Noname manuscript No.
(will be inserted by the editor)

Break the Weakest Rules

Hypothetical Reasoning in Default Logic

Dustin Tucker

Received: date / Accepted: date

Abstract John Horty has proposed using default logic to model reasons and their interactions, an approach with which I am largely sympathetic. Unfortunately, his system has some unwelcome consequences, which are, I think, due to an inability to capture a sort of hypothetical reasoning—roughly, reasoning about the subsequent decisions we will face if we make certain decisions. I develop a new, simpler default logic that does justice to hypothetical reasoning of this sort.

Keywords Default logic · Nonmonotonic logic · Order Puzzle

1 Introduction

In [1],[1] John Horty proposes using default logic to model reasons and their interactions, and I am sympathetic with both the general strategy and many of the details of the proposal. Unfortunately, his system also has some unwelcome consequences. During his discussion of these consequences, Horty briefly mentions but does not pursue the idea that they are due to his logic's inability to capture a sort of hypothetical reasoning—reasoning about the subsequent decisions we will face if we make certain decisions. This idea seems to me to be exactly right. I begin in Sect. 2 with the background I take from Horty. I then develop in Sect. 3 a new, simpler default logic that does justice to hypothetical reasoning of this sort, illustrating the difference between my theory and Horty's in Sect. 4.

D. Tucker
Department of Philosophy
Colorado State University
Fort Collins, CO 80523-1781
E-mail: dustin.tucker@colostate.edu

[1] All page and section references are to this book.

2 Background

I adopt without argument a number of basic definitions and concepts from Horty.[2] Default logic begins with standard propositional logic. I follow Horty in using $\neg$, $\wedge$, $\vee$, $\supset$, and $\equiv$ for the connectives; they have their usual meanings and are governed by the usual rules of inference. Default logic also includes *defaults*, however, which are special, more specific rules of inference. A classic example says that from the information that Tweety is a bird you should conclude by default that Tweety can fly. Of course, not all birds fly, so this must be only a default inference, but it is still reasonable. The premise of this rule of inference is that Tweety can fly and the conclusion is that Tweety is a bird. Following Horty, I represent this default rule with $B \to D$.

Let a *default theory* be an ordered pair $\langle \mathcal{W}, \mathcal{D} \rangle$ where $\mathcal{W}$ is a set of sentences—intuitively, the background information we have—and $\mathcal{D}$ is a set of defaults—intuitively, the defaults we might use. I usually represent these theories with a list. Thus, for example,

δ_1: $B \to F$
δ_2: $P \to \neg F$

$\mathcal{W}$: P, $P \supset B$

represents the default theory in which $\mathcal{W} = \{P, P \supset B\}$ and $\mathcal{D} = \{\delta_1, \delta_2\}$ where δ_1 is $B \to F$ and δ_2 is $P \to \neg F$. We have already seen δ_1; δ_2 captures the idea that it is good to conclude that Tweety cannot fly from the information that Tweety is a penguin. $\mathcal{W}$ represents that in this situation, we know both that Tweety is a penguin and that Tweety is a bird if Tweety is a penguin.

A *scenario* is a set $\mathcal{S}$ of defaults. Say that $\mathcal{S}$ is based on a set A iff $\mathcal{S} \subseteq A$ and that $\mathcal{S}$ is based on a default theory $\langle \mathcal{W}, \mathcal{D} \rangle$ iff it is based on $\mathcal{D}$.[3] $\mathcal{S}$ is intuitively the set of defaults we have decided to use—the set of default rules of inference whose conclusions we have decided to endorse.

Given a scenario $\mathcal{S}$ based on a default theory $\langle \mathcal{W}, \mathcal{D} \rangle$, a default $\delta \in \mathcal{D}$ is *triggered* iff its premise is entailed by $\mathcal{W}$ combined with the conclusions of the defaults in $\mathcal{S}$. In effect, δ is triggered iff we are committed to its premise. Analogously, δ is *conflicted* iff we are committed to the negation of its conclusion.

Let $Premise(P \to Q) = P$ and $Conclusion(P \to Q) = Q$, and given a set of defaults A, let $Conclusion(A) = \{Conclusion(\delta) : \delta \in A\}$. We can then define the sets of triggered and conflicted defaults in a scenario $\mathcal{S}$ based on a default theory $\langle \mathcal{W}, \mathcal{D} \rangle$ as follows.

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

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

Defaults come in different strengths. Tweety being a penguin, for instance, is a stronger reason to believe that Tweety cannot fly than Tweety being a bird is for believing that Tweety can fly. An *ordered default theory* is an ordered triple $\langle \mathcal{W}, \mathcal{D}, < \rangle$

[2] These come from [§1.1] except where noted.

[3] I deviate very slightly from Horty here: he does not define the notion of a scenario being based on a set of defaults, skipping directly to it being based on a default theory.

where $\langle\mathcal{W},\mathcal{D}\rangle$ is a default theory and $<$ is a partial order on $\mathcal{D}$; intuitively we have $\delta < \delta'$ iff *Premise*(δ) is a weaker reason for *Conclusion*(δ) than *Premise*(δ') is for *Conclusion*(δ'). Say that a scenario $\mathcal{S}$ is based on an ordered default theory $\langle\mathcal{W},\mathcal{D},<\rangle$ iff it is based on $\langle\mathcal{W},\mathcal{D}\rangle$.

Finally, given sets A and B of defaults, say $A < B$ iff for every $\delta \in A$ and $\delta' \in B$ we have $\delta < \delta'$, and abbreviate $A < \{\delta\}$ and $\{\delta\} < A$ with $A < \delta$ and $\delta < A$ respectively.[4] Though Horty never does, I also sometimes write $\delta' > \delta$ for $\delta < \delta'$, etc.

This is all I adopt from Horty.

3 A default logic

Consider again the $\mathcal{D}$ from above:

δ_1: $B \to F$
δ_2: $P \to \neg F$

There are four scenarios based on this set:

$\mathcal{S}_1$: $\emptyset$
$\mathcal{S}_2$: $\{\delta_1\}$
$\mathcal{S}_3$: $\{\delta_2\}$
$\mathcal{S}_4$: $\{\delta_1,\delta_2\}$

That is, we can commit to nothing ($\mathcal{S}_1$), commit to Tweety being able to fly because Tweety's a bird ($\mathcal{S}_2$), commit to Tweety not being able to fly because Tweety's a penguin ($\mathcal{S}_3$), or commit both to Tweety being able to fly because Tweety's a bird and to Tweety not being able to fly because Tweety's a penguin.

Of course, which scenario(s) is (are) best intuitively depends on $\mathcal{W}$ and $<$. I introduce two concepts, that of a proper scenario (though this does not mean for me what it means for Horty) and that of an optimal scenario, to capture these two dependencies.

Call a scenario $\mathcal{S}$ based on a default theory $\langle\mathcal{W},\mathcal{D}\rangle$ *proper* iff every default in $\mathcal{S}$ is triggered and no defaults in $\mathcal{S}$ are conflicted—more formally, iff for every $\delta \in \mathcal{S}$ we have $\delta \in \mathit{Triggered}_{\mathcal{W},\mathcal{D}}(\mathcal{S})$ and $\delta \notin \mathit{Conflicted}_{\mathcal{W},\mathcal{D}}(\mathcal{S})$.[5] Thus, whether a scenario is proper can depend on $\mathcal{W}$. $\mathcal{S}_4$ can never be proper because no matter what $\mathcal{W}$ is, δ_1 and δ_2 will be conflicted (because their conclusions contradict each other). And $\mathcal{S}_1$ is trivially proper no matter what $\mathcal{W}$ is (even no matter what $\mathcal{D}$ is). But whether $\mathcal{S}_2$ and $\mathcal{S}_3$ are proper depends on whether δ_1 and δ_2 are triggered. If $\mathcal{W} = \{B\}$, then only $\mathcal{S}_1$

[4] [p. 194], though in fact Horty never defines the former abbreviation and I never use the latter.

[5] Actually, this is not quite right, though it is fine for this paper. For the same reasons Horty does [pp. 222ff.], it is better to define proper scenarios in terms of approximating sequences: $\mathcal{S}_0,\mathcal{S}_1,\mathcal{S}_2,\ldots$ is an *approximating sequence constrained by* $\mathcal{S}$ iff

(i) $\mathcal{S}_0 = \emptyset$ and
(ii) for every $\mathcal{S}_{i+1}$ we have for every $\delta \in \mathcal{S}_{i+1}$ both
  (a) $\delta \in \mathit{Triggered}_{\mathcal{W},\mathcal{D}}(\mathcal{S}_i)$ and
  (b) $\delta \notin \mathit{Conflicted}_{\mathcal{W},\mathcal{D}}(\mathcal{S})$.

$\mathcal{S}$ is *proper* iff $\mathcal{S} = \bigcup_{i\geq 0}\mathcal{S}_i$ for some approximating sequence $\mathcal{S}_0,\mathcal{S}_1,\mathcal{S}_2,\ldots$ constrained by $\mathcal{S}$.

and $\mathcal{S}_2$ are proper; if $\mathcal{W} = \{P\}$, then only $\mathcal{S}_1$ and $\mathcal{S}_3$ are proper; if $\mathcal{W} = \{P, P \supset B\}$, as above, then $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{S}_3$ are proper; etc.

Being proper is a relatively minimal requirement, in effect ruling out only inconsistent commitments and commitments for which we have no reasons. The real work of deciding which scenario to adopt is done by optimality, but we first define the notion of a violated default. Given a scenario $\mathcal{S}$ based on a default theory $\langle \mathcal{W}, \mathcal{D} \rangle$, a default $\delta \in \mathcal{D}$ is *violated* iff it is triggered in $\mathcal{S}$ and not a member of $\mathcal{S}$. As with triggered and conflicted defaults, we can define the set of violated defaults:

$$Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}) = \{\delta \in \mathcal{D} : \delta \in Triggered_{\mathcal{W},\mathcal{D}}(\mathcal{S}) \text{ and } \delta \notin \mathcal{S}\}.$$

Thus, for example, given the familiar default theory (and repeating $\mathcal{S}_1$–$\mathcal{S}_3$ for convenience)

$\delta_1 \colon B \to F$	$\mathcal{S}_1 \colon \emptyset$
$\delta_2 \colon P \to \neg F$	$\mathcal{S}_2 \colon \{\delta_1\}$
$\mathcal{W} \colon P,\ P \supset B$	$\mathcal{S}_3 \colon \{\delta_2\}$

we have

- $Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}_1) = \{\delta_1, \delta_2\}$,
- $Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}_2) = \{\delta_2\}$, and
- $Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}_3) = \{\delta_1\}$.

We come now to the most important definition: A scenario $\mathcal{S}$ based on an ordered default theory $\langle \mathcal{W}, \mathcal{D}, < \rangle$ is *suboptimal* iff for some proper scenario $\mathcal{S}'$ based on $\langle \mathcal{W}, \mathcal{D}, < \rangle$ and some $\delta \in \mathcal{D}$,

(i) $\delta \in Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}) - Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}')$ and
(ii) $\delta > Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}') - Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S})$.

In these circumstances, say that $\mathcal{S}$ is *less optimal than* $\mathcal{S}'$ (*because of* δ). That is, $\mathcal{S}$ is less optimal than $\mathcal{S}'$ because of δ iff δ is a default that is (i) violated in $\mathcal{S}$ but not $\mathcal{S}'$ and (ii) stronger than every default that is violated in $\mathcal{S}'$ but not $\mathcal{S}$. Given this definition, both $\mathcal{S}_1$ and $\mathcal{S}_2$ are less optimal than $\mathcal{S}_3$ because of δ_2. ($\mathcal{S}_1$ is also less optimal than $\mathcal{S}_2$ because of δ_1: $Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}_2) - Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}_1) = \emptyset$ and $\delta_1 > \emptyset$ vacuously.)

Finally, a scenario $\mathcal{S}$ based on an ordered default theory $\langle \mathcal{W}, \mathcal{D}, < \rangle$ is *optimal* iff it is proper and not suboptimal. In the present example, then, only $\mathcal{S}_2$ is optimal—the logic tells us that we ought to endorse the conclusion that Tweety cannot fly. Intuitively, the idea is that when we must break rules, we ought to break the weakest ones we can—when every proper scenario violates at least one default, the optimal ones are those that violate only the weakest defaults.[6]

[6] Horty suggests several amendments to his own prioritized default logic, which should carry over without trouble to the present system; these include variable priorities [§5], exclusionary (and even inclusionary, I suspect, although Horty does not discuss them) defaults [§§5, 8.3.3], and different orderings of sets of defaults [§ 8.3.2]. To adopt the last we would probably want to also amend the definition of suboptimality to look instead for a $\mathcal{D}' \subseteq \mathcal{D}$ such that

(i) $\mathcal{D}' \subseteq Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}) - Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}')$ and
(ii) $\mathcal{D}' > Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}') - Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S})$.

4 Some examples

4.1 A simple example

Suppose that you have received three orders: a weak order to do A, a stronger order to do B if you've done A, and an even stronger order to do C if you've done A. Suppose also that you can't do both B and C. We can represent this with the following ordered default theory.

$\delta_1\colon \top \to A$	$\mathcal{S}_1\colon \emptyset$
$\delta_2\colon A \to B$	$\mathcal{S}_2\colon \{\delta_1\}$
$\delta_3\colon A \to C$	$\mathcal{S}_3\colon \{\delta_1, \delta_2\}$
$\mathcal{W}\colon \neg(B \wedge C)$	$\mathcal{S}_4\colon \{\delta_1, \delta_3\}$
$\delta_1 < \delta_2 < \delta_3$	

Given this theory, $\mathcal{S}_1$–$\mathcal{S}_4$ are the only proper scenarios and we have

- $\mathit{Violated}_{\mathcal{W},\mathcal{D}}(\mathcal{S}_1) = \{\delta_1\}$,
- $\mathit{Violated}_{\mathcal{W},\mathcal{D}}(\mathcal{S}_2) = \{\delta_2, \delta_3\}$,
- $\mathit{Violated}_{\mathcal{W},\mathcal{D}}(\mathcal{S}_3) = \{\delta_3\}$, and
- $\mathit{Violated}_{\mathcal{W},\mathcal{D}}(\mathcal{S}_4) = \{\delta_2\}$.

In light of this,

- $\mathcal{S}_2$ is less optimal than $\mathcal{S}_1$ because of both δ_2 and δ_3,
- $\mathcal{S}_2$ is less optimal than $\mathcal{S}_3$ because of δ_2,
- $\mathcal{S}_2$ is less optimal than $\mathcal{S}_4$ because of δ_3,
- $\mathcal{S}_3$ is less optimal than both $\mathcal{S}_1$ and $\mathcal{S}_4$ because of δ_3, and
- $\mathcal{S}_4$ is less optimal than $\mathcal{S}_1$ because of δ_2,

and so $\mathcal{S}_1$ is the only optimal scenario: the present system recommends doing nothing under these circumstances. Horty's system, in contrast, identifies $\mathcal{S}_4$ as ideal, recommending that you do A and C. For Horty, $\mathcal{S}_1$ is unacceptable because you have no excuse for disobeying δ_1, and $\mathcal{S}_4$ is acceptable because while you have disobeyed δ_2, you have an excuse, namely, the stronger δ_3. In his terminology, δ_1 is not defeated in $\mathcal{S}_1$, but δ_2 is defeated in $\mathcal{S}_4$ by δ_3. Here Horty's system diverges from my intuitions, as I find the following hypothetical reasoning much more natural: you should disobey δ_1 because if you follow it, you put yourself in a situation in which you must disobey one of the stronger δ_2 and δ_3.[7]

Before continuing, two brief technical points are worth noting. First, we can see the importance of not requiring proper scenarios to be maximal: by allowing $\mathcal{S}_1$ to be proper, we capture the right sort of hypothetical reasoning for free, without explicitly representing it anywhere. And second, we can see how suboptimality obviates

[7] Horty himself briefly considers hypothetical reasoning of this sort for the Order Puzzle, which I discuss in Sect. 4.2, attributing the idea to Paul Pietroski [p. 205]. He then writes in a footnote, "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" [p. 205, n. 6]. While answering this call was not my initial motivation, it turns out to exactly capture my aim in this paper.

Horty's notion of defeat. In Horty's system, $\mathcal{S}_3$ is not proper because δ_2 is defeated in it. In the present system, the same underlying information makes $\mathcal{S}_3$ less optimal than $\mathcal{S}_4$.

4.2 The Order Puzzle

Suppose now that you have received a different set of three orders: a weak order to do A, a stronger order to do B, and an even stronger order not to do B if you've done A. This is a simplified version of Horty's Order Puzzle [pp. 201ff.], and the present system sees it as nearly identical to the last example; the only real difference is that while there the conflict between the two stronger orders came from $\mathcal{W}$, here $\mathcal{W}$ is empty and the conflict between δ_1 and δ_2 comes from a stronger order.

$\delta_1\colon \top \to A$	$\mathcal{S}_1\colon \emptyset$	$Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}_1) = \{\delta_1, \delta_2\}$
$\delta_2\colon \top \to B$	$\mathcal{S}_2\colon \{\delta_1\}$	$Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}_2) = \{\delta_2, \delta_3\}$
$\delta_3\colon A \to \neg B$	$\mathcal{S}_3\colon \{\delta_2\}$	$Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}_3) = \{\delta_1\}$
$\delta_1 < \delta_2 < \delta_3$	$\mathcal{S}_4\colon \{\delta_1, \delta_2\}$	$Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}_4) = \{\delta_3\}$
	$\mathcal{S}_5\colon \{\delta_1, \delta_3\}$	$Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}_5) = \{\delta_2\}$

Once again, $\mathcal{S}_1$–$\mathcal{S}_5$ are the only proper scenarios; this time, $\mathcal{S}_3$ is the only optimal one. Here too I diverge from Horty's system, which endorses $\mathcal{S}_5$, reasoning that you have no excuse in $\mathcal{S}_3$ to disobey δ_1 but do have an excuse in $\mathcal{S}_5$ to disobey δ_2, namely, the stronger δ_3. Again, the line of reasoning I aim to capture is hypothetical: we *do* have an excuse to disobey δ_1, namely, that we cannot obey it without disobeying one of the stronger δ_2 and δ_3. Given that we're disobeying δ_1, however, we have no reason to disobey δ_2, so we wind up with $\mathcal{S}_3$.

If we add $\neg(A \wedge B)$ to the currently empty $\mathcal{W}$, we have Horty's example of inappropriate equilibria [§8.3.1]. This situation gives Horty's system trouble because while $\mathcal{S}_3$ is intuitively the best option, his system returns both $\mathcal{S}_3$ and $\mathcal{S}_5$ as acceptable. In contrast, the present system treats the situation as exactly the same as the Order Puzzle: $\mathcal{S}_3$ remains the only optimal scenario.

4.3 Combining the examples

For one more illustration we can combine the two examples, beginning with the Order Puzzle and adding two orders that (i) are both stronger than δ_2 and (ii) have conflicting conclusions:

$\delta_1\colon \top \to A$	$\mathcal{W}\colon \neg(C \wedge D)$	$\mathcal{S}_1\colon \{\delta_1, \delta_3\}$
$\delta_2\colon \top \to B$	$\delta_1 < \delta_2 < \delta_3$	$\mathcal{S}_2\colon \{\delta_2, \delta_4\}$
$\delta_3\colon A \to \neg B$	$\delta_2 < \delta_4 < \delta_5$	$\mathcal{S}_3\colon \{\delta_2, \delta_5\}$
$\delta_4\colon B \to C$		$Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}_1) = \{\delta_2\}$
$\delta_5\colon B \to D$		$Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}_2) = \{\delta_1, \delta_5\}$
		$Violated_{\mathcal{W},\mathcal{D}}(\mathcal{S}_3) = \{\delta_1, \delta_4\}$

$\mathcal{S}_1$–$\mathcal{S}_3$ are not the only proper scenarios, but all the others are less optimal than at least one of them. In fact, even $\mathcal{S}_2$ and $\mathcal{S}_3$ are suboptimal—$\mathcal{S}_1$ is the only optimal choice. Although obeying δ_1 still forces us to disobey one of the stronger δ_2 and δ_3, this time, in contrast to the Order Puzzle, we *do* have a reason to disobey δ_2: obeying it forces us to disobey one of the stronger δ_4 and δ_5. As there is no good way to avoid disobeying δ_2, there is no cost to obeying δ_1 and δ_3, and we wind up with $\mathcal{S}_1$.

References

1. Horty, J.F.: Reasons as Defaults. Oxford Univesrity Press, Oxford (2012)