying either the Major or the Colonel. The argument is interesting, and it would be interesting to try to develop a version of prioritized default logic that allowed this form of hypothetical reasoning (no current theory does). It is clear that any such development would have to involve, on the formal side, entirely different ideas of conflict and defeat, and also that our informal interpretation would have to reflect a much more liberal conception of what counts as an excuse. On the current account, as we have seen, O'Reilly can excuse his actions under a particular scenario only by referring to what he did or did not do in the situation determined by that very scenario. A more liberal notion would have to allow him to excuse his actions by appealing to the choices he would have faced, and the actions he would have been forced to perform, in various hypothetical situations – including, in this particular case, the situation in which he had chosen to obey the Captain's command.

There is one further complication worth noting, both because it highlights the ability the present system to capture an important ambiguity, and also because it may be – I am not certain – that this ambiguity plays some role in accounting for the attractions of the form of hypothetical reasoning just discussed. Suppose that what the Colonel actually says in issuing his command is: "If the heater is on, the window should be open." This statement could naturally be interpreted as a conditional command, along the lines of "If the heater is on, you have an order to open the window," formalized here through our δ_3, the default $H \to O$. But it is also possible to interpret the same statement, not as a conditional command, but as an unconditional, or categorical, command whose content happens to be a conditional, along the lines of "You have an order to open the window if the heater is on." On this latter interpretation, the Colonel's command could best be represented, not through δ_3, but through δ_4, the new default $\top \to (H \supset O)$.

Now imagine that the Colonel's order is interpreted in this way, as a command of a conditional, rather than a conditional command. Imagine,

that is, that δ_4 is substituted for δ_3 in our original description of the order puzzle, so that $\mathcal{W}$ remains unchanged, but the set $\mathcal{D}$ now contains δ_1, δ_2, and δ_4, with $\delta_1 < \delta_2$ and $\delta_2 < \delta_4$. In that case, the unique proper scenario associated with this default theory would be $\mathcal{S}_4 = \{\delta_2, \delta_4\}$, generating the extension $\mathcal{E}_4 = Th(\mathcal{W} \cup \{\neg O, H \supset O\})$, which of course contains the statement $\neg H$.

The two interpretations of the Colonel's command, then, lead to strikingly different results. If the Colonel is interpreted as issuing a conditional command, then, as we have seen, what O'Reilly ought to do is obey the Colonel and the Captain, turning the heater on and opening the window, while disobeying the Major's command to keep the window closed. If the Colonel is interpreted as commanding a conditional, then what O'Reilly ought to do is obey the Colonel and the Major, keeping the window closed but making sure the heater is off, while disobeying the Captain's command to turn the heater on. In both the scenario $\mathcal{S}_4$ associated with the latter interpretation and the scenario $\mathcal{S}_3$, suggested by the process of hypothetical reasoning, O'Reilly obeys the Major and does not necessarily obey the Captain; and it may be – though again, I am not sure – that $\mathcal{S}_3$ gains whatever plausibility is has simply by running together the two distinct ways of interpreting the Colonel's order, as a conditional command or a command of a conditional.

4.3. *Some Difficult Cases*

Having studied a number of situations in which the current theory seems to yield desirable results, or at least results for which some justification can be found, I want to conclude by considering two kinds of situations that raise more difficult issues.

The first can be illustrated with the theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$ in which $\mathcal{W} = \{\neg(A \wedge B)\}$ and $\mathcal{D} = \{\delta_1, \delta_2, \delta_3\}$, where δ_1 is $\top \rightarrow A$, δ_2 is $\top \rightarrow B$, and δ_3 is $A \rightarrow \neg B$, and in which $\delta_1 < \delta_2$ and $\delta_2 < \delta_3$. Again, this theory can usefully be interpreted as a set of commands issued to O'Reilly by the officers, where δ_1 represents the Captain's command to see to it that A, δ_2 represents the Major's command to see to it that B, and δ_3 represents the Colonel's command, conditional on the truth of A, to see to it that $\neg B$. Once more, the Colonel's command is peculiar, since the background information from $\mathcal{W}$ already tells us that A and B are incompatible, but there is nothing to prevent the Colonel from issuing a peculiar command.

On the present approach, this theory supports two proper scenarios. The first is the entirely reasonable $\mathcal{S}_1 = \{\delta_2\}$, generating the extension

$\mathcal{E}_1 = Th(\mathcal{W} \cup \{B\})$. On this scenario, O'Reilly obeys the Major's command δ_2; he disobeys the Captain's command δ_1, but has an excuse, since he is prevented from obeying the Captain by obeying the Major, who outranks the Captain. The Colonel's command δ_3 does not come into play, since it is conditional on the truth of A. There is also, however, a second scenario $\mathcal{S}_2 = \{\delta_1, \delta_3\}$, generating the extension $\mathcal{E}_2 = Th(\mathcal{W} \cup \{A, \neg B\})$. On this scenario, O'Reilly obeys the Captain's command δ_1 and the Colonel's command δ_3; he disobeys the Major's command δ_2, but has an excuse, since he is prevented from obeying the Major by instead obeying the Colonel, who outranks the Major.

Now, although this second scenario $\mathcal{S}_2$ is supported by the theory, and it does, in fact, allow O'Reilly to avoid court martial, there is something anomalous about the scenario all the same. From an intuitive standpoint, it seems almost as if the defaults have been considered in the wrong order. The initial conflict, one wants to say, lies between the Captain's command δ_1 and the Major's command δ_2. This conflict should of course be resolved in favor of the Major, in which case the Colonel's command δ_3 is never even triggered, as in the scenario $\mathcal{S}_1$. In the case of $\mathcal{S}_2$, by contrast, it is as if O'Reilly has made the wrong initial decision, favoring the Captain over the Major, but is absolved from his error by the fact that this incorrect decision triggers the Colonel's command, which provides, in our technical sense, an excuse for his earlier decision to disobey the Major. Once he arrives at the scenario $\mathcal{S}_2$, then, O'Reilly has reached a sort of equilibrium state – the scenario is proper, there is no risk of court martial – but it is not a state he would have arrived at if his reasoning had followed the correct path to begin with.

Let us now turn to another example illustrating the same point, but one that is more disturbing.[15] Since the example is somewhat complicated, we rely on mnemonic abbreviations, focusing on a particular individual, Susan, and letting RC, RN, CC, CU, and VU represent the respective propositions that Susan is a resident of Cuba, a resident of North America, a citizen of Cuba, a citizen of the United States, and a person with voting rights in the United States. We consider the default theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$ in which $\mathcal{W}$ contains the statements RC, $RC \supset RN$, $\neg(CC \wedge CU)$, and $\neg(CC \wedge VU)$; in which $\mathcal{D}$ contains δ_1, δ_2, δ_3, where δ_1 is $RN \rightarrow CU$, δ_2 is $RC \rightarrow CC$, and δ_3 is $RC \rightarrow CC$; and in which the defaults are ordered so that $\delta_1 < \delta_2$ and $\delta_1 < \delta_3$.

The strict information from $\mathcal{W}$ tells us that Susan is a resident of Cuba, and contains instances for Susan of the general facts that residents of Cuba are residents of North America (since Cuba is part of North America), and that citizens of Cuba can neither be citizens of nor have

voting rights in the United States. The set $\mathcal{D}$ contains instances for Susan of three general defaults. First, there is a weak default – with some statistical justification – according to which residents of North America tend to be citizens of the United States. Second, there is a stronger default according to which residents of Cuba tend to be citizens of Cuba. And third, there is a very strong default – stronger than any of the others, and violated only by a few select groups, such as convicted felons – according to which citizens of the United States tend to have voting rights in the United States.

Now, given this information, what are we to conclude about Susan? Well, on the present approach, the theory supports two proper scenarios. The first is $\mathcal{S}_1 = \{\delta_2\}$, generating the extension $\mathcal{E}_1 = Th(\mathcal{W} \cup \{CC\})$, according to which Susan is a citizen of Cuba, rather than the United States, and has no voting rights in the United States; the default δ_1, supporting the proposition that Susan is a citizen of the United States, is defeated by the stronger default δ_2, and the default δ_3, supporting Susan's claim to voting rights, is not triggered. This is, I feel, a reasonable scenario, leading to an intuitively acceptable set of conclusions.

Again, however, there is also a second proper scenario, $\mathcal{S}_2 = \{\delta_1, \delta_3\}$, generating the extension $\mathcal{E}_1 = Th(\mathcal{W} \cup \{CU, VU\})$, according to which Susan is a citizen of the United States, rather than Cuba, and has voting rights; the default δ_2 is now defeated by the stronger default δ_3. This second scenario is less reasonable, and yields conclusions that seem to be clearly incorrect. And again, I would offer the same diagnosis: from an intuitive standpoint, it seems that the defaults are being considered in the wrong order. The initial conflict lies between the defaults δ_1, suggesting that Susan is a citizen of the United States, and δ_2, suggesting that she is a citizen of Cuba. This conflict should be resolved in favor of δ_2, the stronger of the two defaults, in which case δ_3 is not even triggered, as in the reasonable scenario $\mathcal{S}_1$. In the case of the less reasonable $\mathcal{S}_2$, it is as if we have made the wrong initial decision, favoring δ_2 over δ_1, but as a result, the very strong default δ_3 is now triggered, which then defeats δ_1 and provides a sort of justification for the original decision.

What these two examples both illustrate is the need for defining an appropriate order on defaults so that, by considering defaults in that order, we will avoid unintuitive scenarios or extensions, like the scenarios $\mathcal{S}_2$ in each of these theories, and the extensions $\mathcal{E}_2$. This is, as far as I know, an open problem in prioritized default reasoning, and the lack of a solution affects a number of the most promising approaches, as well as this one; for instance, the theory of Brewka and Eiter [6]

generates both the correct extension $\mathcal{E}_1$ and the incorrect $\mathcal{E}_2$ in both of our examples.[16]

The second difficulty I consider raises a different sort of issue, concerning our strength ordering on sets of defaults, according to which one set of defaults $\mathcal{D}'$ is stronger than another set $\mathcal{D}$ just in case $\mathcal{D} < \mathcal{D}'$ – that is, just in case $\delta < \delta'$ for each δ in $\mathcal{D}$ and δ' in $\mathcal{D}'$.

A possible problem for this definition is posed by examples such as $\langle \mathcal{W}, \mathcal{D}_< \rangle$ in which $\mathcal{W}$ is empty and $\mathcal{D} = \{\delta_1, \delta_2, \delta_3, \delta_4\}$, where δ_1 is $\top \to A$, δ_2 is $\top \to \neg A$, δ_3 is $\top \to A$, and δ_4 is $\top \to \neg A$, with the defaults ordered so that $\delta_1 < \delta_2$ and $\delta_3 < \delta_4$. It is useful to think of this theory as representing a set of commands issued to the agent by officials belonging to two separate systems of authority – say, military and ecclesiastical. Let us imagine that δ_1 represents the Captain's command to see to it that A and δ_2 represents the Colonel's command to see to it that $\neg A$, while δ_3 likewise represents the Priest's command to see to it that A and δ_4 represents the Bishop's command to see to it that $\neg A$. The Colonel outranks the Captain and the Bishop outranks the Priest, but the military and ecclesiastical ranks are incomparable.

On the present approach, this theory again supports two perfect scenarios. The first is the reasonable $\mathcal{S}_1 = \{\delta_2, \delta_4\}$, in which the agent obeys the commands of the higher-ranking officials from each of the two systems of authority, the Colonel and the Bishop. The second is the apparently less reasonable $\mathcal{S}_2 = \{\delta_1, \delta_3\}$, in which the agent obeys the lower-ranking officials, the Captain and the Priest.

It is worth pausing at this point to note why $\mathcal{S}_2$ should count even as a stable scenario, let alone perfect. Why is the default δ_1, for instance, not defeated in the context of $\mathcal{S}_2$ by the stronger δ_2, or at least by the defeating set $\mathcal{D}' = \{\delta_2, \delta_4\}$? The reason is that, as we recall from our earlier discussion, a defeating set $\mathcal{D}'$ must be consistent with the set that results when some accommodating set $\mathcal{S}'$ is removed from the current scenario – that is, $\mathcal{S}_2^{\mathcal{D}'/\mathcal{S}'}$ must be consistent – where we require in addition that the defeating set $\mathcal{D}'$ must be stronger than the accommodating set $\mathcal{S}'$. In this case, the only possible accommodating set $\mathcal{S}'$ is, in fact, $\mathcal{S}_2$ itself; and of course, $\mathcal{S}_2^{\mathcal{D}'/\mathcal{S}_2}$ is consistent. But it turns out that $\mathcal{D}'$ is not stronger than $\mathcal{S}_2$ according to our current strength ordering. We do not have $\mathcal{S}_2 < \mathcal{D}'$, since it is not the case that every default from $\mathcal{D}'$ is stronger than every default from $\mathcal{S}_2$; the Colonel's command δ_2 is not stronger than the Priest's command δ_3, and the Bishop's command δ_4 is not stronger than the Captain's command δ_1.

There are two possible reactions to $\mathcal{S}_2$ as a perfect scenario supported by the theory under consideration. It is, first of all, conceivable to

imagine that, although this scenario is apparently less reasonable than $\mathcal{S}_1$, the difficulties are only apparent, and the scenario should indeed be accepted as a legitimate outcome of the theory. Consider, for example, our earlier idea that an agent has an acceptable excuse for disobeying an officer if that agent is forced to do so by obeying other officers of equal or higher rank. This idea works well in the military setting, where the system of ranks forms a total order, but how could it be generalized to apply more broadly? One natural proposal is that an agent should then have an acceptable excuse for disobeying an official if that agent is forced to do so by obeying other officials whose ranks are at least not lower. And in this sense, the agent who adopts the scenario $\mathcal{S}_2$ is, in fact, able to provide acceptable excuses for the neglected commands. The agent is prevented from obeying the Bishop's command δ_4 by instead obeying the command δ_1 issued by the Captain, whose rank is not lower than that of the Bishop; and the agent is prevented from obeying the Colonel's command δ_2 by instead obeying the command δ_3 issued by the Priest, whose rank, again, is not lower than that of the Colonel.

This line of reasoning, of course, supports the current account exactly as it stands, since this account does generate $\mathcal{S}_2$ as an acceptable scenario, along with $\mathcal{S}_1$. Another reaction, however, is simply to reject $\mathcal{S}_2$ as a legitimate outcome. One can imagine the Colonel saying, when δ_3 is offered as an excuse for disobeying δ_2, something along the lines of: "And don't bring up that odd command issued by your Priest – even your Bishop thinks he's wrong."[17] And surely, from an external perspective, it is hard not share the intuition that $\mathcal{S}_1$ is, in some sense, a better scenario than $\mathcal{S}_2$.

What this second reaction suggests is that the current strength ordering on sets of defaults must be modified to reflect this intuition. Our current definition of strength through the $<$ ordering on sets – according to which one set of defaults is stronger than a second only if every member of the first is stronger than every member of the second – is extremely severe. The question is not whether it can be weakened, but which of the various weakenings leads to an acceptable overall theory.

This is, of course, a question that can be answered only after detailed experimentation. But just to hint at the direction in which such a weakening might lead, I display one option that is at least prima facie plausible. Suppose we define a new strength ordering $\prec$ on sets of defaults so that $\mathcal{D} \prec \mathcal{D}'$ just in case: (1) for all δ from $\mathcal{D}$ there is a δ' from $\mathcal{D}'$ such that $\delta < \delta'$; and (2) for all δ' in $\mathcal{D}'$ there is a δ in $\mathcal{D}$ such that $\delta < \delta'$;

and (3) there is no δ from $\mathcal{D}$ and δ' from $\mathcal{D}'$ such that $\delta' < \delta$. Then, returning to the scenarios generated by our example, we can see that $\mathcal{S}_2 \prec \mathcal{S}_1$, as desired; the set of orders issues by the Colonel and the Bishop is preferred to that issued by the Priest and the Captain. And as the reader can verify, with the new $\prec$ relation substituted for the previous $<$ in our definition of defeat, the example now supports only $\mathcal{S}_1$, no longer $\mathcal{S}_2$, as a perfect scenario.

5. CONCLUSION

This paper presents a new approach to prioritized default logic, based on a generalization of previous work on nonmonotonic inheritance reasoning. Although I have not carried out any detailed evaluation, I believe this approach compares favorably to other theories in the area. A number of problems remain to be explored, dividing into three broad categories.

The first category of problems is largely technical. Some of these have already been mentioned. Can we define an appropriately weak notion of stratification, or acyclicity, for which it can then be shown that acyclic ordered default theories must have extensions? Or alternatively, can we generalize the present account, perhaps borrowing techniques from logics of partiality or recent work on the paradoxes, so as to assign natural extensions even to cyclic default theories? Other problems involve defeasible arguments. Although the motivating work on non-monotonic inheritance reasoning defines extensions as sets containing argument paths, the extensions defined here contain only formulas. What would the present account look if extensions were defined as containing defeasible arguments, a generalization of argument paths, rather than formulas? Could we then, as in the theory of nonmonotonic inheritance, define a directly skeptical consequence relation? Finally, there are questions concerning the efficient computation of extensions; and here our focus on scenarios may hold some real benefits, for unlike extensions, which are logically closed belief sets, scenarios will typically be syntactically limited, and in many natural cases, finite.

The second category of problems is more broadly conceptual. As we have seen, the theory developed here still faces certain difficulties, concerning both the order in which defaults are considered and the definition of a preference relation among sets of defaults. I know of no way to address these problems except through experimentation – formulating precise and well-motivated theories of default reasoning, testing them against examples, isolating issues, refining our intuitions.

Finally, there are philosophical issues, centering around the metaphor, frequently appealed to here, of defaults as *reasons*. Can this identification be developed beyond the level of metaphor? Can the simple prioritized default logic presented here be developed into a more robust and general theory of the way in which reasons support conclusions? I believe so, but there is work to be done in a number of areas, of which I mention only three. First, the priorities among defaults are, in this paper, simply taken as given. But one of the things we reason about, and reason about defeasibly, is the priorities among the very defaults that guide our defeasible reasoning. Second, the notion of defeat defined here captures only one form of defeat, sometimes called "rebutting" defeat, in which a stronger reason defeats a weaker reason by contradicting its conclusion. There is at least one other form, sometimes called "undercutting" defeat, in which one reason defeats another, not by contradicting its conclusion, but by undermining its applicability as a reason. And third – an issue perhaps related to the first two – our practical reasoning often seems to involve an appeal to various kinds of "higher-order" reasons, explicitly concerned with the first-order reasons that we should attend to in particular situations. In order for the simple default logic presented here to serve as a basis for a more general theory of practical reasoning, it must be developed to account for these phenomena, and others like them.

Appendix

Proofs of Observations and Theorems

OBSERVATION 1. *Where $\mathcal{S}$ is a scenario based on the ordered default theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$, suppose δ is defeated in $\mathcal{S}$, with $\mathcal{D}'$ as a defeating set and $\mathcal{S}'$ as an accommodating set for $\mathcal{D}'$. Then there is some $\mathcal{S}^* \subseteq \mathcal{S}'$ such that δ is likewise defeated in $\mathcal{S}$ with $\mathcal{D}'$ as a defeating set and $\mathcal{S}^*$ as a minimal accommodating set for $\mathcal{D}'$.*

Proof. Using standard techniques, define $\mathcal{S}''$ as a maximal subset of $\mathcal{S}'$ such that $Conclusion(\mathcal{S}'')$ is consistent with $\mathcal{W} \cup Conclusion((\mathcal{S} - \mathcal{S}') \cup \mathcal{D}')$. Then set $\mathcal{S}^* = \mathcal{S}' - \mathcal{S}''$.   □

OBSERVATION 2. *Where $\mathcal{S}$ is a scenario based on the ordered default theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$, suppose δ is defeated in $\mathcal{S}$, with $\mathcal{D}'$ as a defeating set. Then $\mathcal{S}^* = \emptyset$ is a minimal accommodating set for $\mathcal{D}'$ if and only if $\mathcal{W} \cup Conclusion(\mathcal{S} \cup \mathcal{D}')$ is consistent.*

Proof. First, suppose $\mathcal{S}^* = \emptyset$, where $\mathcal{S}^*$ is a minimal accommodating set for $\mathcal{D}'$. Then since $\mathcal{S}^*$ is an accommodating set for $\mathcal{D}'$, it follows that $\mathcal{W} \cup Conclusion ((\mathcal{S} - \mathcal{S}^*) \cup \mathcal{D}')$ is consistent. So $\mathcal{W} \cup Conclusion (\mathcal{S} \cup \mathcal{D}')$ is consistent, since $\mathcal{S}^* = \emptyset$. Next, suppose $\mathcal{W} \cup Conclusion (\mathcal{S} \cup \mathcal{D}')$ is consistent. Then $\mathcal{S}^* = \emptyset$ is an accommodating set for $\mathcal{D}'$, and so a minimal accommodating set, since it has no proper subsets.   □

OBSERVATION 3. *Where $\mathcal{S}$ is a scenario based on the ordered default theory $\langle \mathcal{W}, \mathcal{D}_< \rangle$, suppose δ is defeated in $\mathcal{S}$, with $\mathcal{D}'$ as a defeating set and $\mathcal{S}^*$ as a minimal accommodating set for $\mathcal{D}'$. Then each default belonging to $\mathcal{S}^*$ is likewise defeated in $\mathcal{S}$, with $\mathcal{D}'$ as a defeating set and $\mathcal{S}^*$ as a minimal accommodating set for $\mathcal{D}'$.*

Proof. If $\mathcal{S}^*$ is empty, the result is trivial, so suppose otherwise, and pick some δ^* belonging to $\mathcal{S}^*$. We show that δ^* is likewise defeated as follows. Since $\mathcal{S}^*$ is an accommodating set for $\mathcal{D}'$, we know that $\mathcal{S}^* < \mathcal{D}'$ by hypothesis, so that (1) $\delta^* < \mathcal{D}'$, since δ^* belongs to $\mathcal{S}^*$. We know that (2a) $\mathcal{W} \cup Conclusion((\mathcal{S} - \mathcal{S}^*) \cup \mathcal{D}')$ is consistent, also by hypothesis. And since $\mathcal{S}^*$ is a minimal accommodating set for $\mathcal{D}'$, we know that $\mathcal{W} \cup Conclusion((\mathcal{S} - (\mathcal{S}^* - \{\delta^*\})) \cup \mathcal{D}')$ – that is, $\mathcal{W} \cup Conclusion ((\mathcal{S} - \mathcal{S}^*) \cup \mathcal{D}' \cup \{\delta^*\})$ – is inconsistent, from which it follows that (2b) $\mathcal{W} \cup Conclusion((\mathcal{S} - \mathcal{S}^*) \cup \mathcal{D}') \vdash \neg Conclusion(\delta^*)$.   □

THEOREM 1. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory and $\mathcal{S}$ a proper scenario based on this theory. Then $\mathcal{S}$ is also stable.*

Proof. Assuming that $\mathcal{S}$ is a proper scenario, so that $\mathcal{S}$ is the limit of a approximating sequence constrained by $\mathcal{S}$, we need to show that $\mathcal{S} = Binding_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$.

So suppose, first, that $\delta \in \mathcal{S}$. Then there is some $\mathcal{S}_{i+1}$ from the approximating sequence for $\mathcal{S}$ such that $\delta \in \mathcal{S}_{i+1}$. From the definition of an approximating sequence, we know, therefore, that $\delta \in$

$Triggered_{\mathcal{W},\mathcal{D}_<}(\mathcal{S}_i)$, that $\delta \notin Conflicted_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$, and that $\delta \notin Defeated_{\mathcal{W},\mathcal{D}_<}$ $(\mathcal{S})$. Because the triggering function is monotonic in its argument, it follows that that $\delta \in Triggered_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$ as well, since $\mathcal{S}_i \subseteq \mathcal{S}$. Hence the conditions are satisfied to have $\delta \in Binding_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$.

Next, suppose $\delta \in Binding_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$, so that we know from the definition of a binding default that $\delta \in Triggered_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$, that $\delta \notin Conflicted_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$, and that $\delta \notin Defeated_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$. Because $\delta \in Triggered_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$, we have $\mathcal{W} \cup Conclusion(\mathcal{S}) \vdash Premise(\delta)$, from which it follows by compactness, along with the fact that the members of the approximating sequence are nested, that $\mathcal{W} \cup Conclusion$ $(\mathcal{S}_i) \vdash Premise(\delta)$ for some $\mathcal{S}_i$ from the sequence. Therefore, $\delta \in Triggered_{\mathcal{W},\mathcal{D}_<}(\mathcal{S}_i)$. The conditions are thus satisfied to have $\delta \in \mathcal{S}_{i+1}$, and so $\delta \in \mathcal{S}$. $\qquad\square$

THEOREM 2. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory and $\mathcal{S}$ a proper scenario based on this theory. Then $\mathcal{S}$ is also grounded in $\mathcal{W}$.*

Proof. Assuming that $\mathcal{S}$ is a proper scenario, so that $\mathcal{S}$ is the limit of a approximating sequence constrained by $\mathcal{S}$, we need to show that $Th(\mathcal{W} \cup Conclusion(\mathcal{S})) \subseteq Th^{\mathcal{S}}(\mathcal{W})$.

To establish this, we show by induction that $Th(\mathcal{W} \cup Conclusion$ $(\mathcal{S}_i)) \subseteq Th^{\mathcal{S}}(\mathcal{W})$ for each i, from which it follows that $Th(\mathcal{W} \cup Conclusion(\mathcal{S})) \subseteq Th^{\mathcal{S}}(\mathcal{W})$ by compactness, along with the fact that the members of the approximating sequence are nested. The base of the induction is obvious, since $\mathcal{S}_0 = \emptyset$. So suppose as inductive hypothesis that $Th(\mathcal{W} \cup Conclusion(\mathcal{S}_i)) \subseteq Th^{\mathcal{S}}(\mathcal{W})$, and consider some formula $A \in Th(\mathcal{W} \cup Conclusion(\mathcal{S}_{i+1}))$.

It then follows that there must be an ordinary proof of A from $\mathcal{W} \cup Conclusion(\mathcal{S}_{i+1})$ – that is, a sequence of formulas $A_1, A_2, \ldots, A_n$ such that A_n is A and, for $j \leq n$, each A_j either satisfies one of the conditions (1), (2), or (3) from Definition 9, or else the following new condition: $(\ast)$ A_j belongs to $Conclusion(\mathcal{S}_{i+1})$. In order to demonstrate that $A \in Th^{\mathcal{S}}(\mathcal{W})$, we show how this ordinary proof can be transformed into an $\mathcal{S}$-proof of A from $\mathcal{W}$. Since the conditions (1), (2), and (3) are already $\mathcal{S}$-proof conditions, we consider only the case in which A_j is justified by the new condition $(\ast)$.

In that case, we know there is some $\delta \in \mathcal{S}_{i+1}$ such that A_j is $Conclusion(\delta)$. By the definition of the approximating sequence, we then know that $\delta \in Triggered_{\mathcal{W},\mathcal{D}_<}(\mathcal{S}_i)$, and by the definition of triggering, that $\mathcal{W} \cup Conclusion(\mathcal{S}_i) \vdash Premise(\delta)$, or put another way, that $Premise(\delta) \in Th(\mathcal{W} \cup Conclusion(\mathcal{S}_i))$. Our inductive hypothesis therefore

tells us that $Premise(\delta) \in Th^{\mathcal{S}}(\mathcal{W})$, so that there is an $\mathcal{S}$-proof $B_1, B_2, \ldots, B_m$ of $Premise(\delta)$ from $\mathcal{W}$. This new proof can then be inserted directly ahead of A_j in the original sequence, and A_j can now be justified by appeal to condition (4) from Definition 9.

Since each appeal to the new condition $(*)$ can be eliminated in this way in favor of an appeal to condition (4), our original proof of A from $\mathcal{W} \cup Conclusion(\mathcal{S}_{i+1})$ can be transformed into an $\mathcal{S}$-proof of A from $\mathcal{W}$. We therefore have $A \in Th^{\mathcal{S}}(\mathcal{W})$ and the induction is complete.   □

THEOREM 3. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory. Then $\mathcal{S}$ is a proper scenario based on this theory if and only if $\mathcal{S}$ is both a stable scenario based on $\langle \mathcal{W}, \mathcal{D}_< \rangle$ and also grounded in $\mathcal{W}$.*

Proof. It follows from Theorems 1 and 2 that $\mathcal{S}$ is stable and grounded if it is proper, and so we need only establish the other direction. Assume, then, that the scenario $\mathcal{S}$ is stable and grounded – that is, that $\mathcal{S} = Binding_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$ and $Th(\mathcal{W} \cup Conclusion(\mathcal{S})) \subseteq Th^{\mathcal{S}}(\mathcal{W})$ – and let $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \ldots$ be an approximating sequence constrained by $\mathcal{S}$. In order to show that $\mathcal{S}$ is proper, we verify that $\mathcal{S} = \bigcup_{i \geq 0} \mathcal{S}_i$.

For the inclusion from right to left, we show by induction that $\mathcal{S}_i \subseteq \mathcal{S}$ for each i, from which it follows that $\bigcup_{i \geq 0} \mathcal{S}_i \subseteq \mathcal{S}$. The base case is obvious, since $\mathcal{S}_0 = \emptyset$. So suppose as inductive hypothesis that $\mathcal{S}_i \subseteq \mathcal{S}$, and consider some default $\delta \in \mathcal{S}_{i+1}$. From our definition of the approximating sequence, we know that $\delta \in Triggered_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S}_i)$, so that $\delta \in Triggered_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$ by inductive hypothesis together with the monotonicity of triggering. From the definition of the sequence, again, we also have $\delta \notin Conflicted_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$ and $\delta \notin Defeated_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$, so that, all together, we now have $\delta \in Binding_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$. Given our initial assumption that $\mathcal{S} = Binding_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$, we can conclude from this that $\delta \in \mathcal{S}$, and the induction is complete.

For inclusion from left to right, suppose $\delta \in \mathcal{S}$. Since $\mathcal{S} = Binding_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$, we know that $\delta \in Triggered_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$, that $\delta \notin Conflicted_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$, and that $\delta \notin Defeated_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$. Given our definition of the approximating sequence, then, we need only show that there is some i such that $\delta \in Triggered_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S}_i)$ in order to establish that $\delta \in \mathcal{S}_{i+1}$ – from which it will then follow that $\delta \in \bigcup_{i \geq 0} \mathcal{S}_i$.

Since $\delta \in Triggered_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$, we know that $\mathcal{W} \cup Conclusion(\mathcal{S}) \vdash Premise(\delta)$, or put another way, that $Premise(\delta) \in Th(\mathcal{W} \cup Conclusion(\mathcal{S}))$. Given our assumption that $Th(\mathcal{W} \cup Conclusion(\mathcal{S})) \subseteq Th^{\mathcal{S}}(\mathcal{W})$, we therefore have $Premise(\delta) \in Th^{\mathcal{S}}(\mathcal{W})$. But we can now show that $(*)$

for any formula A, if $A \in Th^{\mathcal{S}}(\mathcal{W})$, there is some i such that $A \in Th(\mathcal{W} \cup Conclusion(\mathcal{S}_i))$. Since we have $Premise(\delta) \in Th^{\mathcal{S}}(\mathcal{W})$, this allows us to conclude in particular that there is some i such that $Premise(\delta) \in Th(\mathcal{W} \cup Conclusion(\mathcal{S}_i))$, or put another way, that $\mathcal{W} \cup Conclusion(\mathcal{S}) \vdash Premise(\delta)$. From this, we get the desired result that there is some i such that $Premise(\delta) \in Triggered_{\mathcal{W},\mathcal{D}_<}(\mathcal{S}_i)$, completing the proof.

Our verification of $(*)$ proceeds by induction on length of $\mathcal{S}$-proofs. We show that for any $\mathcal{S}$-proof which establishes that some formula belongs to $Th^{\mathcal{S}}(\mathcal{W})$, there is some i such that the very same proof sequence is an ordinary proof establishing that the same formula belongs to $Th(\mathcal{W} \cup Conclusion(\mathcal{S}_i))$. In the base case, where the $\mathcal{S}$-proof is of length 1, the result is obvious, since the single formula belonging to the proof must be either an axiom or a member of $\mathcal{W}$. So suppose as inductive hypothesis that, for each $\mathcal{S}$-proof of length less than or equal to j establishing that some formula belongs to $Th^{\mathcal{S}}(\mathcal{W})$, there is some i such that the same sequence counts as an ordinary proof establishing that the same formula belongs to $Th(\mathcal{W} \cup Conclusion(S_i))$.

Now consider some $\mathcal{S}$-proof $A_1, \ldots, A_j, A_{j+1}$ establishing that A_{j+1} belongs to $Th^{\mathcal{S}}(\mathcal{W})$, with length $j + 1$. By hypothesis, there is some i such that the sequence $A_1, \ldots, A_j$ is an ordinary proof establishing that A_j belongs to $Th(\mathcal{W} \cup Conclusion(S_i))$. If the formula A_{j+1} is justified by condition (1), (2), or (3) of Definition 9, then of course $A_1, \ldots, A_j, A_{j+1}$ is likewise an ordinary proof showing that A_{j+1} belongs to $Th(\mathcal{W} \cup Conclusion(\mathcal{S}_i))$. So suppose A_{j+1} is justified by condition (4) – that is, that there is some δ from $\mathcal{S}$ such that $Conclusion(\delta)$ is A_{j+1} and $Premise(\delta)$ is a previous member of the sequence. Again, the inductive hypothesis allows us to conclude that $Premise(\delta)$ belongs to $Th(\mathcal{W} \cup Conclusion(S_i))$, so that $\delta \in Triggered_{\mathcal{W},\mathcal{D}_<}(\mathcal{S}_i)$. Since $\delta \notin Conflicted_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$ and $\delta \notin Defeated_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$, the definition of the sequence tells us that $\delta \in \mathcal{S}_{i+1}$, so that $A_{j+1} \in Conclusion(\mathcal{S}_{i+1})$. This allows us to conclude that $A_1, \ldots, A_j, A_{j+1}$ is an ordinary proof showing that A_{j+1} belongs to $Th(\mathcal{W} \cup Conclusion \ (\mathcal{S}_{i+1}))$, and so the induction is complete. □

THEOREM 4. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory with $\mathcal{E}$ as an extension. Then $\mathcal{E}$ is consistent if and only if $\mathcal{W}$ is consistent.*

Proof. Since $\mathcal{E}$ is an extension, we have $\mathcal{E} = Th(\mathcal{W} \cup Conclusion(\mathcal{S}))$, where $\mathcal{S}$ is some proper scenario based on $\langle \mathcal{W}, \mathcal{D}_< \rangle$. If $\mathcal{W}$ is inconsistent, $\mathcal{E}$ must be inconsistent as well, since $\mathcal{W} \subseteq \mathcal{E}$. If $\mathcal{E}$ is inconsistent, then the set $\mathcal{W} \cup Conclusion(\mathcal{S})$ entails every formula. Therefore, $Conflicted_{\mathcal{W},\mathcal{D}_<}(\mathcal{S}) = \mathcal{D}$ – every default from $\mathcal{D}$ is conflicted in $\mathcal{S}$ – and so $Binding_{\mathcal{W},\mathcal{D}_<}$

$(S) = \emptyset$. It follows that $\mathcal{S} = \emptyset$ as well, since $\mathcal{S}$ is a stable scenario, so that *Conclusion*$(S) = \emptyset$, of course. We thus have $\mathcal{E} = Th(\mathcal{W})$, so that $\mathcal{W}$ must be inconsistent.   □

THEOREM 5. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory in which the ordering $<$ is empty. Then $\mathcal{E}$ is an extension of $\langle \mathcal{W}, \mathcal{D}_< \rangle$ if and only if $\mathcal{E}$ is a Reiter extension of $\langle \mathcal{W}, \mathcal{D} \rangle$, the corresponding normal default theory.*

Proof (*Sketch*). Left to right. Suppose $\mathcal{E}$ is an extension of $\langle \mathcal{W}, \mathcal{D}_< \rangle$. Then $\mathcal{E} = Th(\mathcal{W} \cup Conclusion(S))$, where $\mathcal{S}$ is a proper scenario based on $\langle \mathcal{W}, \mathcal{D}_< \rangle$ – that is, $\mathcal{S} = \bigcup_{i \geq 0} \mathcal{S}_i$, where $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \ldots$ is an approximating sequence constrained by $\mathcal{S}$. Define the sequence $\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}_2, \ldots$ by putting

$$\mathcal{E}_0 = \mathcal{W},$$

$$\mathcal{E}_{i+1} = Th(\mathcal{E}_i) \cup Conclusion(\mathcal{S}_{i+1}),$$

and let $\mathcal{E}' = \bigcup_{i \geq 0} \mathcal{E}_i$. It is easy to see that $\mathcal{E}' = \mathcal{E}$. Hence, it is necessary only to show that $\mathcal{E}'$ is a Reiter extension, by verifying that the $\mathcal{E}_i$ sequence meets the conditions of Definition 11.

We begin by noting that (∗) $Th(\mathcal{E}_i) = Th(\mathcal{W} \cup Conclusion(\mathcal{S}_i))$ for each i, and also that (∗∗) $Th(\mathcal{E}) = Th(\mathcal{W} \cup Conclusion(\mathcal{S}))$. The first of these results can be established by induction. The base case, with $i = 0$, is evident from the definition of the $\mathcal{E}_i$ sequence. As inductive hypothesis, suppose that $Th(\mathcal{E}_i) = Th(\mathcal{W} \cup Conclusion(\mathcal{S}_i))$ for some i. The inductive step can then be established through the chain of reasoning

$$Th(\mathcal{E}_{i+1}) = Th(Th(\mathcal{E}_i) \cup Conclusion(\mathcal{S}_{i+1}))$$

$$= Th(Th(\mathcal{W} \cup Conclusion(\mathcal{S}_i)) \cup Conclusion(\mathcal{S}_{i+1}))$$

$$= Th(Th(\mathcal{W}) \cup Conclusion(\mathcal{S}_{i+1}))$$

$$= Th(\mathcal{W} \cup Conclusion(\mathcal{S}_{i+1})),$$

in which the first equation follows from the definition of the $\mathcal{E}_i$ sequence, the second from the inductive hypothesis, the third from the fact that $\mathcal{S}_i \subseteq \mathcal{S}_{i+1}$, and the fourth from general properties of the *Th* operator.

The second result can be established by reasoning as follows

$$Th(\mathcal{E}) = Th\left(\bigcup_{i \geq 0} \mathcal{E}_i\right)$$

$$= \bigcup_{i \geq 0} (Th(\mathcal{E}_i))$$

$$= \bigcup_{i \geq 0} (Th(\mathcal{W} \cup Conclusion(\mathcal{S}_i)))$$

$$\text{so:} = Th\left(\mathcal{W} \cup Conclusion\left(\bigcup_{i \geq 0} \mathcal{S}_i\right)\right)$$

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

where the first equation holds because $\mathcal{E} = \bigcup_{i \geq 0} \mathcal{E}_i$, the second by compactness and because the $\mathcal{E}_i$ sequence is nested, the third due to the previous $(*)$, the fourth by compactness and because the $\mathcal{S}_i$ sequence is nested, and the fifth because $\mathcal{S} = \bigcup_{i \geq 0} \mathcal{S}_i$.

In order to verify that the $\mathcal{E}_i$ sequence meets the conditions of Definition 11, it is enough to verify the equation

$$\mathcal{S}_{i+1} = \{\delta \in \mathcal{D} : \mathcal{E}_i \vdash Premise(\delta),$$
$$\mathcal{E} \nvdash \neg Conclusion(\delta)\}.$$

Because the $<$ ordering is empty, no default can be defeated in any scenario. Hence, by the definition of the $\mathcal{S}_i$ sequence, $\delta \in \mathcal{S}_{i+1}$ just in case $\delta \in Triggered_{\mathcal{W},\mathcal{D}_<}(\mathcal{S}_i)$ and $\delta \notin Conflicted_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$. By definition, $\delta \in Triggered_{\mathcal{W},\mathcal{D}_<}(\mathcal{S}_i)$ just in case $\mathcal{W} \cup Conclusion(\mathcal{S}_i) \vdash Premise(\delta)$, which is equivalent by $(*)$ to the condition that $\mathcal{E}_i \vdash Premise(\delta)$. And $\delta \notin Conflicted_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$ just in case $\mathcal{W} \cup Conclusion(\mathcal{S}) \nvdash \neg Conclusion(\delta)$, which is equivalent by $(**)$ to the condition that $\mathcal{E} \nvdash \neg Conclusion(\delta)$. The equation is therefore established.

Right to left (sketch). Suppose $\mathcal{E}$ is a Reiter extension of $\langle \mathcal{W}, \mathcal{D} \rangle$. Then $\mathcal{E} = \bigcup_{i \geq 0} \mathcal{E}_i$, with the sequence $\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}_2, \ldots$ specified as in Definition 11. Define the sequence $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \ldots$ by putting

$$\mathcal{S}_0 = \emptyset,$$
$$\mathcal{S}_{i+1} = \{\delta \in \mathcal{D} : \mathcal{E}_i \vdash Premise(\delta),$$
$$\mathcal{E} \nvdash \neg Conclusion(\delta)\};$$

let $\mathcal{S} = \bigcup_{i \geq 0} \mathcal{S}_i$, and let $\mathcal{E}' = Th(\mathcal{W} \cup Conclusion(\mathcal{S}))$. The result can then be verified by showing that $\mathcal{E}' = \mathcal{E}$, and that the $\mathcal{S}_i$ sequence is an approximating sequence constrained by the scenario $\mathcal{S}$. $\qquad\qquad\square$

OBSERVATION 4. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory, and suppose $\mathcal{S}$ is a stable scenario based on this theory. Then if some default δ is defeated in $\mathcal{S}$, with $\mathcal{D}'$ as a defeating set, it follows that $Conclusion(\mathcal{S} \cup \mathcal{D}')$ is consistent.*

Proof. Assume that the default δ is defeated in the scenario $\mathcal{S}$ with $\mathcal{D}'$ as a defeating set and $\mathcal{S}'$ as an accommodating set for $\mathcal{D}'$. By Observation 1, it follows that there is some $\mathcal{S}^* \subseteq \mathcal{S}'$ – so that $\mathcal{S}^* \subseteq \mathcal{S}$, of course – such that δ is likewise defeated with $\mathcal{D}'$ as a defeating set and $\mathcal{S}^*$ as a minimal accommodating set for $\mathcal{D}'$. Now suppose $Conclusion(\mathcal{S} \cup \mathcal{D}')$ is not consistent. By Observation 2, it follows that $\mathcal{S}^*$ is nonempty, and by Observation 3, that each default belonging to $\mathcal{S}^*$ is itself defeated in $\mathcal{S}$. But this is impossible, since $\mathcal{S}^* \subseteq \mathcal{S}$, and, because $\mathcal{S} = Binding_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$, no default belonging to $\mathcal{S}$ can be defeated. $\square$

OBSERVATION 5. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory in which the set $\mathcal{D}$ of defaults is finite, and suppose $\mathcal{S}$ is a stable scenario based on this theory. Then if some default δ is defeated in $\mathcal{S}$, with $\mathcal{D}'$ as a defeating set, it follows that $\mathcal{D}' \subseteq \mathcal{S}$.*

Proof. Since $\mathcal{D}$ is finite, we can define the degree of a default δ-written, $degree(\delta)$ – as follows: if there is no δ' such that $\delta < \delta'$, then $degree(\delta) = 0$, and otherwise,

$$degree(\delta) = 1 + maximum(\{degree(\delta') : \delta < \delta'\}).$$

The result can then be established by induction on the degree of the defeated default. The base case, with $degree(\delta) = 0$, is trivial, since defaults can be defeated only by other defaults having higher priority. But if $degree(\delta) = 0$, then δ has a maximal priority, and so can never be defeated.

As inductive hypothesis, suppose we know that, for any default whose degree is less than n, whenever that default is defeated in some scenario $\mathcal{S}$, any defeating set for the default must be a subset of that scenario. And where δ is a particular default with $degree(\delta) = n$, suppose that δ is defeated in the scenario $\mathcal{S}$ with defeating set $\mathcal{D}'$.

From the definition of defeat – and also from Observation 4, which tells us that $Conclusion(\mathcal{S} \cup \mathcal{D}')$ is itself consistent, so that the accommodating set can be empty – we know that $\mathcal{D}'$ is a subset of $Triggered_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$, and also that (1) $\delta < \mathcal{D}'$, that (2a) $\mathcal{W} \cup Conclusion(\mathcal{S} \cup \mathcal{D}')$ is consistent, and that (2b) $\mathcal{W} \cup Conclusion(\mathcal{S} \cup \mathcal{D}') \vdash \neg Conclusion(\delta)$. In order to show that $\mathcal{D}'$ is a subset of $\mathcal{S}$, pick some default

δ' from $\mathcal{D}'$. We know that $\mathcal{D}'$ is triggered in the scenario $\mathcal{S}$, and also, from (2a), that it is not conflicted. Because $\mathcal{S} = Binding_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$, therefore, δ must belong to $\mathcal{S}$ unless it is defeated.

Assume, then, that δ' is defeated in $\mathcal{S}$, with $\mathcal{D}''$ as a defeating set. Then from the definition of defeat and Observation 4, again, we know that $\mathcal{D}''$ is also a subset of $Triggered_{\mathcal{W}, \mathcal{D}_<}(\mathcal{S})$, and as before, that $(1')$ $\delta' < \mathcal{D}''$, that $(2a')$ $\mathcal{W} \cup Conclusion(\mathcal{S} \cup \mathcal{D}'')$ is consistent, and that $(2b')$ $\mathcal{W} \cup Conclusion(\mathcal{S} \cup \mathcal{D}'') \vdash \neg Conclusion(\delta')$. From the fact that $degree(\delta) = n$, as well as (1) above, we know that $degree(\delta') < n$. Our inductive hypothesis therefore tells us that $\mathcal{D}'' \subseteq \mathcal{S}$, which together with $(2b')$ allows us to conclude that $\mathcal{W} \cup Conclusion(\mathcal{S}) \vdash \neg Conclusion(\delta')$. Since δ' belongs to $\mathcal{D}'$, however, this contradicts the previous (2a), and so the assumption that δ' is defeated fails.

Therefore δ' belongs to $\mathcal{S}$, and the proof is complete.   □

OBSERVATION 6. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory in which the set $\mathcal{D}$ of defaults is finite, and suppose $\mathcal{S}$ is a stable scenario based on this theory. Then any default that is defeated in $\mathcal{S}$ must also be conflicted in $\mathcal{S}$.*

Proof. Suppose δ is defeated in $\mathcal{S}$, with $\mathcal{D}'$ as a defeating set. Then by the definition of defeat, we know, among other things, that $\mathcal{W} \cup Conclusion(\mathcal{S} \cup \mathcal{D}') \vdash \neg Conclusion(\delta)$. Observation 5 tells us that $\mathcal{D}' \subseteq \mathcal{S}$. Therefore $\mathcal{W} \cup Conclusion(\mathcal{S}) \vdash \neg Conclusion(\delta)$ as well, so that δ is conflicted in $\mathcal{S}$.   □

THEOREM 6. *Let $\langle \mathcal{W}, \mathcal{D}_< \rangle$ be an ordered default theory in which the set $\mathcal{D}$ of defaults is finite. Then if $\mathcal{E}$ is an extension of $\langle \mathcal{W}, \mathcal{D}_< \rangle$, it follows that $\mathcal{E}$ is also an extension of $\langle \mathcal{W}, \mathcal{D} \rangle$, the corresponding normal default theory.*

Proof. The proof follows the pattern of the first half of the proof of the earlier Theorem 5. We begin, as before, by noting that $\mathcal{E} = Th(\mathcal{W} \cup Conclusion(\mathcal{S}))$ where $\mathcal{S}$ is a proper scenario – that is, $\mathcal{S} = \bigcup_{i \geq 0} \mathcal{S}_i$, where $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \ldots$ is an approximating sequence constrained by $\tilde{\mathcal{S}}$. As before, we define the sequence $\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}_2, \ldots$ by putting

$$\mathcal{E}_0 = \mathcal{W},$$

$$\mathcal{E}_{i+1} = Th(\mathcal{E}_i) \cup Conclusion(\mathcal{S}_{i+1}).$$

Setting $\mathcal{E}' = \bigcup_{i \geq 0} \mathcal{E}_i$, it is again easy to see that $\mathcal{E}' = \mathcal{E}$. Hence, it remains only to show that $\mathcal{E}'$ is a Reiter extension, by verifying that the $\mathcal{E}_i$ sequence meets the conditions of Definition 11, which we can accomplish, as before, by showing that

$$\mathcal{S}_{i+1} = \{\delta \in \mathcal{D} : \mathcal{E}_i \vdash Premise(\delta),$$

$$\mathcal{E} \nvdash \neg Conclusion(\delta)\}.$$

By definition of the $\mathcal{S}_i$ sequence, we have $\delta \in \mathcal{S}_{i+1}$ just in case $\delta \in Triggered_{\mathcal{W},\mathcal{D}_<}(\mathcal{S}_i)$, and $\delta \notin Conflicted_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$, and $\delta \notin Defeated_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$. It is again possible to establish the earlier $(*)$ and $(**)$, and then to use these preliminary facts to verify that $\delta \in Triggered_{\mathcal{W},\mathcal{D}_<}(\mathcal{S}_i)$ if and only if $\mathcal{E}_i \vdash Premise(\delta)$, and that $\delta \notin Conflicted_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$ if and only if $\mathcal{E} \nvdash \neg Conclusion(\delta)$. The right hand side of the equation therefore contains those defaults that are triggered in $\mathcal{S}_i$ and not conflicted in $\mathcal{S}$, exactly as before.

In this new case, however, since the priority ordering $<$ is no longer empty, it is now possible for a default to be defeated in $\mathcal{S}$, and as we have seen, the membership conditions for $\mathcal{S}_{i+1}$ specify that $\delta \notin Defeated_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$. Since defaults that are defeated in $\mathcal{S}$ cannot belong to the left hand side of the equation, we must be able to show that they cannot belong to the right hand side either. Fortunately, Observation 6 allows us to conclude that any default that is defeated in $\mathcal{S}$ is also conflicted – that $Defeated_{\mathcal{W},\mathcal{D}_<}(\mathcal{S}) \subseteq Conflicted_{\mathcal{W},\mathcal{D}_<}(\mathcal{S})$. By ruling out conflicted defaults, the right hand side therefore rules out defeated defaults as well, and the result is established. $\square$

NOTES

[1] 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.'

[2] This concept, couched in slightly different notation, is explored in detail in Section 3.1 of Marek and Truszczyński [20], and plays a central role in their presentation of default logic. See also Chapter 4 of Makinson [19] for an analysis of default logic along similar lines.

[3] This reasoning strategy was first labelled as "credulous" by Touretzky et al. [27], and as the "choice" option by Makinson [18]; it had earlier been characterized as "brave" by McDermott [21].

[4] The resulting logic generalizes that of van Fraassen [29]. The interpretation of van Fraassen's account within default logic was first established in my [11]; a defense of the overall approach can be found in my [15].

[5] The label is again due to Touretzky et al. [27]; the same reasoning strategy had earlier been described as "cautions" by McDermott [21].

[6] An argument that the skeptical approach, as defined here, presents the only coherent option for epistemic default reasoning is presented by Pollack [22]; some of my own doubts can be found in [14].

[7] See, for example, Antonelli [1], which adapts ideas from Kripke's treatment of the paradoxes to modify Reiter's original default logic, without priorities, so that existence of extensions is guaranteed. It is reasonable to hope that Antonelli's approach might be applicable in the present setting, since it is based, originally, in a study of nonmonotonic inheritance networks, and draws on many of the ideas and formulations at work here.

[8] It is easy to see, for example, that the "PDL-extensions" defined by Brewka [5] for his prioritized default logic form a subset of the corresponding Reiter extensions, and a similar result is established as by Baader and Hollunder [3]. Rintanen [25] simply defines his "preferred extensions" as Reiter extensions that satisfy a complex preferential criterion. Brewka and Eiter [6] likewise build it into their definition of the "prioritized extensions" that these are a subset of the Reiter extensions, as do Marek and Truszczyński [20].

[9] This example arose in discussion with Bijan Parsia and Michael Morreau.

[10] The same point could be made by focusing on the problem of *reinstatement*, which I have discussed at length in my [13]; the correct extension of the Microsoft example from that paper, for instance, is not a Reiter extension of the corresponding normal default theory.

[11] This definition of an active default is due to Baader and Hollunder [3].

[12] For a more precise statement of the constraint at work here, see Principle I from Brewka and Eiter [6].

[13] See Sections 3.1 and 4.2 of Delgrande and Schaub [7] for discussion of this example.

[14] This line of reasoning was suggested to me by Paul Pietroski.

[15] This example is drawn from my [13].

[16] In Horty et al. [17], a "degree" ordering is defined on the defaults present in the very simple language of defeasible inheritance networks, and the correct results are generated when defaults are considered in order of their degree; but this notion of degree has not been successfully extended to richer languages.

[17] A response suggested by Jörg Hansen.

REFERENCES

1. Antonelli, G. A. (1999): A directly cautious theory of defeasible consequence for default logic via the notion of a general extension, *Artificial Intelligence* **109**, 71–109.
2. Apt, K. R., Blair, H. A., and Walker, A. (1998): Towards a theory of declarative knowledge, in J. Minker (ed.), *Foundations of Deductive Databases and Logic Programming*, Morgan Kaufmann, San Mateo, CA, pp. 89–148.
3. Baader, F. and Hollunder, B. (1995): Priorities on defaults with prerequisites, and their applications in treating specificity in terminological default logic. *Journal of Automated Reasoning* **15**, 41–68.

4. Brewka, G. (1994): Adding priorities and specificity to default logic, in *Proceedings of the European Workshop on Logics in Artificial Intelligence (JELIA-94)*, Springer Verlag Lecture Notes in Artificial Intelligence, Springer, Berlin Heidelberg New York, pp. 247–260.

5. Brewka, G. (1994): Reasoning about priorities in default logic, in *Proceedings of the Twelveth National Conference on Artificial Intelligence (AAAI-94)*, AAAI/MIT, pp. 940–945.

6. Brewka, G. and Eiter, T. (2000): Prioritizing default logic, in St. Hölldobler (ed.), *Intellectics and Computational Logic: Papers in Honor of Wolfgang Bibel*, Kluwer.

7. Delgrande, J. and Schaub, T. (2000): Expressing preferences in default logic, *Artificial Intelligence* **123**, 41–87.

8. Delgrande, J. and Schaub, T. (2000): The role of default logic in knowledge representation, in J. Minker (ed.), *Logic Based Artificial Intelligence*, Kluwer, Dordrecht, pp. 107–126.

9. Delgrande, J., Schaub, T., Tompits, H., and Wang, K. (2004): A classification and survey of preference handling approaches in nonmonotonic reasoning, *Computational Intelligence* **20**, 308–334.

10. Etherington, D. and Reiter, R. (1983): On inheritance hierarchies with exceptions, in *Proceedings of the Third National Conference on Artificial Intelligence (AAAI-83)*, pp. 104–108.

11. Horty, J. (1994): Moral dilemmas and nonmonotonic logic, *Journal of Philosophical Logic* **23**, 35–65.

12. Horty, J.: Some direct theories of nonmonotonic inheritance, in D. Gabbay, C. Hogger, and J. Robinson (eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning*, Oxford University Press, 1994, pp. 111–187.

13. Horty, J. (2001): Argument construction and reinstatement in logics for defeasible reasoning, *Artificial Intelligence and Law* **9**, 1–28.

14. Horty, J. (2002): Skepticism and floating conclusions, *Artificial Intelligence* **135**, 55–72.

15. Horty, J. (2003): Reasoning with moral conflicts, *Nous* **37**, 557–605.

16. Horty, J. (2006): Reasons as defaults, Manuscript.

17. Horty, J., Thomason, R, and Touretzky, D. (1990): A skeptical theory of inheritance in nonmonotonic semantic networks, *Artificial Intelligence* **42**, 311–348.

18. Makinson, D. (1994): General patterns in nonmonotonic reasoning, in D. Gabbay, C. Hogger, and J. Robinson (eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning*, Oxford University Press, pp. 35–110.

19. Makinson, D. (2005): *Bridges from Classical to Nonmonotonic Logic*, Texts in Computing, Volume 5, King's College Publications.

20. Marek, W. and Truszczyński, M. (1993): *Nonmonotonic Logic: Context-dependent Reasoning*, Springer, Berlin Heidelberg New York.

21. McDermott, D. (1982): Non-monotonic logic II, *Journal of the Association for Computing Machinery* **29**, 33–57.

22. Pollock, J. (1995): *Cognitive Carpentry: A Blueprint for How to Build a Person*, MIT, Cambridge, MA.

23. Reiter, R. (1980): A logic for default reasoning, *Artificial Intelligence* **13**, 81–132.

24. Reiter, R. and Criscuolo, G. (1981): On interacting defaults, in *Proceedings of the Seventh International Joint Conference on Artificial Intelligence (IJCAI-81)*, pp. 270–276.

25. Rintanen, J. (1998): Lexicographic priorities in default logic, *Artificial Intelligence* **106**, 221–265.
26. Touretzky, D. (1986): *The Mathematics of Inheritance Systems*, Morgan Kaufmann, San Mateo, CA.
27. Touretzky, D., Horty, J., and Thomason, R. (1987): 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)*, Morgan Kaufmann, San Mateo, CA, pp. 476–482.
28. Touretzky, D., Thomason, R., and Horty, J. (1991): A skeptic's menagerie: conflictors, preemptors, reinstaters, and zombies in nonmonotonic inheritance, in *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI-91)*, Morgan Kaufmann, San Mateo, CA, pp. 478–483.
29. van Fraassen, B. (1973): Values and the heart's command, *The Journal of Philosophy* **70**, 5–19.

Philosophy Department and
Institute for Advanced Computer Studies,
University of Maryland,
College Park, MD 20742, USA
E-mail: horty@umiacs.umd.edu
URL: http://www.umiacs.umd.edu/users/horty