Dynamic Foundations for Deontic Logic*

Malte Willer

Abstract

Several authors have observed that the tools and techniques developed within the field of nonmonotonic logic provide a fruitful framework for the theoretical study of deontic discourse and reasoning. The prominent sources of inspiration for the resulting work in deontic logic are the classical nonmonotonic analyses of reasoning with defeasible generalizations. But while the study of reasoning with defaults may help us understand the nature of prima facie obligations, it arguably does not generalize to address other major sources of nonmonotonicity in deontic discourse and reasoning: the violability of obligations and the sensitivity of obligations to epistemic uncertainty. I demonstrate that the tools and techniques developed within the field of dynamic semantics provide a more comprehensive foundation for deontic logic, the underlying observation being that the semantics of deontic *ought* is sensitive to the presence or absence of epistemic possibilities in discourse and reasoning. The nonmonotonicity of deontic thought and talk, so the key message of this paper, can be illuminated in terms of the familiar nonmonotonicity of epistemic thought and talk that finds a natural articulation in dynamic semantics.

1 The Plot

Several authors—most notably Horty (1994, 1997, 2003, 2007, 2012) but also Asher and Bonevac (1996, 1997), Belzer (1986), Bonevac (1998), McCarty (1994), Nute (1997), and Ryu and Lee (1997)—have observed that the tools and techniques developed within the field of nonmonotonic logic provide a fruitful framework for the theoretical study of deontic discourse and reasoning.[1] The classical motivation for this approach is the idea that certain obligations are only *prima facie* obligations and may thus be *overridden* by other considerations (see Ross 1930). Here is a familiar case. Jones has promised to meet Mary for lunch and so ought to meet her for lunch. But it may very well be that Jones finds himself in a situation in which he would need to miss lunch to save a drowning child, and then he is permitted, arguably even ought, to miss lunch and instead save the drowning child. Such situations, while perfectly ordinary, pose trouble for the standard monotonic approach to deontic reasoning.[2]

*Forthcoming in *Deontic Modality*, ed. N. Charlow and M. Chrisman, Oxford University Press

[1] Some classical frameworks in the nonmonotonic tradition: *model-preference* theories (beginning with McCarthy's (1980) theory of circumscription); *fixed-point* theories (see, e.g., McDermott and Doyle 1980 and Reiter 1980); logics for *argument-based* defeasible reasoning (see Nute 1988, Pollock 1987, and Touretzky 1986 for seminal discussions).

[2] Classical deontic logic goes back to the foundational work in von Wright 1951, 1956. See Åqvist 2002 and McNamara 2006 for overviews.

To see what the issue is, assume that the logical consequence relation is monotonic in the following sense:

$$\textsc{Monotonicity}: \quad \text{If } \phi_1, \ldots, \phi_n \vDash \psi, \text{ then } \phi_1, \ldots, \phi_n, \phi_{n+1} \vDash \psi$$

The initial observation is that (1)–(4) are consistent:

(1) If Jones has promised to meet Mary for lunch, he ought to meet Mary for lunch.

(2) If a drowning child needs Jones's help, then he ought to help the drowning child.

(3) Jones has promised to meet Mary for lunch.

(4) A drowning child needs Jones's help.

Here (1) and (2) are instances of the general principles that one ought to keep one's promises and that one ought to help children in need, respectively. Clearly, (1)–(4) entail (5) and (6):

(5) Jones ought to meet Mary for lunch.

(6) Jones ought to help the drowning child.

This is as it should be, yet monotonicity requires that if (1)–(4) entail (5) and (6), so does the set of premises consisting of (1)–(4) and (7):

(7) Jones would need to miss lunch with Mary to save the drowning child.

And that leaves no room for the intuition that (7) relieves Jones from the obligation to meet Mary for lunch in the sense that breaking his promise is perfectly *permissible* (perhaps even required) if this is what it takes to save the drowning child. One may, of course, say that we should have never subscribed to (1) in the first place since, after all, (7) might come out true. But then it would be just as mistaken to rely on (2) in concluding that Jones ought to save the drowning child—this obligation may be overridden in certain cases as well—and in general it would be hard to see how prima facie obligations could play *any* interesting role in deontic reasoning.

Those sympathetic to Harman's (1986) strict distinction between logic and reasoned change in view may insist that the issue under consideration has nothing to do with principles of semantic entailment and is better left to principles about how to revise one's beliefs in discourse and reasoning. Here I will mostly focus on what form a semantic explanation of the nonmonotonicity of deontic discourse and reasoning *could* take, leaving a discussion of whether it *should* receive such an explanation to the concluding section. But that the overridability of prima facie obligations is of *some* importance for semantic theorizing should be uncontroversial since it underwrites the consistency of (1) and (8), which (by everyone's agreement) any satisfying semantics for *ifs* and *oughts* must explain:

(8) If Jones has promised to meet Mary for lunch but needs to break his promise to save a drowning child, he may/ought to miss lunch with Mary.

Accounting for this fact, it turns out, motivates a departure from the classical analysis of modals and conditionals—one that relies on a proper model of how defeasible norms interact with what is taken for granted in discourse and reasoning to determine what is deontically ideal. Once this is done it is only a small step from the observation that conditionals fail to be monotonic in the antecedent (resist antecedent strengthening) to a nonmonotonic semantic consequence relation.[3]

Previous attempts to account for the nonmonotonicity of deontic discourse and reasoning draw their principle inspiration from the striking parallel between prima facie obligations and defeasible generalizations: the former may be overridden while the latter may be *defeated*.[4] For instance, the assumption that Tweety is a bird together with the general principle that birds fly licenses the inference that Tweety flies, but this inference is only tentative and to be retracted in case we also assume that, say, Tweety is a penguin. The parallel, of course, is that penguins are an exception to the general rule that birds fly just as situations in which one would need to break a promise to save a drowning child are an exception to the general rule that one ought to keep one's promises. Reasoning with obligations, so the hypothesis then goes, is nonmonotonic precisely because certain obligations are alike to defeasible generalizations in that they allow for exceptions, and this is good news since we may now rely on the well-established nonmonotonic analyses of how to reason with rules that allow for exceptions to get the facts about deontic reasoning straight. But the parallel between obligations that may be overridden and defeasible generalizations highlights only one *particular* source of the nonmonotonic nature of deontic reasoning: the phenomenon, or so I shall argue here, is much more general and thus the lessons from default logic, important as they are, provide a too narrow conceptual foundation for exploring the phenomenon of nonmonotonicity in deontic reasoning.

One way to see the point starts with the obvious fact that obligations may not only allow for exceptions but also be outright *violated*. What is more, our misdeeds are alike to exceptions in that they also have the potential to transform certain duties into actions that are wrong. Here is a case that is inspired by Chisholm's (1963) classical discussion of contrary-to-duty obligations: obligations telling us what to do in case we neglect our duties so that we make the best of the bad situations to which our misdeeds have led. Consider the classical inference rule of *deontic detachment* (Greenspan 1975):

$$\text{Deontic Detachment:}\quad (\text{If }\phi)(\text{Ought }\psi), \text{Ought }\phi \models \text{Ought }\psi$$

Deontic detachment licenses the inference of (11) from (9) and (10):

(9) Jones ought to go to the aid of his neighbors.

(10) If Jones goes to the aid of his neighbors, then he ought to tell them he is coming.

(11) Jones ought to tell his neighbors that he is coming.

But suppose that:

(12) Jones does not go to the aid of his neighbors.

Then it may very well be that Jones ought *not* tell his neighbors that he is coming: leaving them hanging is already bad enough, but creating false expectations in addition to that

[3] Charlow (2013a) argues that deontic conditionals such as "If you want to go to Harlem, you ought to take the A-train" exhibit a resistance to antecedent strengthening that is grounded in the goal sensitivity of their consequents and cannot be captured by the classical analysis of modals and conditionals. I set such conditionals aside but everything I am about to say here is sympathetic to Charlow's discussion.

[4] An exception is the proposal by van der Torre and Tan (1998), which resembles mine because of its dynamic spirit but differs substantially in motivation, execution, and scope.

would make things even worse.[5] Once again it looks as if strengthening a set of premises with an extra bit of information blocks an inference that the original set licenses.

Earlier I said that there is a striking parallel between prima facie obligations and defeasible generalizations: the former may be overridden while the latter may be defeated. What we have seen just now is that there is also a striking parallel between exceptional circumstances and violations of obligations: both may transform a duty such as keeping a promise or telling one's neighbors that one is coming into a wrong action. The obvious thing to say then is that reasoning with violable obligations is just as susceptible to monotonicity failures as is reasoning with obligations that allow for exceptions. But while the lessons from default reasoning may very well help us understand why reasoning with prima facie obligations exhibits monotonicity failures, they do not explain why reasoning with violable obligations should exhibit the same phenomenon: it is, after all, a truism that violations of obligations are not exceptional circumstances in which the obligation is no longer binding (Prakken and Sergot 1996, 1997). In particular, a situation in which Jones does not go to the aid of his neighbors is not an exceptional circumstance that relieves him from the duty articulated by (9)—his obligation to go—and it cannot be an exception to the conditional obligation articulated by (10) either since the latter does not even pertain to situations in which Jones does not go. Reasoning with violable obligations has a nonmonotonic flavor alright, but insights from the logic of reasoning with exceptions do not immediately illuminate why this should be so.

Reasoning with obligations under epistemic uncertainty arguably has a nonmonotonic flavor as well, though the case is slightly more involved. Consider the miners paradox from Kolodny and MacFarlane (2010). Ten miners are trapped either in shaft A or in shaft B, but we do not know which one. Water threatens to flood the shafts. We only have enough sandbags to block one shaft but not both. If one shaft is blocked, all of the water will go into the other shaft, killing every miner inside. If we block neither shaft, both will be partially flooded, killing one miner.

Action	if miners in A	if miners in B
Block A	All saved	All drowned
Block B	All drowned	All saved
Block neither shaft	One drowned	One drowned

In this scenario, (13) seems to be the right thing to say since all we know is that (14) is true. However, we are also willing to accept (15) and (16):

(13) We ought to block neither shaft.

(14) Either the miners are in shaft A or they are in shaft B.

(15) If the miners are in shaft A, we ought to block shaft A.

(16) If the miners are in shaft B, we ought to block shaft B.

[5] The intended interpretation of deontic *ought* here is the *deliberative* one in the sense of Thomason 1981: the one that figures prominently in advice and takes the facts as given. The claim that deontic *ought* thus interpreted is subject to monotonicity failures—the advice we give to Jones is not preserved under information strengthening—is compatible with the intuition that it also has interpretations less sensitive to what is taken for granted.

(13)–(16) seem consistent. But let Γ be the set of premises consisting of (13)–(16) and observe that (13) intuitively entails both (17) and (18):

(17) It is not the case that we ought to block shaft A.

(18) It is not the case that we ought to block shaft B.

(17) and (18) just negate the consequents of the conditionals in (15) and (16), respectively, and so given monotonicity and modus ponens, $\Gamma \cup \{inA\} \models \perp$ and hence $\Gamma \models \neg inA$ by reductio. For parallel reasons, $\Gamma \models \neg inB$. Hence Γ is inconsistent after all since it entails the negation of (14).

The paradoxical argument just given appeals to a monotonic conception of logical consequence—specifically, it assumes that if Γ entails (13), then so do $\Gamma \cup \{inA\}$ and $\Gamma \cup \{inB\}$—and so one might hope for a nonmonotonic escape route. But reasoning with obligations under epistemic uncertainty is not another instance of reasoning with obligations that allow for exceptions. For suppose that the miners' being in shaft A is just a case in which the obligation to block neither shaft is no longer in force: then so would be a case in which the miners are in shaft B, and since the miners are in one of the two shafts we know that we are facing an exceptional situation and thus the obligation to block neither shaft should not hold in the first place.

One reaction to the issues I have just outlined is to say that nonmonotonicity in deontic reasoning is indeed limited to cases in which prima facie obligations are in play: Chisholm's scenario and the miners paradox highlight other problems with the classical conception of deontic logic.[6] But another reaction—the one I will pursue here—is to say that nonmonotonicity in deontic reasoning is a pervasive phenomenon that does not reduce to reasoning with rules allowing for exceptions.[7] This reaction is appealing for reasons other than theoretical simplicity. The key conceptual driver behind nonmonotonic logic is the idea that the validity of an inference is not only sensitive to the *presence* of information—as in classical logic—but also to the *absence* of information, and it does not take much to apply this idea to the cases just discussed. Prima facie obligations are binding unless there are exceptional circumstances, violable obligations may entail other duties unless they are violated, and obligations under epistemic uncertainty may no longer hold in case the underlying uncertainty is resolved by additional information. In brief, while the cases discussed are at some level of description quite disparate, there is more than a merely initial appeal to the idea that they also have something important in common that a nonmonotonic analysis of deontic logic is ideally suited to capture.

What is needed, then, is a comprehensive perspective on deontic discourse and reasoning that captures the variety of ways in which deontic inference is sensitive to the absence

[6]Rejecting deontic detachment blocks Chisholm's paradox but the prominent proposals by Lewis (1973) and Kratzer (1991, 2012) preserve the validity of this rule (though they deny modus ponens for deontic conditionals, which is sometimes labeled "factual detachment"). Appealing to tense is another option but does not help with "timeless" Chisholm scenarios (see Carmo and Jones 2002 and McNamara 2006 and references therein). Rejecting modus ponens is Kolodny and MacFarlane's preferred solution to their miners paradox, but Willer (2012) argues that this is an overreaction and also says a bit about Bledin's (2014) and Yalcin's (2012) suggestion to respond to the paradox by rejecting reductio (see also §5 for additional remarks). Silk (2013) critically discusses Dowell's (2012) and von Fintel's (2012) suggestion that the deontic modal *ought* receives different interpretations in (13) and (14)/(15), respectively.

[7]See also the discussion by van der Torre and Tan (1997), who argue that the violability of obligations should be understood as just such a kind of defeasibility, though they focus on a quite narrow conception of defeasibility as failure of antecedent strengthening for deontic conditionals.

of information. Such a perspective, I maintain, arises in a dynamic semantic analysis of *ifs* and *oughts* that, while being non-classical, incorporates ideas from the familiar linguistics literature by analyzing deontic modals and conditionals using possible worlds and in light of a context determining what is deontically ideal.[8] The resulting framework is flexible enough to allow for the interpretation of deontic *ought* in light of contextually provided norms that may be overridable, violable, or information-sensitive. While the potential overridability, violability, and information-sensitivity of norms are distinct features that are captured by different aspects of the formal apparatus, the semantic framework also highlights a common denominator between them: they all induce deontic commitments that carry in their wake distinct epistemic commitments and since the latter are defeasible, so are the latter. More precisely, in the framework I propose the previously observed sensitivity of deontic inference to the absence of information uniformly manifests in entailment relations between deontic *ought* and epistemic *might*. Once we have explained why epistemic *might* allows for monotonicity failures—a fact that finds a natural explanation in dynamic semantics—we also know why deontic *ought* does so as well.

My strategy is straightforward. The next §2 briefly outlines a basic dynamic semantics for epistemic modals and conditionals, highlights its nonmonotonic outlook on logical consequence, and briefly explains how this outlook may serve as a comprehensive foundation for a nonmonotonic perspective on deontic discourse and reasoning. §3 offers a simple dynamic semantics for deontic *ought* that makes sense of reasoning with prima facie obligations and with obligations under epistemic uncertainty by highlighting its sensitivity to epistemic possibilities. §4 extends the simple proposal to avoid a few shortcomings when it comes to the analysis of contrary-to-duty obligations.[9] §5 concludes the discussion by briefly comparing the story told here with alternative views that treat nonmonotonic effects as a topic for belief revision theory rather than semantics.

2 Basics

A basic dynamic story about conditionals and epistemic modals has been told before (see Veltman 1996 and Gillies 2004 for seminal discussion), and so I will just briefly highlight its key ideas and explain why they matter for current purposes. Semantic values are stated in terms of context change potentials (CCPs): functions that take an information state as input and return another state—the result of *updating* the input state with the sentence under consideration—as output.[10] Thinking of semantic values in such a way

[8]See Lewis 1974 for an overview of early studies of deontic logic within possible world frameworks (including his own proposal in Lewis 1973). In the linguistics literature, the work by Kratzer (see Kratzer 2012 and references therein) is seminal.

[9]Nothing I am about to say here *by itself* enforces the validity of deontic detachment (which, recall, figures prominently in Chisholm's paradox about contrary-to-duty obligations) but the upcoming story can be modified so that it does. Motivating the required twists and turns in detail and explaining why they do the trick goes beyond the scope of this paper—in brief, what is needed are some plausible restrictions on the notion of a deontic context (§3.2) and some minimal modifications of the notion of retraction (§4)—and so I will focus here on the (no less important) task of telling a story about Chisholm's case that is *compatible* with the assumption that deontic detachment is valid and explains the data from a nonmonotonic perspective. See Willer 2014 for a detailed discussion of deontic detachment that lives happily with everything said in this paper but remains silent on many of the issues that matter here.

[10]Classical dynamic frameworks include *Discourse Representation Theory* (see Kamp et al. 2011 for an up-to-date discussion), *Dynamic Predicate Logic* (Groenendijk and Stokhof 1991), *File Change Semantics* (Heim 1982), and *Update Semantics* (Veltman 1985, 1996).

owes some inspiration to the classical story about assertion from Stalnaker (1978), who observes that context-content interaction is not a one-way road: context affects what proposition an assertion expresses, but an assertion in turn affects the context by adding the proposition expressed to the common ground. In Stalnaker's picture the latter effect is a pragmatic afterthought to classical truth-conditional semantics, but here the facts about context-content interaction are integral to how we model semantic values.

Thinking of semantic values dynamically as CCPs motivates an equally dynamic conception of logical consequence as guaranteed preservation of rational commitment. On this view, an argument is valid just in case its conclusion is accepted by any carrier of information σ under the supposition of its premises:

$$\phi_1, \ldots, \phi_n \vDash \psi \text{ just in case for all } \sigma\text{: } \sigma[\phi_1]\ldots[\phi_n] \vDash \psi$$

Here $\sigma[\phi]$ is the result of strengthening σ with the information carried by ϕ, and to say that $\sigma \vDash \phi$ is to say that σ is rationally committed to ϕ. The question whether logical consequence is nonmonotonic now hinges on the question whether additional premises in discourse and reasoning are guaranteed to preserve existing rational commitments and, what is more, considerations about epistemic *might* strongly suggest that information agglomeration fails to be commitment preserving in such a way: the belief that, say, Mary might be in Chicago needs to go once one learns that Mary is not in Chicago (see Levi 1988, Fuhrmann 1989, and Rott 1989 for seminal discussion). More generally:

$$\sigma \vDash \textsc{Might}\ \phi \not\rightarrow \sigma[\neg\phi] \vDash \textsc{Might}\ \phi$$

A state may be committed to $\ulcorner\textsc{Might}\ \phi\urcorner$ but fail to be thus committed if strengthened with $\ulcorner\neg\phi\urcorner$. And that already leads to monotonicity failures given the following minimal assumptions:

Acceptance:	For all σ: $\sigma \vDash \phi$ just in case $\sigma[\phi] = \sigma$
Epistemic Success:	For all σ: $\sigma[\textsc{Might}\ \phi] \vDash \textsc{Might}\ \phi$

The first principle says that ϕ is accepted in σ just in case σ already carries the information carried by ϕ. The second principle requires that the result of updating some carrier of information with an epistemic modal claim accepts that claim. It follows immediately that $\textsc{Might}\ \phi \vDash \textsc{Might}\ \phi$ for all ϕ, and we also know that for some σ and ϕ, $\sigma \vDash \textsc{Might}\ \phi$ yet $\sigma[\neg\phi] \nvDash \textsc{Might}\ \phi$ and thus $\sigma[\textsc{Might}\ \phi][\neg\phi] \nvDash \textsc{Might}\ \phi$. Accordingly, $\textsc{Might}\ \phi \vDash \textsc{Might}\ \phi$ yet $\textsc{Might}\ \phi, \neg\phi \nvDash \textsc{Might}\ \phi$ and thus the proposed notion of logical consequence is nonmonotonic.

To make all of this more precise, define the initial target of our semantic analysis as a simple propositional language extended with the conditional connective *if* ($\Rightarrow$) and epistemic *might* ($\Diamond_e$):

Definition 1 (Basic Language) $\mathcal{L}$ is the smallest set that contains a set of atomic sentences $\mathcal{A} = \{p, q, r, \ldots\}$ and is closed under negation ($\neg$), conjunction ($\wedge$), epistemic *might* ($\Diamond_e$), and the conditional connective *if* ($\Rightarrow$). Inclusive and exclusive disjunction ($\vee$, $\dot\vee$), the material conditional ($\supset$), the biconditional ($\equiv$), and epistemic *must* ($\Box_e$) are defined in the usual way. $\mathcal{L}_0$ is defined as the propositional fragment of $\mathcal{L}$.

Semantic values take a state as input and return one as output. Such carriers are sets of indices consisting of a possible world coupled with a *frame* (more on frames momentarily):

Definition 2 (Information States) σ is an *information state* iff $\sigma \subseteq I$, where I is the set of all *indices*. w_i is the *possible world parameter* of some index i, where w is a possible world iff $w\colon \mathcal{A} \mapsto \{0,1\}$ and W is the set of all possible worlds. Σ is the set of all information states. σ_0 is the *initial* information state and identical with I.

Information states are sets of indices, and to avoid the need for redefinitions at a later stage we will not simply identify an index with a plain possible world, though this would be good enough if we were only interested in the semantics for $\mathcal{L}$.

Recursively define the update rules for elements of $\mathcal{L}$ as follows:

Definition 3 (Basic Update Rules) Associate with each $\phi \in \mathcal{L}$ a context change rule $[\phi]\colon \Sigma \mapsto \Sigma$ as follows:

(1) $\sigma[p] = \{i \in \sigma\colon w_i(p) = 1\}$

(2) $\sigma[\neg\phi] = \sigma \backslash \sigma[\phi]$

(3) $\sigma[\phi \wedge \psi] = \sigma[\phi][\psi]$

(4) $\sigma[\phi \Rightarrow \psi] = \{i \in \sigma\colon \sigma[\phi][\psi] = \sigma[\phi]\}$

(5) $\sigma[\Diamond_e \phi] = \{i \in \sigma\colon \sigma[\phi] \neq \varnothing\}$

According to (1), updating σ with an atomic sentence eliminates all indices with a world-parameter at which that sentence is false. Updating σ with $\ulcorner\neg\phi\urcorner$ comes down to subtracting from σ the result of updating σ with ϕ (cf. (2)). The clause in (3) requires that updating with a conjunction amounts to an update with the first conjunct followed by an update with the second conjunct. The clause in (4) takes Ramsey's (1931) test procedure for conditionals as a guide to the semantics of *if*. Ramsey suggested that a conditional is to be accepted just in case its consequent is (hypothetically) accepted under the supposition of its antecedent, and here we say that a conditional tests whether the input state σ has a certain informational structure: once we have updated σ with the antecedent, updating with the consequent idles, which is just to test whether the consequent is accepted under the supposition of the antecedent. If this is so, the test is passed and the conditional is accepted; otherwise, the conditional is rejected. Finally, the meaning of epistemic *might* is given in terms of acceptance conditions as well: it tests whether its prejacent is compatible with the input state.

With each element of $\mathcal{L}_0$ we may associate a proposition in the familiar way:

Definition 4 (Propositions) The function $[\![\cdot]\!]\colon \mathcal{L}_0 \mapsto W$ assigns to each $\phi \in \mathcal{L}_0$ a *proposition* understood as a set of possible worlds:

(1) $[\![p]\!] = \{w \in W\colon w(p) = 1\}$

(2) $[\![\neg\phi]\!] = W \backslash [\![\phi]\!]$

(3) $[\![\phi \wedge \psi]\!] = [\![\phi]\!] \cap [\![\psi]\!]$

Propositions will not play their classical role as carriers of meaning but be useful at a later stage.

The following definition just restates the dynamic conception of logical consequence that was mentioned earlier but also says what it takes for a commitment to be defeasible.

Definition 5 (Acceptance, Defeasibility, Entailment) Consider any $\sigma \subseteq \Sigma$ and $\phi_1, \ldots \phi_n, \psi \in \mathcal{L}$:

1. σ *accepts* ϕ, $\sigma \models \phi$, iff $\sigma[\phi] = \sigma$
2. σ *defeasibly* accepts ϕ iff $\sigma \models \phi$ and for some $\sigma' \subseteq \sigma$: $\sigma' \not\models \phi$
3. ψ *defeats* a commitment to ϕ in σ iff $\sigma \models \phi$ but $\sigma[\psi] \not\models \phi$
4. $\phi_1, \ldots \phi_n$ *entails* ψ, $\phi_1, \ldots \phi_n \models \psi$, iff for all $\sigma \in \Sigma$: $\sigma[\phi_1] \ldots [\phi_n] \models \psi$

A state σ accepts ϕ just in case updating σ with ϕ idles. This commitment is defeasible just in case there is some strengthening of σ that no longer accepts ϕ, and ψ is a defeater just in case updating σ with ψ results in such a state. An argument is valid just in case any information state that is updated with its premises is committed to its conclusion.

It will be helpful to highlight three facts about the outlined framework, the first one being that modus ponens is valid:

Fact 1 $\phi \Rightarrow \psi, \phi \models \psi$

Take any σ and consider $\sigma[\phi \Rightarrow \psi][\phi]$: if $\sigma[\phi \Rightarrow \psi] = \emptyset$, the conclusion is accepted for trivial reasons. Otherwise, $\sigma[\phi \Rightarrow \psi] = \sigma$ and $\sigma[\phi] \models \psi$, which just proves the point.

Second, observe that the dynamic logical consequence relation is nonmonotonic:

Fact 2 For all contingent $\phi \in \mathcal{L}_0 : \Diamond_e\phi \models \Diamond_e\phi$ and $\Diamond_e\phi, \neg\phi \not\models \Diamond_e\phi$.

The reason is familiar: *might*-commitments are not guaranteed to be preserved as new information is acquired, and in particular a commitment to $\Diamond_e p$ fails to be preserved by the information that p is not the case.

The third observation of relevance is that the defeasibility of epistemic *might* may enforce the defeasibility of other rational commitments:

Fact 3 Consider any $\sigma \in \Sigma$ such that $\sigma \models \phi$ and suppose that $\phi \models \psi$: if χ defeats a commitment to ψ in σ, then χ defeats a commitment to ϕ in σ.

This is just to say that whenever a commitment to ϕ brings in its wake a commitment to ψ, then the commitment to ϕ is defeasible in case the commitment to ψ is. And that is important since there is every reason to think that commitments to deontic *ought* frequently bring in their wake a commitment to epistemic *might*: the defeasibility of deontic commitments can be illuminated in terms of the by now established dynamic defeasibility to certain epistemic commitments. Let us go through the details.

3 Simple Oughts

This section presents a simple dynamic framework that makes sense of reasoning with prima facie obligations but is also flexible enough to make sense of some other nonmonotonic effects in deontic reasoning. Doing so requires that we offer a dynamic analysis of the deontic modal *ought* ($\Box_d$) as well as of a defeasible conditional connective ($>$) that will be used in stating prima facie obligations. The extended target language is defined as follows:

Definition 6 (Extended Language) Define a language $\mathcal{L}_1$ so that $\mathcal{L}_0 \subseteq \mathcal{L}_1$ and whenever $\phi, \psi \in \mathcal{L}_0$, then $\Box_d \phi \in \mathcal{L}_1$ and also $\phi > \psi \in \mathcal{L}_1$. $\mathcal{L}^+$ is the smallest set that contains all elements of $\mathcal{L}_1$ and is closed under negation ($\neg$), conjunction ($\wedge$), epistemic *might* ($\Diamond_e$), and the conditional connective *if* ($\Rightarrow$). Other connectives are again defined as usual.

Earlier I stated the semantics of epistemic modals and conditionals in terms of acceptance conditions, and it makes good sense to do the same when it comes to deontic *ought*. The proposal is that deontic *ought* is accepted just in case its prejacent is accepted in the state got by focussing on those indices that accord best with the rules articulating what is deontically ideal, where those rules are determined by context and may very well be defeasible. The idea will be slightly refined at a later stage, but it is good enough to illustrate why the setup pursued here has some real promise.

Definition 7 (Deontic Ought (Basic Version)) Extend the update rules for $\mathcal{L}$ with the following entry:

(6) $\sigma[\Box_d \phi] = \{i \in \sigma : \sigma_d \Vdash \phi\}$

Here σ_d is the set of indices compatible with σ that are deontically ideal in light of some deontic context d, and $\sigma \Vdash \phi$ holds whenever σ *expects* ϕ to be true: ϕ is true in every possible world that is compatible with σ and accords best with the defeasible generalizations accepted by that state. I will first say more about defeasible generalizations (§3.1) and then explain what a deontic context does (§3.2).

3.1 Defeasible Rules

Defeasible generalizations articulate what to expect in case certain conditions are satisfied, and since we do not care too much about sentential structure we can make this more precise by saying that such conditionals select a set of default worlds from the possibilities depicted by the antecedent. Taking some inspiration from Veltman (1996) the semantic analysis is provided using frames:

Definition 8 (Frames) A *frame* π maps each *scenario* $s \subseteq W$ to a subset of s. $[\![\phi]\!]$ is a *default* in $\pi(s)$ iff $\pi(s) \subseteq [\![\phi]\!]$. w is a *normal* world in $\pi(s)$, $w \in \mathcal{N}\pi(s)$, iff $w \in s$ and for all $s' \subseteq s$: if $w \in s'$, then $w \in \pi(s')$. π is *coherent* iff for every nonempty $s \subseteq W$, there is a normal world in $\pi(s)$. Π is the set of all such coherent π's.

Frames allow us to say, for instance, that it normally rains but there is an *exception:* if there is an easterly wind, it normally does not rain. If this is all we want to say, we have $\pi(W) = [\![\text{RAIN}]\!]$, $\pi([\![\text{EAST}]\!]) = [\![\text{EAST}]\!] \backslash [\![\text{RAIN}]\!]$, and $\pi(s) = s$ for all other scenarios. Consider the following distribution of truth-values across possible worlds:

	RAIN	EAST
w_1	T	T
w_2	T	F
w_3	F	T
w_4	F	F

Then $\pi(W) = \{w_1, w_2\}$ and $\pi([\![\text{EAST}]\!]) = \{w_3\}$. Since w_1 violates the expectations that come with an easterly wind—$w_1 \in [\![\text{EAST}]\!]$ yet $w_1 \notin \pi([\![\text{EAST}]\!])$—it does not count as a normal world. In contrast, w_2 does not violate any of our expectations—notice that $w_2 \notin [\![\text{EAST}]\!]$ and so even though it rains at w_2, no expectation is violated—and so $\mathcal{N}\pi(W) = \{w_2\}$: normally, it rains and there is no easterly wind. But assuming that there is an easterly wind, we expect no rain since $\mathcal{N}\pi([\![\text{EAST}]\!]) = \{w_3\}$.

We may now make the notion of an index precise and determine the update rule for the defeasible conditional connective:

Definition 9 (Indices) i is an index iff $i \in W \times \Pi$, that is, an index is a pair consisting of a possible world w and a coherent frame π. π_i is the *frame parameter of i*.

Definition 10 (Defeasible Conditional Connectives) Extend the previous update rules for $\mathcal{L}$ and deontic *ought* with the following entry:

(7) $\sigma[\phi > \psi] = \{i \in \sigma : \pi_i([\![\phi]\!]) \subseteq [\![\psi]\!]\}$

An information state keeps track of the hard information modeled by a set of possible worlds but also of what is to be expected, which is represented by a set of coherent frames. A formula of the form $\ulcorner \phi > \psi \urcorner$ eliminates from an information carrier all those indices whose frame parameter fails to treat $[\![\psi]\!]$ as a $[\![\phi]\!]$-default.

Veltman (1996) offers a model for how to reason with defeasible generalizations in dynamic semantics, but for our purposes there is no need to repeat his story here. I am first and foremost interested in the role of prima facie obligations for deontic discourse and reasoning, and for this I need to say precisely what is expected by a carrier of information. As a preparation, observe that we can associate with each such carrier a unique scenario and frame as follows:

Definition 11 (Depicted Scenarios and Frames) Consider arbitrary $\sigma \in \Sigma$:

1. The *scenario depicted by* σ is defined as $s_\sigma = \{w_i : i \in \sigma\}$.
2. The *frame depicted by* σ is defined so that $[\![\phi]\!]$ is a default in $\pi_\sigma(s)$ iff for all $i \in \sigma$: $[\![\phi]\!]$ is a default in $\pi_i(s)$.

The scenario depicted by σ is simply the set of possible worlds compatible with the hard information carried by σ. And σ treats $[\![\psi]\!]$ as a $[\![\phi]\!]$-default just in case σ accepts the defeasible generalization articulated by $\ulcorner \phi > \psi \urcorner$.

Suppose then that one accepts a set of default rules. The simple intuition is that a default rule should play a role in forming expectations just in case the rule is *triggered*—one's information entails the premise of the default rule—and at the same time *undefeated*—one's information does not depict an exceptional scenario in which the rule is no longer binding. The only wrinkle we have to add to the story is that default rules may sometimes lead to *conflicting* expectations, and so we must define what it takes for a set of default rules to be unconflicted in expectation forming.[11] The key notions involved

[11]My story about how to form expectations on the basis of accepted defaults owes inspiration to Veltman's (1996) and Horty's (2012) accounts. But the difference beween the upcoming proposal and theirs are substantial, not least because it offers a possible world analysis of the defeasible conditional connective (thus differing from Horty's) and ties the applicability of a default rule to a triggering condition (thus differing from Veltman's).

in this characterization of how expectations are formed on the basis of accepted default rules can be made more precise as follows:

Definition 12 (Triggers, Defeats, Conflicts) Consider arbitrary $\pi \in \Pi$, $\sigma \in \Sigma$, and say that a world w *complies* with a $[\![\phi]\!]$-default $[\![\psi]\!]$ iff $w \notin [\![\phi \wedge \neg\psi]\!]$:

1. A $[\![\phi]\!]$-default $[\![\psi]\!]$ is *triggered* in σ iff $\sigma \vDash \phi$.
2. A $[\![\phi]\!]$-default $[\![\psi]\!]$ is *undefeated* in σ iff for every s such that $s_\sigma \subseteq s$: there is some $w \in \mathcal{N}\pi(s)$ that complies with the $[\![\phi]\!]$-default $[\![\psi]\!]$.
3. A set of defaults is *unconflicted* in σ iff for every s such that $s_\sigma \subseteq s$: there is some $w \in \mathcal{N}\pi(s)$ that complies with each of its members.

To say that a world complies with a default rule is just to say that it does not violate the expectation articulated by that rule. For a default rule to be triggered in σ, its condition must be entailed by the scenario depicted by σ. But the rule is defeated in σ if another, more specific rule is triggered that conflicts with it, and a set of default rules is conflicted in σ just in case no normal world in the scenario depicted complies with each of its members. Note here that whenever a single default rule is defeated in σ, no set of rules containing it can be unconflicted in σ.

To give a very simple example of the notion of defeat, go back to the case in which we believe that it normally rains but if there is an easterly wind, the weather is normally dry: $[\![\textsc{Rain}]\!]$ is W-default but $[\![\neg\textsc{Rain}]\!]$ is an $[\![\textsc{East}]\!]$-default. Then intuitively a scenario in which the wind comes from the east is an exception to the rule that it normally rains, and this is just what the framework predicts. For suppose σ accepts that the wind comes from the east: then $s_\sigma \subseteq [\![\textsc{East}]\!]$ and of course every normal $[\![\textsc{East}]\!]$-world is one at which the weather is dry. So no normal $[\![\textsc{East}]\!]$-world complies with the W-default $[\![\textsc{Rain}]\!]$, which is just to say that the default rule is defeated in σ.

To illustrate the possibility of conflicting expectations, and to go through an example that will be of relevance in the upcoming discussion, start with the initial information carrier σ_0 and consider the state $\sigma = \sigma_0[\textsc{Promise} > \textsc{Meet}][\textsc{Need} > \textsc{Help}][\textsc{Promise} \wedge \textsc{Need}]$, with the atomic sentences being translated as follows:

promise:	Jones has promised to meet Mary for lunch.
meet:	Jones will meet Mary for lunch.
need:	The drowning child needs Jones's help.
help:	Jones will help the drowning child.

Then clearly σ accepts that Jones will, all things being equal, meet Mary for lunch if he promised to do so and that he will, all things being equal, help a child if it is in need of his help. Not surprisingly, both default rules are triggered and undefeated in the scenario we consider—that Jones has promised to meet Mary for lunch and that there is a drowning child in need of Jones's help. The worlds compliant with both rules are just those at which Jones keeps his promise *and* saves the drowning child.

Things become more interesting if we assume that Jones cannot do both, which I will do here by considering $\tau = \sigma[\textsc{meet} \dot{\vee} \textsc{help}]$.[12] Both default rules remain triggered and

[12]This is simplifying a bit: it is desirable to distinguish the claim that John makes a choice between keeping his promise and saving the child—which is what "$\textsc{meet} \dot{\vee} \textsc{help}$" literally says—from the stronger

undefeated—the scenario depicted by τ is not an *exception* to any of those rules—but they conflict with each other: no world in the scenario under consideration complies with both expectations. For at any such world, Jones does not meet Mary for lunch even though he promised to do so—thus violating the expectation that he will keep his promise—or fails to help the child even though it is drowning—thus violating the expectation that he helps the child in need. It makes sense to expect that Jones will at least keep his promise or save the drowning child (rather than doing neither) but there is, without additional assumptions about Jones's character anyway, no reason to favor one default rule over the other in thinking about what he is going to do in the scenario under consideration.

Now suppose we want to say that in Jones's case, the default rules are not of equal strength: Jones will not let the child in need drown just to keep his promise to meet Mary for lunch. One way to go is to appeal to a priority relation between default rules (see Horty 2012 for recent discussion) but here I exploit the already existing feature that more specific rules trump less specific ones in case of a conflict. For consider $\chi = \tau[((\textsc{promise} \wedge \textsc{need}) \wedge (\textsc{meet} \dot{\vee} \textsc{help})) > \textsc{help}]$, that is, the result of strengthening τ with the additional default rule that Jones will help the child in need if faced with the unhappy choice between doing so and keeping his promise. This update would have no interesting effect if it were epistemically possible for Jones to both keep his appointment and help the drowning child, but it makes a big difference in the case under consideration.

The crucial observation about χ is that it differs from τ in treating a scenario in which Jones must choose as an *exception* to the rule that, all things being equal, he will meet Mary for lunch if he has promised to do so. To see this, simply observe that a world now counts as normal in $\pi_\chi((\llbracket \textsc{promise} \wedge \textsc{need}) \wedge (\textsc{meet} \dot{\vee} \textsc{help})\rrbracket)$ only if John helps the drowning child—and thus breaks his promise to meet Mary for lunch— at that world. No such world can thus comply with the $\llbracket\textsc{promise}\rrbracket$-default $\llbracket\textsc{meet}\rrbracket$, which just means that the default rule is defeated in the scenario under consideration. Since the other default rules remain triggered and undefeated, we expect that Jones will help the drowning child.

Much more could be said about how we form expectations in everyday discourse and reasoning, but the emerging picture is good enough for our purposes. What matters here is that the role of a default rule in expectation forming is sensitive to the presence or absence of global features of an information state that is not preserved under the process of strengthening. Accordingly, expectations may be defeated as new information comes into view, and the following definition makes it easier to see why this is so.

Definition 13 (Expectations) Consider any $\sigma \subseteq \Sigma$ depicting a frame π:

1. $i \in \sigma_o^0$ iff $i \in \sigma$ and w_i complies with a maximal set of π-defaults such that (i) the set is unconflicted in σ and (ii) each member is triggered in σ.
2. $i \in \sigma_o^{n+1}$ iff $i \in \sigma_o^n$ and w_i complies with a maximal set of π-defaults such that (i) the set is unconflicted in σ_o^n and (ii) each member is triggered in σ_o^n.
3. The set of *optimal* indices in σ is defined as $\sigma_o = \bigcap_{n \geqslant 0} \sigma_o^n$.
4. σ *expects* ϕ, $\sigma \Vdash \phi$, iff $\sigma_o \vDash \phi$.

To say that ϕ is expected in σ is to say that ϕ is accepted by the state got by focussing on the optimal indices in σ. The optimal indices are defined in a step-wise fashion: first,

one that Jones must make this choice. Here and throughout it will no do harm to avoid additional complications and simply assume that Jones makes the choice only if he cannot do both.

determine σ_o^0, that is, the indices whose world-parameter complies with any maximal unconflicted set of triggered default rules in σ. Then determine σ_o^1, that is, the indices in σ_o^0 that come with a world-parameter complying with any maximal unconflicted set of triggered default rules in σ_o^0, and so on. The set of optimal indices—defined as $\bigcap_{n \geqslant 0} \sigma_o^n$—is just the state of information that the deliberating agent will arrive at after carrying out the reasoning process indefinitely.

Whenever the applicability of a default rule is defeasible, so are expectations formed on the basis of that rule. And the special case that interests us is the one we considered earlier: given certain expectation patterns, the $[\![\textsc{promise}]\!]$-default $[\![\textsc{meet}]\!]$ is defeated in some carrier of information σ such that $\sigma \vDash \textsc{need}$ unless for some $i \in \sigma$, $w_i \in [\![\textsc{meet} \wedge \textsc{help}]\!]$. Specifically:

Fact 4 Take any σ so that $\pi_\sigma = \pi_\chi$: if $\sigma \Vdash \textsc{meet}$, then $\sigma \vDash \Diamond_e(\neg\textsc{need} \vee (\textsc{meet} \wedge \textsc{help}))$

Expectations that derive from defeasible generalizations, in short, may be sensitive to the presence or absence of certain epistemic possibilities, and this is just how things should be. This is the moral we need to make sense of the nonmonotonicity of thought and talk about prima facie obligations. The purpose of the next section is to explain why.

3.2 Deontic Ought

Earlier I said that deontic *ought* is accepted just in case its prejacent is optimal in light of some deontic context. It is a familiar idea from the linguistics literature that a deontic context determines what is deontically ideal by fixing an ordering source. Here I will adopt this idea but not think of ordering sources classically as sets of propositions but dynamically as sets of context change potentials.

Definition 14 (Deontic Contexts) A deontic context d determines for each $\sigma \in \Sigma$ a deontically ideal state σ_d by providing an ordering source $o \subseteq \Sigma^\Sigma$, that is, a set of CCPs. o_d is the ordering source provided by d. Given some $\sigma \in \Sigma$, $\sigma_\Delta = \{i \in \sigma_0 \colon \exists i' \in \sigma.\, w_i = w_{i'}\}$ and $i <_d^\sigma i'$ iff $i, i' \in \sigma_\Delta$ and (i) for all $[\phi] \in o_d$: if $i' \in \sigma[\phi]$, then $i \in \sigma[\phi]$ and (ii) for some $[\phi] \in o_d$: $i \in \sigma[\phi]$ and $i' \notin \sigma[\phi]$. i is *minimal* in σ given d iff $\neg\exists i' \colon i' <_d^\sigma i$. σ_d is then just the set of indices that are minimal in σ given d.[13]

Whenever the ordering source exclusively consists of elements of $\mathcal{L}_0$, we may just as well have propositions fix what is deontically ideal. The current framework is more flexible, however, and in particular we will put the possibility of stating ordering sources using epistemic modals and defeasible conditionals to good use. Let me explain.

Start with Jones, who has promised to meet Mary for lunch but also stumbles upon a drowning child. Suppose that the deontic context fixes the following ordering source:

$$o_d = \{[\textsc{promise} > \textsc{meet}], [\textsc{need} > \textsc{help}], [((\textsc{promise} \wedge \textsc{need}) \wedge (\textsc{meet} \dot{\vee} \textsc{help})) > \textsc{help}]\}$$

Here the first member of o_d articulates the rule that Jones's promise to meet Mary for lunch creates the prima facie obligation to meet her for lunch, while the second says that

[13] A disclaimer with a familiar ring: analyzing deontic *ought* as a quantifier over a set of possible worlds that are minimal in σ given d is technically convenient but requires that there is always such a set. The classical analogue is the Limit Assumption (Lewis 1973).

a child's need for help creates the prima facie obligation to help that child. The third rule effectively resolves a potential conflict between the first and second prima facie obligation in favor of helping the child in need.

Take any $\sigma \in \Sigma$ and let $\pi = \pi_{\sigma_d}$: then $\pi(\llbracket \textsc{promise} \rrbracket) = \llbracket \textsc{promise} \wedge \textsc{meet} \rrbracket$, $\pi(\llbracket \textsc{need} \rrbracket) = \llbracket \textsc{need} \wedge \textsc{help} \rrbracket$, $\pi(\llbracket (\textsc{promise} \wedge \textsc{need}) \wedge (\textsc{meet} \dot{\vee} \textsc{help}) \rrbracket) = \llbracket (\textsc{promise} \wedge \textsc{need}) \wedge (\neg\textsc{meet} \wedge \textsc{help}) \rrbracket$, and $\pi(s) = s$ for all other scenarios. The good news is that we already know what it takes for it to be the case that $\sigma_d \Vdash \textsc{meet}$ from the earlier discussion in §3.1: of course, it is required that $\sigma \vDash \textsc{promise}$, but for the $\llbracket \textsc{promise} \rrbracket$-default $\llbracket \textsc{meet} \rrbracket$ to apply in σ_d, it must *also* be the case that $\sigma \nvDash \textsc{need} \wedge \neg(\textsc{meet} \wedge \textsc{help})$—otherwise, the $\llbracket \textsc{promise} \rrbracket$-default $\llbracket \textsc{meet} \rrbracket$ is defeated. And that is just to say that a commitment to (1) needs to go in light of the additional information that (i) there is a drowning child in need of Jones's help and (ii) Jones cannot help that child *and* keep his lunch appointment. It follows immediately that (1) and (3) entail (5) but no longer do so if strengthened by the information carried by (7). Here is a summary of the predictions:

Fact 5 Consider d as fixed for Jones's scenario:

1. $\textsc{promise} \Rightarrow \Box_d\textsc{meet}, \textsc{promise} \vDash \Box_d\textsc{meet}$
2. $\textsc{promise} \Rightarrow \Box_d\textsc{meet} \vDash \Diamond_e(\neg\textsc{need} \vee (\textsc{meet} \wedge \textsc{help}))$
3. $\textsc{promise} \Rightarrow \Box_d\textsc{meet}, \textsc{promise}, \textsc{need} \wedge (\textsc{meet} \dot{\vee} \textsc{help}) \nvDash \textsc{promise} \Rightarrow \Box_d\textsc{meet}$
4. $\textsc{promise} \Rightarrow \Box_d\textsc{meet}, \textsc{promise}, \textsc{need} \wedge (\textsc{meet} \dot{\vee} \textsc{help}) \nvDash \Box_d\textsc{meet}$

The first observation follows from the general validity of modus ponens while the second is a direct consequence of the applicability conditions for defeasible generalizations together with the semantics for *ifs* and *oughts*. It follows from the defeasibility of epistemic *might* and its connection with deontic *ought* that the information carried by (7) defeats a commitment to (5), and this yields the final observation together with the semantics of the Ramsey conditional. It is easy to verify that Jones's obligation to help the drowning child, while induced by a norm that is in principle defeasible, is not affected by (7).[14]

I am about to conclude that the framework developed so far offers a straightforward treatment of prima facie obligations: such obligations may be overridden, and this is not surprising since they stem from deontic rules that are articulated using the defeasible conditional connective. But one may wonder whether the story told here really captures Ross's (1930) idea that prima facie obligations, even if overridden, continue to count in favor of some action rather than lose their status as a reason for action altogether. The answer, I think, depends on what it means to accept the antecedent of a defeasible norm (that is, to be in a state that triggers the default norm). We are free to suggest that doing so provides a reason for action, albeit one that may be outweighed by other

[14]One might complain that the proposal made here hardly tells us everything we want to know about Jones's case since it simply hardwires our intuitions about what he is normally required to do and about what counts as an exceptional case into an ordering source. However, it is arguably not the task of a framework for deontic reasoning to *explain* our normative intuitions but rather to correctly predict which inferences, *given* those intuitions, are valid and invalid, and here I add the additional twist that these predictions derive from plausible assumptions about the semantics of modals, conditionals, and defeasible generalizations. Integrating techniques from frameworks such as Horty's (2012) allows us to derive the ordering source for Jones's scenario from its first two rules and the assumption that the requirement to save a child in need takes priority over the requirement to keep a lunch appointment, but clearly one would then *still* have to rely normative intuitions that are not derived from more basic ethical principles.

reasons. Since the state of accepting the antecedent of a defeasible norm is guaranteed to be preserved even if new information is acquired, it follows that the overridability of prima facie obligations is compatible with their persistence as moral reasons for action. But this is not the only available interpretation: perhaps it is better to say that moral valence is entirely a matter of what norms are triggered *and* undefeated in the scenario under consideration, thus allowing for the possibility that a feature counting in favor some action given one scenario may not be a reason at all given another, or perhaps even count against choosing that action. If so, accepting the antecedent of a defeasible norm by itself does not give one a moral reason for acting in one way or another. Both views have some intuitive appeal, and both can be accommodated by the story told here.[15]

All of that preserves the advocated connection between prima facie obligations and defeasible generalizations, and it translates the classical idea that deontic reasoning is sensitive to the absence of information into the claim that deontic commitments are sensitive to epistemic commitments: since the latter may be defeated by additional information—in particular, commitments involving epistemic *might* are not necessarily preserved by information agglomeration—it is not surprising that deontic commitments are no less susceptible to preservation failure. The even better news is that the attested sensitivity of deontic *ought* to epistemic *might* highlights other potential nonmonotonic effects in deontic discourse and reasoning that would remain hidden if we merely focussed on the parallel between prima facie obligations and defeasible generalizations. Let me explain.

Taking a nonmonotonic perspective on discourse and reasoning, I have said, promises an attractive response to the problem surrounding the miners paradox. But whatever the underlying mechanism is, it cannot be that we are dealing with a prima facie obligation to block neither shaft that is overridden whenever the miners are in shaft A (shaft B). To see this, suppose we think of the deontic context in the miners scenario as follows:

$$o_d = \{[(\textsc{inA} \vee \textsc{inB}) > \neg(\textsc{blA} \vee \textsc{blB})], [\textsc{inA} > \textsc{blA}], [\textsc{inB} > \textsc{blB}]\}$$

Any frame that treats $[\![\neg(\textsc{blA} \vee \textsc{blB})]\!]$ as an $[\![\textsc{inA} \vee \textsc{inB}]\!]$-default while treating $[\![\textsc{blA}]\!]$ as an $[\![\textsc{inA}]\!]$-default and $[\![\textsc{blB}]\!]$ as an $[\![\textsc{inB}]\!]$-default is incoherent. For let $s = [\![\textsc{inA} \vee \textsc{inB}]\!]$ and consider any $w \in s$: then $w \in \mathcal{N}\pi(s)$ only if $w \in [\![\neg(\textsc{blA} \vee \textsc{blB})]\!]$. But clearly either $w \in [\![\textsc{inA}]\!]$ and in that case $w \notin \pi([\![\textsc{inA}]\!])$ or $w \in [\![\textsc{inB}]\!]$ and in that case $w \notin \pi([\![\textsc{inB}]\!])$, which is just to say that $\mathcal{N}\pi(s) = \varnothing$.

The last observation does not show that there is something wrong with our story about defeasible generalizations and prima facie obligations, but just that a nonmonotonic escape route from the miners paradox cannot rely on what this particular story has to say. The good news is that the dynamic framework developed so far does not tie nonmonotonic effects in deontic discourse and reasoning to the defeasibility of certain generalizations but rather to the defeasibility of epistemic *might*. The following deontic context does justice

[15] The first interpretation is in line with Horty's (2012) conception of reasons as triggered defaults, though he develops his framework further so that it leaves room for valence switching. The second interpretation is in line with a contextualist perspective on reasons, which has been developed in detail by, for instance, Lance and Little in a series of papers, starting with Lance and Little 2004 (which they originally labeled "particularist"). What all these views have in common is that they appeal to general, albeit defeasible, principles in the derivation of obligations, thus being at odds with moral particularism. See Dancy 2004 for a recent articulation of the case for particularism that also criticizes the appeal to defeasible principles as guides to reasoning about obligations. Alas, discussing the merits of these criticisms goes beyond the scope of this paper.

to our intuitions about the miners scenario while allowing for nonmonotonic effects to play a role in reasoning about that scenario:

$$o_d = \{[\text{BLA} \equiv \Box_e \text{INA}], [\text{BLB} \equiv \Box_e \text{INB}], [\neg(\text{BLA} \vee \text{BLB}) \equiv (\Diamond_e \text{INA} \wedge \Diamond_e \text{INB})]\}$$

Let σ be the information we have about the miners' whereabouts. Then $\sigma \models \Diamond_e \text{INA} \wedge \Diamond_e \text{INB}$ and accordingly $\sigma \models \neg\Box_e \text{INA}$ and $\sigma \models \neg\Box_e \text{INB}$. It follows that the minimal indices in σ are those at which we block neither shaft. But consider $\sigma' = \sigma[\text{INA}]$: then $\sigma' \models \Box_e \text{INA}$ and accordingly, the minimal indices in σ' are those at which we block shaft A. For parallel reasons, $\sigma[\text{INB}] \models \Box_d \text{BLB}$. Here is a summary of the output:

Fact 6 Consider d as fixed for the miners paradox and let σ be the information we have about the miners' whereabouts:

1. $\sigma \models \Box_d \neg(\text{BLA} \vee \text{BLB})$
2. $\sigma \models \text{INA} \Rightarrow \Box_d \text{BLA}$
3. $\sigma \models \text{INB} \Rightarrow \Box_d \text{BLB}$
4. $\sigma \models \text{INA} \vee \text{INB}$

The framework thus accounts for our intuitions about the miners scenario and resolves its air of paradox. For while $\sigma \models \Box_d \neg(\text{BLA} \vee \text{BLB})$, we also know that $\sigma[\text{INA}] \not\models \Box_d \neg(\text{BLA} \vee \text{BLB})$ and so even though $\sigma[\text{INA}] \models \Box_d \text{BLA}$, nonetheless $\sigma[\text{INA}] \not\models \bot$ and so $\sigma \not\models \neg\text{INA}$. For similar reasons, $\sigma \not\models \neg\text{INB}$. What makes all this possible is that adding the information that the miners are in shaft A (in shaft B) defeats one's rational commitment to the claim that we ought to block neither shaft, and the fact underlying *this* observation is the following:

Fact 7 Consider d as fixed for the miners paradox: then for all $\sigma \in \Sigma$ such that $\sigma \models \Diamond_e \text{BLA} \wedge \Diamond_e \text{BLB}$ it holds that if $\sigma \models \Box_d \neg(\text{BLA} \vee \text{BLB})$, then $\sigma \models \Diamond_e \text{INA} \wedge \Diamond_e \text{INB}$

A commitment to the claim that we ought to block neither shaft (even though we might block either of them) is dependent on one's ignorance about the miners' whereabouts. Once this ignorance is removed the deontic commitment has to go as well.[16]

We may now highlight what Jones's case has in common with the miners scenario and where they differ. Both scenarios require that an action that is deontically optimal with respect to some state σ fails to be deontically optimal with respected to a contracted state even though that action is still choosable. The possibility of meeting Mary for lunch, albeit deontically optimal in case Jones might do so *and* help the child in need,

[16]Carr (2012) and Cariani (forthcoming) argue that the semantics of deontic modality is sensitive to probabilistic considerations, and it would be possible to expand the story developed here so that ordering sources can be stated using probability operators. This would not change the fact that the framework developed here is "quietist" in Charlow's (forthcoming) sense since it does not explicitly encode decision-theoretic considerations in the semantics: rather, such considerations play an implicit role in selecting the relevant ordering source for the case under consideration. Charlow argues that so much quietism is not a good thing, but for our purposes there is no need to resolve this issue here: whether or not decision theoretic considerations are to be explicitly represented in the semantics, the point remains that our intuitions about what ought to be done in the miners scenario are sensitive to the presence or absence of certain epistemic possibilities in a way that the framework developed here is well-equipped to capture.

fails to be deontically optimal once we assume that he needs to *choose* between keeping his appointment and helping the child—even though he still might do the former instead of the latter. Likewise, the possibility of blocking neither shaft, albeit deontically optimal in case we do not know where the miners are, fails to be deontically optimal once we assume that the miners are in shaft A—even though we still might block neither shaft. So in both cases deontic *ought* exhibits a sensitivity to epistemic *might*, but for different reasons. In Jones's scenario, the underlying fact is that information strengthening may override certain obligations since certain rules may be defeated as additional information becomes available. This is not so in the miners scenario since, for instance, the rule to block none of the shafts just in case the miners might be in either of them is not defeasible in the first place. Instead, what is going on here is that contraction may result in the rule requiring an action—blocking at least one of the shafts—that it forbids in light of a weaker carrier of information (and similarly for the other rules in the ordering source). In other words, the rules that are at play in the miners scenario are *absolute* in the sense that they apply within any scenario, but they may require different actions since they are stated using epistemic modals.

Horty (2014) considers various obstacles toward integrating the full insights from default logic in a Kratzerian possible worlds analysis of deontic modals. I have not demonstrated that all these obstacles can be overcome—for instance, I have said nothing about higher-order default norms in deontic reasoning—but the progress made here should create some confidence that such a project is anything but futile. Let me briefly highlight one particular issue that Horty discusses in some detail. Suppose that there is a general prohibition against eating with your fingers ($\top > \neg F$) with one exception: if you are served cold asparagus, eating with your fingers is required ($A > F$).[17] Horty's preferred framework for conditional oughts predicts that in light of these norms, "You ought not eat with your fingers" and "If you are served cold asparagus, you ought to eat with your fingers" are accepted in an out-of-the-blue context. It also predicts that the norms under consideration do not license what he labels the "asparagus inference" to the conclusion that you ought not be served cold asparagus. And this is a good result since rules with exceptions should not render their exceptions as deontically sub-ideal by design.

Horty also correctly observes that Kratzer-style possible worlds analyses of deontic conditionals license the unfortunate asparagus inference. The inference pattern, however, is grounded in a "stability" feature of deontic contexts that is characteristic of classical frameworks but *not* of the one presented here: that if an index i is optimal in σ_d and included in some $\tau \subseteq \sigma$, then i is guaranteed to be optimal in τ_d as well.[18] It is well-worn story that so much stability is problematic when it comes to the miners paradox (see Cariani et al. 2013, Charlow 2013b, and Kolodny and MacFarlane 2010)—the crucial observations here are, first, that the feature is no less problematic when it comes to prima facie obligations, and, second, that it is avoided by the story about defaults told here. For consider $o_d = \{[\top > \neg F], [A > F]\}$ and let σ be the initial information state: then clearly $\sigma_d \Vdash \neg F$ yet $\sigma_d \nVdash \neg A$ even though $\sigma[A]_d \Vdash F$, and so $\sigma \vDash \Box_d \neg F$ and $\sigma \vDash A \Rightarrow \Box_d F$

[17] Following standard protocol, unconditional default rules are understood as defaults for the trivial scenario W; $\top$ is just any tautology.

[18] To see the connection, suppose that $\sigma \nvDash \Box_d \neg\phi$ and thus that for some optimal i in σ_d, $w_i \in [\![\phi]\!]$. Clearly $i \in \sigma[\phi]$ and thus, assuming that d is stable, optimal in $\sigma[\phi]_d$. And if $\sigma \vDash \Box_d \neg\psi$ as well, then $w_i \in [\![\neg\psi]\!]$ and hence $\sigma \nvDash \phi \Rightarrow \Box_d \psi$. Accordingly, if $\sigma \vDash \Box_d \neg\psi$ and $\sigma \vDash \phi \Rightarrow \Box_d \psi$, then $\sigma \vDash \Box_d \neg\phi$ under the assumption that d is stable

yet $\sigma \not\models \Box_d \neg A$, as desired.[19] It follows that a possible worlds analysis of conditionals and deontic modals—if developed with a dynamic spin—is not committed to licensing the unattractive asparagus inference.

A dynamic analysis of deontic *ought* offers an attractive foundation for a nonmonotonic perspective on discourse and reasoning. It readily handles cases in which prima facie obligations are overridden—cases that naturally call for a nonmonotonic analysis—but also offers a nonmonotonic perspective on thought and talk about *ought* under epistemic uncertainty, and all of this while keeping track of what these cases have in common and of where they differ. I take this result to be motivation enough to look at how the dynamic framework can be refined so that it captures the connection between nonmonotonicity, epistemic possibility, and the violability of obligations.

4 Loose Ends

The framework developed so far has *something* useful to say about contrary-to-duty obligations, though the story does not turn out to be entirely satisfying. To see the positive aspect of the story, fix the ordering source for Chisholm's scenario as follows:

$$o_d = \{[\textsc{go}], [\textsc{go} \supset \textsc{tell}], [\neg\textsc{go} \supset \neg\textsc{tell}]\}$$

Since we are only concerned here with absolute rules, considerations about frames are irrelevant for current purposes and so may simply consider the following distribution of truth-values across possible worlds:

	GO	TELL
w_1	T	T
w_2	T	F
w_3	F	T
w_4	F	F

If $s_\sigma = \{w_1, w_2, w_3, w_4\}$, then for any $i \in \sigma_d$, $w_i = w_1$ and accordingly $\sigma \models \Box_d\textsc{tell}$. But consider $\tau = \sigma[\neg\textsc{go}]$: then $s_\tau = \{w_3, w_4\}$ and so for any $i \in \tau_d$, $w_i = w_4$. Hence $\sigma[\neg\textsc{go}] \not\models \Box_d\textsc{tell}$ and in fact, $\sigma[\neg\textsc{go}] \models \Box_d\neg\textsc{tell}$, which just captures our intuition that Jones ought not tell his neighbors that he is coming under the assumption that he does not go. The general observations are the following:

Fact 8 Consider d as fixed for Chisholm's scenario:

1. $\Box_d\textsc{go}, \textsc{go} \Rightarrow \Box_d\textsc{tell} \models \Box_d\textsc{tell}$
2. $\Box_d\textsc{go}, \textsc{go} \Rightarrow \Box_d\textsc{tell}, \neg\textsc{go} \not\models \Box_d\textsc{tell}$

[19] The underlying fact here is that only the W-default $[\![\neg F]\!]$ but not the $[\![A]\!]$-default $[\![F]\!]$ is triggered in σ_d. Accordingly, optimality in σ_d merely requires compliance with the W-default $[\![\neg F]\!]$ and so an index may be optimal even if its world-parameter is an A-world. When it comes to the miners paradox, suppose again that σ is the information we have about the miners paradox: then all the indices in σ_d are optimal since no default norms are involved and for some $i \in \sigma_d$, $w_i \in [\![(\neg\textsc{bl}\mathrm{A} \wedge \neg\textsc{bl}\mathrm{B}) \wedge \textsc{in}A]\!]$. Such an index is an element of $\sigma[\textsc{in}\mathrm{A}]$ but not included in $\sigma[\textsc{in}\mathrm{A}]_d$, which shows that the deontic context for the miners scenario is just as unstable as the one for the asparagus case.

Once again we can predict that thought and talk about deontic *ought* is nonmonotonic: in the present case, an additional bit of information that licenses the derivation of a contrary-to-duty obligation—Jones's obligation not to tell his neighbors that he is coming—defeats one's commitment to a conflicting obligation—Jones's obligation to tell his neighbors that he is coming. The validity of deontic detachment and thus of the move from (9) and (10) to (11) is compatible with the intuition that (9) and (10) no longer license the inference of (11) under the additional assumption that Jones does not go.

All of this is good news, but there remains the question *why* reasoning with violable obligations is susceptible to monotonicity failures, and it is not clear that the story told here gives quite the right answer. For notice that the current framework predicts that $\Box_d\phi \vDash \Diamond_e\phi$ for all $\phi \in \mathcal{L}$, and so the reason for why $\neg$GO defeats $\Box_d$TELL is that $\neg$GO defeats $\Box_d$GO. But of course we want to say that Jones ought to go regardless of whether or not he goes, and so the treatment of the case under consideration is not fully satisfying. Let me outline how the basic framework may be expanded so that we can do better (see Willer 2014 for additional discussion).

I rely on Frank's (1997) proposal that deontic *ought* requires that the input context be "non-trivial" in the sense that its prejacent as well as its negation must be open possibilities. This requirement was satisfied in all the cases we considered earlier but it naturally becomes an issue when we look at cases in which obligations are violated. Frank suggests that nontriviality violations result in the retraction of information to arrive at an appropriate state. How to retract information from a state is a very complex issue, and here I choose a very simple approach and assume that context associates with each $\sigma \in \Sigma$ a system of spheres $S(\sigma)$ that is ordered by $\subseteq$ and centered on σ. The intuitive role of $S(\sigma)$ is to capture which commitments stand and fall together, and it is then possible to define a downdating operation on carriers of information as follows:

Definition 15 (Downdating) Consider arbitrary $\sigma \in \Sigma$ and $\phi \in \mathcal{L}$. $S(\sigma) \circ \phi = \{\sigma' \in S(\sigma) \colon \sigma' \nvDash \phi\}$. The result of *downdating* σ with ϕ, $\sigma \downarrow \phi$, is the minimal element of $S(\sigma) \circ \phi$ in case $S(\sigma) \circ \phi \neq \varnothing$, and σ otherwise.

Downdating σ with ϕ removes any commitment to ϕ by weakening σ to its minimal revision that is no longer committed to ϕ. Notice that downdating idles whenever the input state already fails to be committed to ϕ.

The modified proposal then is that deontic *ought* is a universal quantifier over the set of possible worlds that are deontically optimal in light of a carrier of information that leaves room for the prejacent as well as its negation to be a possibility. Precisely:

Definition 16 (Deontic Ought with Retraction) Extend the update rules for $\mathcal{L}$ with the following entry:

(6) $\sigma[\Box_d\phi] = \{i \in \sigma \colon (\sigma \downarrow \phi \downarrow \neg\phi)_d \Vdash \phi\}$

Downdating thus guarantees that we consider an appropriate carrier of information even if nontriviality is violated by the input state (as long as the prejacent is contingent).

It is easy to see that the refined proposal avoids the unfortunate result that any update with $\ulcorner\neg\phi\urcorner$ defeats a commitment to $\ulcorner\Box_d\phi\urcorner$. For consider again the distribution of truth-values across possible worlds from the beginning of this section, let $s_\sigma = \{w_1, w_2, w_3, w_4\}$

and consider $\sigma[\neg\text{GO}]$: then downdating with the prejacent of $\Box_d\text{TELL}$ or $\Box_d\neg\text{TELL}$ idles and so $\sigma[\neg\text{GO}] \vDash \Box_d\neg\text{TELL}$. But downdating with the prejacent of $\Box_d\text{GO}$ or $\Box_d\neg\text{GO}$ re-introduces some indices at which Jones does go to the help of his neighbors since $\sigma[\neg\text{GO}] \vDash \neg\text{GO}$. In particular, observe that on the most natural conception of the fallback relation, $\sigma[\neg\text{GO}] \downarrow \text{GO} = \sigma$, the intuitive idea being that if Jones were to go, he might and might not tell his neighbors he is coming. But if that is right, then for any $i \in (\sigma \downarrow \text{GO} \downarrow \neg\text{GO})_d$, $w_i = w_1$. Given minimal assumptions about the fallback relation that figures in downdating, we can thus predict that, assuming that Jones does not go, he ought not tell his neighbors that he is coming, but he (still) ought to go.

The revised story is not entirely complete since it is currently incompatible with the validity of deontic detachment: as before $\sigma[\neg\text{GO}] \vDash \Box_d\neg\text{TELL}$, but now we also have $\sigma[\neg\text{GO}] \vDash \Box_d\text{GO}$ and $\sigma[\neg\text{GO}] \vDash \text{GO} \Rightarrow \Box_d\text{TELL}$. So if this were the final word, there would be no nonmonotonic story to be told here in the first place; however, it is a familiar assumption that conditionals in the indicative mood (such as the one that figures in Chisholm's case) presuppose that their antecedent is compatible with the input context:

Definition 17 (Presupposition) For all $\sigma \in \Sigma$: $\sigma[\phi \Rightarrow \psi]$ is defined iff $\sigma[\phi] \neq \varnothing$.

Here presupposition is modeled as a definedness condition on updating.[20] In these lights, the notion of logical consequence should be refined as follows (see Starr 2014 and also Beaver 2001 and von Fintel 1999):

Definition 18 (Logical Consequence with Presupposition) $\phi_1, \ldots \phi_n \vDash \psi$ iff for all $\sigma \in \Sigma$: if $\sigma[\phi_1]\ldots[\phi_n][\psi]$ is defined, then $\sigma[\phi_1]\ldots[\phi_n] \vDash \psi$.

Logical consequence remains understood as preservation of rational commitment, but we now set aside those input states for which updating with the premises and then with the conclusion is undefined. This is all we need for the final proposal.

Observe that while the commitment to the claim that Jones ought to go is not sensitive to the epistemic possibility that Jones in fact goes, the conditional obligation is:

Fact 9 Consider d as fixed for Chisholm's scenario: then for all $\sigma \in \Sigma$ it holds that if $\sigma \vDash \text{GO} \Rightarrow \Box_d\text{TELL}$, then $\sigma \vDash \Diamond_e\text{GO}$.

The hypothesis then is that the phenomenon of nonmonotonicity in Chisholm's scenario can once again be explained in terms of the sensitivity of deontic commitments to the existence of epistemic possibilities: the conditional obligation licensing the inference to the conclusion that Jones ought to tell his neighbors that he is coming *presupposes* that Jones might in fact go to the help of his neighbors.

Earlier I said that just as reasoning with prima facie obligations is sensitive to the absence of information, so is reasoning with violable obligations. The informal proposal, remember, was that such obligations may entail other obligations unless they are violated, and I can now make this more precise. Say that an *inference* presupposes ϕ just in case any state for which updating with its premises and its conclusion is defined is committed

[20] See Heim 1982 and, among others, von Fintel 1998, Gillies 2009, 2010, and Starr 2014 for the presuppositional analysis of conditionals. The basic idea goes back to Stalnaker (1975).

to ϕ: then an inference licensed by deontic detachment—the rule allowing the step from (9) and (10) to (11)—presupposes that the satisfaction of the unconditional obligation is compatible with the input state. It follows that even if deontic detachment is valid, the inferences it licenses may be defeated through information agglomeration.

On the view I have proposed here, then, an obligation $[\![\phi]\!]$ is *violated but binding* in σ just in case $\sigma \models \neg\phi \wedge \Box_d\phi$. Such cases differ sharply from those in which an obligation $[\![\phi]\!]$ is *defeated* in σ, which are such that $\sigma \models \neg\Box_d\phi$ yet for some σ' so that $\sigma \subseteq \sigma'$, $\sigma' \models \Box_d\phi$. And intuitively, whenever $[\![\phi]\!]$ is a prima facie obligation conditional on $[\![\psi]\!]$ but defeated in σ, this is because σ depicts an *exceptional circumstance* just in case $\sigma \models \psi \wedge \neg\Box_d\phi$.

So in our Chisholm-style scenario, Jones violates his obligation to go but his obligation to tell his neighbors is defeated and thus no longer binding. And once again we have identified a nonmonotonic effect in deontic discourse and reasoning that cannot be explained by appealing to the fact that certain obligations allow for exceptions. To see this, suppose an alternative deontic context d' providing the following ordering source consisting of defeasible rules:

$$o_{d'} = \{[\top > \textsc{go}], [\textsc{go} > \textsc{tell}], [\neg\textsc{go} > \neg\textsc{tell}]\}$$

Then clearly any $\sigma_{d'}$ treats $[\![\textsc{go}]\!]$ as a W-default, $[\![\textsc{tell}]\!]$ as a $[\![\textsc{go}]\!]$-default, and $[\![\neg\textsc{tell}]\!]$ as a $[\![\neg\textsc{go}]\!]$-default. The simple observation then is that $\sigma[\neg\textsc{go}] \models \Box_d\textsc{go}$, which is just to say that $\sigma[\neg\textsc{go}]$ does not depict an exceptional circumstance in which the default obligation to go fails to be binding—this just captures the intuition that we should not collapse violations and exceptions. And while the prima facie obligation $[\![\textsc{tell}]\!]$ conditional on $[\![\textsc{go}]\!]$ is defeated in $\sigma[\neg\textsc{go}]$, this is not so because $\sigma[\neg\textsc{go}]$ depicts an exceptional circumstance either since $\sigma[\neg\textsc{go}] \not\models \textsc{go}$—this just captures the intuition that a prima facie obligation conditional on $[\![\textsc{go}]\!]$ fails to pertain to situations in which Jones does not go, and so a scenario in which Jones does not go hardly qualifies as an exceptional circumstance. That certain obligations allow for exceptions does not explain why reasoning with violable obligations has a nonmonotonic flavor.

The fact that the framework developed here has no trouble accounting for the connection between defeasibility and violability demonstrates once again that it captures a rich variety of nonmonotonic effects in discourse and reasoning without reducing them to exceptional circumstances. While providing a promising account of violable obligations requires adding some non-trivial complications to the basic dynamic account told in §3, its success at fulfilling a range of key desiderata for a nonmonotonic story about reasoning with violable obligations suggests that the story is on the right track.

5 Conclusion

Deontic discourse and reasoning, so a familiar story goes, is nonmonotonic, but the variety of ways in which it is sensitive to the absence of information has traditionally been underappreciated. The traditional focus on prima facie obligations overlooks that information agglomeration may trigger nonmonotonic effects without creating contexts in which a prima facie obligation no longer applies, and I have highlighted two reasons for why this is so: the fact that obligations may be violated and the one that obligations may be sensitive to epistemic uncertainty. The analysis developed here differs from what has happened before in the literature in that it offers a comprehensive perspective on the nonmonotonic nature of deontic discourse and reasoning, and it does so by translating the sensitivity of deontic inferences to the absence of information into the sensitivity of deontic *ought* to epistemic *might*. Since information agglomeration may defeat commitments to epistemic *might*, it may also defeat commitments to deontic *ought*: the nonmonotonicity of deontic thought and talk can be illuminated in terms of the familiar nonmonotonicity of epistemic thought and talk that finds a natural articulation in dynamic semantics.

I have remained silent on another popular motivation for nonmonotonic approaches to deontic logic: the possibility of genuine moral dilemmas. This is deliberate, for even if deontic logic must leave room for such dilemmas there is some reason to think that their existence is compatible with a monotonic outlook (see, e.g., van Fraassen (1973), Cariani (2013), and von Fintel (2012)). But let me briefly come back to the question of whether attempting to arrive at a suitably nonmonotonic logical consequence relation puts an explanatory burden on semantic theorizing that is better reserved for some other component of a complete story about meaning, communication, and reasoning.

The question connects with the issue whether dynamic semantics has any advantages over a truth-conditional alternative coupled with an adequate *pragmatic* story about conversational dynamics (see Rothschild and Yalcin (2012) for discussion). To get the issue into better view, suppose we recursively define truth-conditions for our target language relative to an index i and information state σ that delivers the following result for all ϕ of our target language: if $i \in \sigma$, then $[\![\phi]\!]^{i,\sigma}$ is true just in case $i \in \sigma[\phi]$. We may then define two distinct notions of logical consequence:

1. Neoclassical Consequence: $\phi_1, \ldots, \phi_n \vDash \psi$ iff for all i and σ such that $i \in \sigma$: if $[\![\phi_1]\!]^{i,\sigma}$ is true and ... and $[\![\phi_n]\!]^{i,\sigma}$ is true, then $[\![\psi]\!]^{i,\sigma}$ is true.
2. Informational Consequence: $\phi_1, \ldots, \phi_n \vDash \psi$ iff $\Box_e\psi$ is a neoclassical consequence of $\Box_e\phi_1, \ldots, \Box_e\phi_n$.

These semantic entailment relations have been extensively discussed in the context of the miners paradox: neoclassical consequence avoids the problem by denying modus ponens (see Kolodny and MacFarlane 2010) while informational consequence does so by allowing for reductio failures (see Bledin 2014 and Yalcin 2012). But what about Ross- and Chisholm-style scenarios, which stand at odds with the fact that neoclassical and informational consequence are monotonic by design? The most promising strategy is to appeal to the already familiar story about conversational dynamics from Stalnaker (1978): additional information in discourse and reasoning affects, as a matter of pragmatics, the informational parameter in light of which subsequent utterances are evaluated. More precisely, say that $\sigma + \phi = \sigma \cap [\![\phi]\!]^{\sigma}$. Then we may define a pragmatic inference relation (inspired by Stalnaker 1975) that treats the inference of ψ from $\phi_1, \ldots \phi_n$ as *reasonable* (rather than valid) just in case for all i and σ: if $i \in \sigma + \phi_1 + \ldots \phi_n$, then $i \in [\![\psi]\!]^{\sigma+\phi_1+\ldots\phi_n}$. Given the striking similarity between reasonable inference and dynamic logical consequence, it should not be surprising that the former is just as nonmonotonic as the latter is: epistemic and deontic commitments are, as we have seen, defeasible and hence a sentence may be true in light of some carrier of information but fail to be true with respect to some stronger state. We may thus leave it to pragmatics to explain why additional information in discourse and reasoning can defeat existing deontic commitments, and of course this shift of explanatory burden would live happily with the suggestion that belief revision theorists, rather than those concerned with semantic entailment, should worry about nonmonotonicity. Why let semantics do all the work?

Let me make two points in response. First, many key aspects of the previous discussion remain unaffected by the decision of how to divide the labor between semantics and pragmatics (or between logic and belief revision theory). Specifically, the point remains that the nonmonotonicity of deontic discourse and reasoning stands in need of an explanation that goes beyond the one provided by theories that take their inspiration from default logic, and that such an explanation *can* be provided by a possible worlds semantics for conditionals and deontic modals that takes dynamic effects in discourse and reasoning seriously. Furthermore, I contend that any such explanation needs a model of how norms are sensitive to the flow of information along the lines I have provided in this paper. These aspects of my story, as well as the key ideas behind their formal elaboration, remain in place regardless of the semantics-pragmatics distinction.

Playing a bit more offense, the choice of a monotonic consequence relation has repercussions that go beyond the need to put substantial explanatory burden onto a pragmatic story about information dynamics. Most strikingly, *any* monotonic conception of logical consequence will inevitably be at odds with the semantic fact that conditionals are nonmonotonic in the antecedent (resist antecedent strengthening), and in particular we have cases in which $\Gamma \vDash \psi$ yet $\Gamma \nvDash \phi \Rightarrow \psi$ for some $\phi, \psi \in \mathcal{L}^+$ and $\Gamma \subseteq \mathcal{L}^+$.[21] But of course on a monotonic conception of logical consequence, $\Gamma, \phi \vDash \psi$ whenever $\Gamma \vDash \psi$, which is just to say that any such conception must flat out reject the semantic equivalent of the deduction theorem: if $\Gamma, \phi \vDash \psi$ then $\Gamma \vDash \phi \Rightarrow \psi$. As Yalcin (2012, fn. 14) observes, this is a problematic feature of neoclassical consequence—it disconnects conditionals from consequence in unexpected ways—but it is in fact shared by all monotonic entailment relations, including informational consequence. And of course this is just what we expect if we ban dynamic effects from the semantic entailment relation but also follow the standard protocol of adopting a Ramsey-inspired semantics for conditionals that evaluates such constructions by evaluating their consequents in light of the result of *updating* some carrier of information with the antecedent. Only a dynamic entailment relation along the lines I have suggested here preserves the intuitive match between the semantic evaluation procedure for conditionals and the one for logical arguments. And since this conception of logical consequence essentially relies on the assumption that semantic values are relations between information states rather than indices of evaluation, we also have a point here at which a dynamic perspective on semantic theorizing differs non-trivially from a truth-conditional alternative that pushes dynamic effects into the pragmatics.[22]

There is, to be clear, *some* legitimacy to a monotonic perspective on discourse and reasoning. Specifically, there is every reason to think that logical consequence should be insensitive to the absence of information if validity amounts to guaranteed preservation of truth at a point of evaluation, and in fact dynamic logical consequence turns out to be monotonic in case we restrict attention to the propositional fragments of our target languages. The point here is that validity as guaranteed preservation of truth is just a special instance of the more general conception of logical consequence as guaranteed preservation of rational commitment, and this connects with the semantic fact that truth-conditional

[21] To choose an equivalent formulation that carries strengthening failures on its sleeves: $\Gamma \vDash \top \Rightarrow \psi$ yet $\Gamma \nvDash (\top \wedge \phi) \Rightarrow \psi$. One example would be the premises of the miners paradox.

[22] For sure, one may rely on CCPs in semantic theorizing and just arrive at a framework that makes the same predictions about consistency and entailment as its truth-conditional alternative (see von Fintel and Gillies 2008). Such proposals arguably fail to be *interestingly* dynamic, but clearly the one I have developed here is not one of them.

meaning is just a special instance of meaning understood as context change potential.[23] Just like the classical conception of meaning as truth-conditional meaning, monotonicity has some place in our best theory about discourse and reasoning—the dynamic perspective on meaning and communication, so the moral of the story, makes sense of the scope as well as the limit of the classical view.

The comprehensiveness of the framework developed here does not reduce to the fact that it has something nonmonotonic to say about thought and talk pertaining to obligations that may be overridden, violated, or are sensitive to epistemic uncertainty. It showcases what these cases have in common and where they differ, and in addition assigns the proper place to important ideas from the classical literature. The fact that we have every reason to take seriously a nonmonotonic perspective on deontic discourse and reasoning is well-established. I submit that we have just as much reason to take seriously a dynamic perspective on deontic discourse and reasoning.

Acknowledgments Key ideas of this paper were presented at a workshop on nonstandard modals at the University of Leeds (organized by Paolo Santorio). Special thanks to the audience for insightful comments and discussion, as well as to Fabrizio Cariani, Nate Charlow, Matthew Chrisman, Frank Veltman, and an anonymous peer commentator for detailed feedback that proved to be invaluable in preparing the final manuscript. Support from the Franke Institute for the Humanities is gratefully acknowledged.

References

Åqvist, Lennart. 2002. "Deontic Logic." In *Handbook of Philosophical Logic, 2nd Edition, Volume 8*, ed. Dov M. Gabbay and Franz Guenthner, 147–264. Dordrecht: Kluwer. http://dx.doi.org/10.1007/978-94-010-0387-2_3.

Asher, Nicholas, and Daniel Bonevac. 1996. "Prima Facie Obligation." *Studia Logica: An International Journal for Symbolic Logic* 57(1): 19–45. http://dx.doi.org/10.1007/BF00370668.

———. 1997. "Common Sense Obligation." In *Defeasible Deontic Logic*, ed. Donald E. Nute, 159–203. Dordrecht: Kluwer. http://dx.doi.org/10.1007/978-94-015-8851-5_8.

Beaver, David I. 2001. *Presupposition and Assertion in Dynamic Semantics*. Stanford: CSLI Publications.

Belzer, Marvin. 1986. "Reasoning with Defeasible Principles." *Synthese* 66(1): 135–158. http://dx.doi.org/10.1007/BF00413584.

van Benthem, Johan. 1986. *Essays in Logical Semantics*. Dordrecht: Reidel. http://dx.doi.org/10.1007/978-94-009-4540-1.

[23] The underlying observation from van Benthem (1986) is that updating with an element of $\mathcal{L}_0$ is always mediated by a classical proposition. This is no longer the case once we consider modal formulas and state their semantics in terms of acceptance conditions that are sensitive to global features of the input state.

Bledin, Justin. 2014. "Logic Informed." *Mind* 490(123). http://dx.doi.org/10.1093/mind/fzu073.

Bonevac, Daniel. 1998. "Against Conditional Obligation." *Noûs* 32(1): 37–53. http://dx.doi.org/10.1111/0029-4624.00086.

Cariani, Fabrizio. 2013. "'Ought' and Resolution Semantics." *Noûs* 47(3): 534–558. http://dx.doi.org/10.1111/j.1468-0068.2011.00839.x.

———. Forthcoming. "Deontic Modals and Probabilities: One Theory to Rule Them All?" In *Deontic Modality*, ed. Nate Charlow and Matthew Chrisman. Oxford University Press.

Cariani, Fabrizio, Magdalena Kaufmann, and Stefan Kaufmann. 2013. "Deliberative Modality under Epistemic Uncertainty." *Linguistics and Philosophy* 36(3): 225–259. http://dx.doi.org/10.1007/s10988-013-9134-4.

Carmo, José, and Andrew J. I. Jones. 2002. "Deontic Logic and Contrary-to-Duties." In *Handbook of Philosophical Logic, 2^{nd} Edition, Volume 8*, ed. Dov M. Gabbay and Franz Guenthner, 265–343. Dordrecht: Kluwer. http://dx.doi.org/10.1007/978-94-010-0387-2_4.

Carr, Jennifer. 2012. "Deontic Modals without Decision Theory." In *Proceedings of Sinn und Bedeutung 17*, 167–182.

Charlow, Nate. 2013a. "Conditional Preferences and Practical Conditionals." *Linguistics and Philosophy* 36(6): 463–511. http://dx.doi.org/10.1007/s10988-013-9143-3.

———. 2013b. "What We Know and What To Do." *Synthese* 190(12): 2291–2323. http://dx.doi.org/10.1007/s11229-011-9974-9.

———. Forthcoming. "Decision Theory: Yes! Truth Conditions: No!" In *Deontic Modality*, ed. Nate Charlow and Matthew Chrisman. Oxford University Press.

Chisholm, Roderick M. 1963. "Contrary-to-Duty Imperatives and Deontic Logic." *Analysis* 24(2): 33–36. http://dx.doi.org/10.2307/3327064.

Dancy, Jonathan. 2004. *Ethics Without Principles*. New York: Oxford University Press. http://dx.doi.org/10.1093/0199270023.001.0001.

Dowell, Janice L. 2012. "Contextualist Solutions to Three Puzzles about Practical Conditionals." In *Oxford Studies in Metaethics*, ed. Russ Shafer-Landau, vol. 7, 271–303. New York: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199653492.003.0009.

von Fintel, Kai. 1998. "The Presupposition of Subjunctive Conditionals." In *The Interpretive Tract (MIT Working Papers in Linguistics 25)*, ed. Uli Sauerland and Orin Percus, 29–44. Cambridge, MA: MITWPL.

———. 1999. "NPI Licensing, Strawson Entailment, and Context Dependency." *Journal of Semantics* 16(2): 97–148. http://dx.doi.org/10.1093/jos/16.2.97.

———. 2012. "The Best We Can (Expect to) Get? Challenges to the Classic Semantics for Deontic Modals." http://mit.edu/fintel/fintel-2012-apa-ought.pdf. Manuscript, MIT.

von Fintel, Kai, and Anthony S. Gillies. 2008. "An Opinionated Guide to Epistemic Modality." In *Oxford Studies in Epistemology 2*, ed. Tamar Gendler and John Hawthorne, 32–62. New York: Oxford University Press.

Frank, Anette. 1997. *Context Dependence in Modal Constructions*. Ph.D. thesis, University of Stuttgart.

Fuhrmann, André. 1989. "Reflective Modalities and Theory Change." *Synthese* 81(1): 115–134. http://dx.doi.org/10.1007/BF00869347.

Gillies, Anthony S. 2004. "Epistemic Conditionals and Conditional Epistemics." *Noûs* 38(4): 585–616. http://dx.doi.org/10.1111/j.0029-4624.2004.00485.x.

———. 2009. "On Truth-Conditions for *If* (but Not Quite Only *If*)." *Philosophical Review* 118(3): 325–349. http://dx.doi.org/10.1215/00318108-2009-002.

———. 2010. "Iffiness." *Semantics and Pragmatics* 3(4): 1–42. http://dx.doi.org/10.3765/sp.3.4.

Greenspan, Patricia S. 1975. "Conditional Oughts and Hypothetical Imperatives." *The Journal of Philosophy* 72(10): 259–276. http://dx.doi.org/10.2307/2024734.

Groenendijk, Jeroen, and Martin Stokhof. 1991. "Dynamic Predicate Logic." *Linguistics and Philosophy* 14(1): 39–100. http://dx.doi.org/10.1007/BF00628304.

Harman, Gilbert. 1986. *Change in View*. Cambridge, MA: MIT Press.

Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, University of Massachusetts, Amherst.

Horty, John F. 1994. "Moral Dilemmas and Nonmonotonic Logic." *Journal of Philosophical Logic* 23(1): 35–65. http://dx.doi.org/10.1007/BF01417957.

———. 1997. "Nonmonotonic Foundations for Deontic Logic." In *Defeasible Deontic Logic*, ed. Donald E. Nute, 17–44. Dordrecht: Kluwer. http://dx.doi.org/10.1007/978-94-015-8851-5_2.

———. 2003. "Reasoning with Moral Conflicts." *Noûs* 37(4): 557–605. http://dx.doi.org/10.1046/j.1468-0068.2003.00452.x.

———. 2007. "Reasons as Defaults." *Philosophers' Imprint* 7(3): 1–28. http://dx.doi.org/2027/spo.3521354.0007.003.

———. 2012. *Reasons as Defaults*. New York: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199744077.001.0001.

———. 2014. "Deontic Modals: Why Abandon the Classical Semantics?" *Pacific Philosophical Quarterly* 4(95): 424–460. http://dx.doi.org/10.1111/papq.12044.

Kamp, Hans, Josef van Genabith, and Uwe Reyle. 2011. "Discourse Representation Theory." In *Handbook of Philosophical Logic*, 2nd ed., vol. 15, ed. Dov M. Gabbay and Franz Guenthner, 125–394. Dordrecht: Springer. http://dx.doi.org/10.1007/978-94-007-0485-5_3.

Kolodny, Niko, and John MacFarlane. 2010. "Ifs and Oughts." *Journal of Philosophy* 107(3): 115–143.

Kratzer, Angelika. 1991. "Modality." In *Semantics: An International Handbook of Contemporary Research*, ed. Armin von Stechow and Dieter Wunderlich, 639–650. Berlin: Walter de Gruyter & Co.

———. 2012. *Modals and Conditionals*. New York: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199234684.001.0001.

Lance, Mark N., and Margaret O. Little. 2004. "Defeasibility and the Normative Grasp of Context." *Erkenntnis* 61(2-3): 435–455. http://dx.doi.org/10.1007/s10670-004-9286-2.

Levi, Isaac. 1988. "Iteration of Conditionals and the Ramsey Test." *Synthese* 76(1): 49–81. http://dx.doi.org/10.1007/BF00869641.

Lewis, David K. 1973. *Counterfactuals*. Cambridge, MA: Harvard University Press.

———. 1974. "Semantic Analyses for Dyadic Deontic Logic." In *Logical Theory and Semantic Analysis: Essays Dedicated to Stig Kanger on His Fiftieth Birthday*, ed. Sören Stenlund, 1–14. Dordrecht: Reidel. http://dx.doi.org/10.1007/978-94-010-2191-3_1.

McCarthy, John. 1980. "Circumscription—A Form of Non-Monotonic Reasoning." *Artificial Intelligence* 13(1-2): 27–39. http://dx.doi.org/10.1016/0004-3702(80)90011-9.

McCarty, L. Thorne. 1994. "Defeasible Deontic Reasoning." *Fundamenta Informaticae* 21(1-2): 125–148. http://dx.doi.org/10.3233/FI-1994-21128.

McDermott, Drew, and Jon Doyle. 1980. "Non-Monotonic Logic I." *Artificial Intelligence* 13(1-2): 41–72. http://dx.doi.org/10.1016/0004-3702(80)90012-0.

McNamara, Paul. 2006. "Deontic Logic." In *The Handbook of the History of Logic, vol. 7: Logic and Modalities in the Twentieth Century*, ed. Dov M. Gabbay and John H. Woods, 197–288. Amsterdam: Elsevier. http://dx.doi.org/10.1016/S1874-5857(06)80029-4.

Nute, Donald E. 1988. "Defeasible Reasoning and Decision Support Systems." *Decision Support Systems* 4(1): 97–110. http://dx.doi.org/10.1016/0167-9236(88)90100-5.

———. 1997. "Apparent Obligation." In *Defeasible Deontic Logic*, ed. Donald E. Nute, 287–315. Dordrecht: Kluwer. http://dx.doi.org/10.1007/978-94-015-8851-5_12.

Pollock, John L. 1987. "Defeasible Reasoning." *Cognitive Science* 11(4): 481–518. http://dx.doi.org/10.1016/S0364-0213(87)80017-4.

Prakken, Henry, and Marek Sergot. 1996. “Contrary-to-duty Obligations.” *Studia Logica* 57(1): 73–90. http://dx.doi.org/10.1007/BF00370671.

———. 1997. “Dyadic Deontic Logic and Contrary-to-Duty Obligations.” In *Defeasible Deontic Logic*, ed. Donald E. Nute, 223–262. Dordrecht: Kluwer. http://dx.doi.org/10.1007/978-94-015-8851-5_10.

Ramsey, Frank P. 1931. *The Foundations of Mathematics*. London: Routledge and Kegan Paul.

Reiter, Raymond. 1980. “A Logic for Default Reasoning.” *Artificial Intelligence* 13(1-2): 81–132. http://dx.doi.org/10.1016/0004-3702(80)90014-4.

Ross, William D. 1930. *The Right and the Good*. Oxford: Oxford University Press. http://dx.doi.org/10.1093/0199252653.001.0001.

Rothschild, Daniel, and Seth Yalcin. 2012. “On the Dynamics of Conversation.” http://danielrothschild.com/dynamics.pdf. Manuscript, Oxford/Berkeley.

Rott, Hans. 1989. “Conditionals and Theory Change: Revisions, Expansions, and Additions.” *Synthese* 81(1): 91–113. http://dx.doi.org/10.1007/BF00869346.

Ryu, Young U., and Ronald M. Lee. 1997. “Deontic Logic Viewed as Defeasible Reasoning.” In *Defeasible Deontic Logic*, ed. Donald E. Nute, 123–137. Dordrecht: Kluwer. http://dx.doi.org/10.1007/978-94-015-8851-5_6.

Silk, Alex. 2013. “Truth Conditions and the Meanings of Ethical Terms.” In *Oxford Studies in Metaethics, Volume 8*, ed. Russ Shafer-Landau, 195–222. New York: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199678044.003.0008.

Stalnaker, Robert C. 1975. “Indicative Conditionals.” *Philosophia* 5(3): 269–286. http://dx.doi.org/10.1007/BF02379021.

———. 1978. “Assertion.” In *Syntax and Semantics 9*, ed. Peter Cole, 315–332. New York: New York Academic Press.

Starr, William B. 2014. “A Uniform Theory of Conditionals.” *Journal of Philosophical Logic* 43(6): 1019–1064. http://dx.doi.org/10.1007/s10992-013-9300-8.

Thomason, Richmond H. 1981. “Deontic Logic as Founded on Tense Logic.” In *New Studies in Deontic Logic*, ed. Risto Hilpinen, 165–176. Dordrecht: Reidel. http://dx.doi.org/10.1007/978-94-009-8484-4_7.

van der Torre, Leendert W. N., and Yao-Hua Tan. 1997. “The Many Faces of Defeasibility.” In *Defeasible Deontic Logic*, ed. Donald E. Nute, 79–121. Dordrecht: Kluwer. http://dx.doi.org/10.1007/978-94-015-8851-5_5.

———. 1998. “An Update Semantics for Deontic Reasoning.” In *Proceedings of DEON 1998*, 409–426.

Touretzky, David S. 1986. *The Mathematics of Inheritance Systems*. Los Altos, CA: Morgan Kaufmann.

van Fraassen, Bas. 1973. "Values and the Heart's Command." *Journal of Philosophy* 70(1): 5–19. http://dx.doi.org/10.2307/2024762.

Veltman, Frank. 1985. *Logics for Conditionals*. Ph.D. thesis, University of Amsterdam.

———. 1996. "Defaults in Update Semantics." *Journal of Philosophical Logic* 25(3): 221–261. http://dx.doi.org/10.1007/BF00248150.

Willer, Malte. 2012. "A Remark on Iffy Oughts." *Journal of Philosophy* 109(7): 449–461.

———. 2014. "Dynamic Thoughts on Ifs and Oughts." *Philosophers' Imprint* 14(28): 1–30. http://dx.doi.org/2027/spo.3521354.0014.028.

von Wright, George H. 1951. "Deontic Logic." *Mind* 60(237): 1–15. http://dx.doi.org/10.1093/mind/LX.237.1.

———. 1956. "A Note on Deontic Logic and Derived Obligation." *Mind* 65(260): 507–509. http://dx.doi.org/10.1093/mind/65.1.507.

Yalcin, Seth. 2012. "A Counterexample to Modus Tollens." *Journal of Philosophical Logic* 41(6): 1001–1024. http://dx.doi.org/10.1007/s10992-012-9228-4.