Prima Facie Obligation

Abstract. This paper presents a nonmonotonic deontic logic based on commonsense entailment. It establishes criteria a successful account of obligation should satisfy, and develops a theory that satisfies them. The theory includes two conditional notions of prima facie obligation. One is constitutive; the other is epistemic, and follows nonmonotonically from the constitutive notion. The paper defines unconditional notions of prima facie obligation in terms of the conditional notions.

Key words: deontic logic, nonmonotonic logic.

To attack the sophists’ conception of weakness of will, Aristotle cites Sophocles’ Philoctetes (Nicomachean Ethics VII, 2, 1146a19). As the play opens, Philoctetes has suffered for years with a disfiguring disease; he had wandered into a forbidden garden, through no fault of his own, and had been punished by the gods. Banished to a remote island, he has nothing left but his bow. But the gods reveal to Odysseus that only that bow can win the Trojan War. So, Odysseus orders Neoptolemus to trick Philoctetes out of his bow. Neoptolemus obeys. Overcome with regret, however, he decides to return the bow. Neoptolemus tricks Philoctetes for serious reasons: to obey Odysseus, his commander, and to win the war. But those reasons, he concludes, cannot justify his cruelty to the anguished Philoctetes.

It seems natural to analyze this situation in terms of prima facie obligation. A prime facie obligation holds under normal circumstances. It holds all other things being equal, becoming actual unless some other moral consideration intervenes ([22, 9]). Prima facie obligations can conflict, while actual obligations cannot. A system of norms may of course yield conflicting obligations in particular cases. Faced with such a conflict, however, we can always ask what should be done to resolve the conflict. An existing system of norms may give no answer, but an ideal system surely would. We shall think of actual obligations as mandated by such an ideal normative system. (For further arguments that actual obligations cannot conflict, see [6].) So, in the case at hand, we can say that Neoptolemus has a prima facie obligation to trick Philoctetes, and another not to trick him. At most one can be actual. Neoptolemus, returning the bow, decides that his obligation not to trick Philoctetes takes precedence.
The story suggests criteria that an adequate account of obligation should satisfy.

(I) *Default Detachment.* Inferences of the form

\[
\begin{align*}
F \text{ should } G & \quad \text{If } A, \text{ then it ought to be the case that } B \\
A & \quad \text{A} \\
G & \quad \text{It ought to be the case that } B
\end{align*}
\]

should count as acceptable if no other moral considerations apply. They are not valid, but they are legitimate default inferences; we may draw the conclusion if no other rules intervene. If Neoptolemus knows only that Odysseus has commanded him to get Philoctetes’ bow to save the Greeks, and that he should obey Odysseus’s commands, he should conclude that he should get the bow.

(II) *Conditional Conflict.* In cases where conditional *prima facie* principles conflict, we should be able in general to draw no conclusions:

\[
\begin{align*}
F \text{ should } G & \quad \text{If } A, \text{ then it ought to be the case that } C \\
H \text{ should not } G & \quad \text{If } B, \text{ then it ought not to be the case that } C \\
A & \quad \text{A} \\
B & \quad \text{B} \\
\end{align*}
\]

Given that Neoptolemus has both *prima facie* obligations, logic alone should not decide between them.

(III) *Deontic Specificity.* More specific *prima facie* obligations should take precedence over less specific ones.

\[
\begin{align*}
F \text{ should } G & \quad \text{If } A, \text{ then it ought to be the case that } C \\
H \text{ should not } G & \quad \text{If } B, \text{ then it ought not to be the case that } C \\
\text{All } F \text{ are } H & \quad \text{If } A, \text{ then } B \\
A & \quad \text{A} \\
B & \quad \text{B} \\
\end{align*}
\]

Given that commands should be obeyed, that unjust commands should not be obeyed, and that Odysseus’ command is unjust, we should be able to conclude, unless other considerations intervene, that it should not be obeyed.

(IV) *Unwanted Implications.* Any system of deontic logic should handle paradoxes of deontic logic. The robber paradox, for example, like the gentle murder paradox ([8]), leads from seemingly innocuous premises to troubling conclusions:

\[
\begin{align*}
\text{a. Neoptolemus is a robber.} & \quad \text{Robbers should repent.} \\
\text{One can repent only if one has done wrong.} & \quad \text{Neoptolemus should have done wrong.}
\end{align*}
\]
Prima facie obligation

b. If you murder Jones, you should murder him gently.
   You murder Jones.
   If you murder Jones gently, you murder him.
   You should murder Jones.

The contrary-to-duty paradox confronts us with assertions that appear consistent and mutually independent, but, when symbolized in standard deontic logic, are neither ([4]):

(5) Arabella ought to visit her grandmother.
   It ought to be the case that, if she visits, she calls to say she’s coming.
   Arabella doesn’t visit her grandmother.
   If she doesn’t visit, she shouldn’t call to say she’s coming.

An adequate deontic logic should avoid, or at least explain, these conclusions.

(V) Unconditional Actual Obligation. Unconditional statements of actual obligation should not be able to conflict. This pair should be inconsistent:

(6) a has an actual obligation to F
    a has an actual obligation not to F

Neoptolemus’ decision that his obligation not to trick Philoctetes is actual is incompatible with a conclusion that his obligation to trick him is actual.

(VI) Unconditional Prima Facie Obligation. It should be possible for unconditional prima facie obligations to conflict. This pair should be consistent:

(7) a has a prima facie obligation to F
    a has a prima facie obligation not to F

We want to be able to say that Neoptolemus has, at the same time, prima facie obligations to trick Philoctetes and not to trick him.

There is a natural model for the behavior of conditional obligation: nonmonotonic inference. As Roderick Chisholm ([5]) and John Hory ([14, 15]) have argued, prima facie principles license conclusions that might have to be withdrawn in the face of further information. Like Hory, we shall use nonmonotonic logic to elucidate prima facie obligation. We shall, however, choose a different nonmonotonic system of reasoning: commonsense entailment, a modal theory of nonmonotonic reasoning developed by Nicholas Asher and Michael Morreau ([1, 2]). Hory’s system counts too many inferences valid. All the following “cut” inferences are valid in an obvious extension of Hory’s default logic to conditionals. (Here and throughout, ‘*’ indicates that a conclusion does not follow, monotonically or nonmonotonically.)
(8)  a. I should never do anything immoral.
If I never do anything immoral, I should be revered.
I should be revered. *

b. Those who remain alert should be rewarded
Police officers should remain alert
Pat is a police officer
Pat should be rewarded *

c. Those who remain alert don’t work the night shift
Police officers should remain alert
Pat is a police officer
Pat shouldn’t work the night shift *

In our theory, none of these inferences hold. Similarly, we reject defeasible transitivity inferences such as the following, which Horty’s scheme, again extended in obvious ways, counts valid:

(9) Midas is a person
    People should be generous
    Generous people should be thanked
    Midas should be thanked*

1. **Truth and Monotonic Consequence**

We symbolize formulas expressing conditional obligation as:

\[ A > OB \]

or

\[ \forall x(F(x) > OG(x)) \]

or as:

\[ A >_O B \]

or

\[ \forall x(F(x) >_O G(x)) \].

Because only initial universal quantification interests us in this paper, we shall generally discuss only the simpler, sentential forms.

Our base language is a first-order language \( L \) with \( T \) and \( \bot \), a unary obligation operator \( O \), a binary connective, \( > \), and a binary connective \( >_O \). We interpret \( O \) in the standard way: \( OA \) holds at a world \( w \) if and only if \( A \) holds in all of \( w \)’s “ideal” worlds. The connective \( > \) is a doxastic, nonmoral, generic conditional: \( A > B \) means that, where \( A \) holds, \( B \) normally holds
Prima facie obligation

too. The connective $>_O$ is a generic deontic conditional: $A>_O B$ means that, where $A$ holds, then, other things being equal, $B$ should hold.

Morally normal circumstances are those where other things are equal, morally speaking. They are morally pure or uncomplicated in the sense that only one kind of moral consideration pertains to them; conflicts involving $A$ do not arise. In ideal worlds, all actual obligations are fulfilled; in morally normal worlds, all prima facie obligations become actual. Asher and Morreau construe normality in generic contexts in terms of an accessibility relation $*$; we introduce another accessibility relation, $\bullet$, which assigns to each world $w$ and proposition $p$ a set $\bullet(w,p)$ of “good-and-simple” worlds.

We distinguish two kinds of prima facie obligation. $A>_O OB$ and $A>_O B$ mean different things. Each plays an important role in moral reasoning. But the roles differ. Sentences of the form $A>_O B$ are ceteris paribus principles. They say that, if $A$, then, all other things being equal, $B$ is obligatory. They are thus constitutive: $A>_O B$ implies that the truth of $A$ is a reason for doing $B$. In other words, roughly, $A>_O B$ implies that, if $A$ were true, $B$ would be obligatory because $A$. It may be possible to analyze this constitutive deontic conditional in terms of unary obligation and a causal conditional stronger than $>$. Since we have not yet worked out the theory of such a conditional, however, we have opted for the safer course of a conditional obligation operator.

Sentences of the form $A>_O OB$, in contrast, generalize about circumstances in which $A$ is true, saying that, generally, where $A$ is true, then $B$ is obligatory. They are thus epistemic: they imply that the truth of $A$ is a good indicator that $B$ is obligatory. To put it another way, sentences of the form $A>_O B$ say what is true ceteris paribus, all other things being equal; those of the form $A>_O OB$ say what is true normally.

Constitutive principles are nonmonotonically stronger than epistemic principles. If the truth of $A$ is a reason for $B$’s being obligatory, then, normally, the truth of $A$ is a good indicator that $B$ is obligatory. The reverse, however, does not hold. Some good indicators are reasons for what they indicate, but some are not. Constitutive and epistemic principles thus have precisely the logical relation that causal assertions and generalizations have. If one kind of event causes another, then, normally, events of the first kind will generally be followed by events of the second. This connection is only nonmonotonic; in some worlds there may routinely be countervailing and overriding causes. In any case, generalizations do not imply causal assertions even nonmonotonically. That events of one kind are routinely followed by events of another kind does not allow us to conclude, monotonically or otherwise, that events of the first kind cause events of the second kind. In-
deed, such an inference counts traditionally as a post hoc ergo propter hoc fallacy.

Nevertheless, constitutive and epistemic principles have much the same logic, as we might expect from their sharing the same forms of expression in natural language. Truth conditionally, however, they are distinct in philosophically important ways. Here are five illustrations of the difference between constitutive and epistemic principles in moral contexts:

(a) Kantians and utilitarians both defend many traditional moral principles. They not only argue for them on different grounds, however; they read the principles themselves differently. Traditional principles find support in utilitarian theory, John Stuart Mill argues, because they are generalizations about good action, good character, and the good life, based on a vast range of experience. In short, they are generalizations about actual obligation. A principle of the form

If A, then it ought to be the case that B

from this perspective asserts that, where A holds, there is normally an actual obligation that B. This is well-represented by A > OB.

Kantians, in contrast, see such a principle as a hypothetical imperative or maxim justified by the categorical imperative, "on grounds valid for every rational being as such" ([16]). They may interpret it as strictly universal or as nonmonotonic: in conflicts between reason and inclination, "the universality of the principle is changed into mere generality, whereby the practical principle of reason meets the maxim halfway." Even in the latter case, however, a situation’s being such that A itself provides a reason for an obligation that B. This corresponds to a constitutive ceteris paribus principle, well-represented by A >OB.

The Kantian reading of moral principles tends to imply the utilitarian reading. If promises ought to be kept in the sense that the promise itself creates, ceteris paribus, an obligation to keep it, then we can expect that, if an agent makes a promise, he or she generally has an actual obligation to keep it. The implication, however, is nonmonotonic. There may be unusual worlds in which promises, while still having real moral force, are generally outweighed by other moral factors.

(b) Mill argues for the importance of traditional moral rules in defending utilitarianism against common objections. There is not enough time, his hypothetical opponent objects, to perform all the calculations that utilitarianism requires. Not so, Mill replies; "there has been ample time, namely, the whole past duration of the human species. During all that time mankind have been learning by experience the tendencies of actions" ([19]). We ought
to obey traditional moral rules, in general, because they are products of experience, experimentation, and reasoned choice. In our terms, Mill advocates a principle of the form $\forall x (\text{traditional} - \text{rule}(x) > \text{follow}(x))$: traditional rules generally ought to be followed. Being traditional is not itself a right-making property; it is merely a good indicator of such properties.

Edmund Burke, in contrast, might be read as holding the generally stronger position that traditional rules ought to be followed precisely because they are traditional. He maintains, in other words, that being traditional is itself right-making, in the sense that traditional rules, *ceteris paribus*, ought to be followed: $\forall x (\text{traditional} - \text{rule}(x) >_O \text{follow}(x))$. Again, this constitutive position is stronger, but only nonmonotonically. The tradition could be so corrupt that other defects generally outweigh the benefits of tradition.

(c) Many writers argue for affirmative action programs as compensating for prior injustice. Some (e.g., [20]) defend the use of racial criteria as an administrative convenience: blacks, for example, are generally worse off than they might have been had there been no racial discrimination. Some blacks may not be, however; their inclusion is justified only by the headaches it would require to identify them. On this view, we should, *prima facie*, compensate blacks for discrimination, in the sense that blacks generally ought to be compensated. This is precisely what we would represent as $\forall x (\text{black}(x) > \text{Ocompensated}(x))$. Other writers (e.g., [24], [7]) defend the use of racial criteria differently. They contend that, because discrimination was directed at blacks as a group, blacks deserve compensation as a group. They contend, in other words, that being black is itself a reason for compensation, not merely a convenient proxy for the reason. All other things being equal, they maintain, blacks ought to be compensated. This we represent as the constitutive $\forall x (\text{black}(x) >_O \text{compensated}(x))$.

These two positions are not equivalent. Ordinarily, the latter is stronger; to say that being black itself justifies receiving compensation implies, but is not implied by, the assertion that blacks generally ought to be compensated. The implication is nonmonotonic, however, for it is possible to imagine a situation in which blacks ought, *ceteris paribus*, to be compensated, but in which, due to grave budget problems, for example, it is not true that blacks generally ought to be compensated.

(d) Nominalists have sometimes wanted to understand a sentence such as (10)a by translating it into the putatively equivalent (10)b or, to simplify somewhat, (10)c:

(10) 
   a. Honesty is a virtue
   b. Honest acts are virtuous
   c. Honest acts ought to be done.
Those with other motives have sometimes agreed, arguing that there must be some logical relation between talk of honesty and talk of honest actions, or we could never apply abstract sentences to concrete situations ([23]). Platonists have objected that honesty is only one of the virtues; not all honest actions ought to be performed. Clearly, the nominalist needs to say that \textit{honesty is a virtue} is equivalent to \textit{Honest acts, other things being equal, ought to be done}. This we might represent as $\forall x(\text{honest}(x) >_O \text{do}(x))$. It would not do to translate the sentence as the epistemic $\forall x(\text{honest}(x) > \text{Odo}(x))$, for this says something nonmonotonically weaker— that honest acts normally ought to be done. It might be contingently true, for example, that acts done with a smile normally ought to be done, but that would not make \textit{Smiling is a virtue} true. That an act is honest is itself a reason for doing it, but that it is done with a smile is not.

(e) The distinction between constitutive and epistemic principles is crucial to understanding the consequences of moral conflict. Paul Pietroski ([21]), following Bernard Williams ([27]), contends that regret is justified if and only if an agent has and fails to fulfill a \textit{prima facie} obligation. Such obligations carry moral cost even when overridden by other considerations. So, Neoptolemus should regret his deception of Philoctetes, whether or not that deception is ultimately justified. And, after returning the bow, he should regret disobeying Odysseus, even if his decision is correct. It is thus plausible to hold that \textit{prima facie} obligations understood as stemming from \textit{ceteris paribus} principles, having the form $A >_O B$, justify regret. But the same does not hold of such obligations understood as stemming from generalizations having the form $A > OB$. Suppose acts performed with a smile generally ought to be done. It does not follow that one ought to regret failures to smile.

The distinction between constitutive and epistemic generics is not limited to deontic contexts; it extends to all generics. The contrast between causal assertions and generalizations is an analogous instance of a more general distinction, recognition of which goes back at least as far as Aristotle’s discussion of \textit{qua} phrases. The account we offer here is a special case of a more general theory we hope to develop elsewhere.

1.1. Frames

We now express these ideas more formally. Throughout the following we take $w, w'$, etc. to be arbitrary worlds in $W$ and $p, q$, etc. to be arbitrary subsets of $W$. (We identify propositions with sets of worlds.) $\oplus w$ is the set of $w$'s ideal worlds; $\oplus(X)$, where $X$ is a set of worlds, is $\cup \oplus(w)$ for $w \in X$; $(w, p)$ is the set of $w - p$-normal worlds, and $(w, p)$ is the set of $w - p$-good-
and-simple worlds. To preview our truth conditions for constitutive and epistemic principles: $A >_O B$ is true in world $w$ if $B$ is true in all good-and-simple worlds relative to $A$ and $w$; that is, if (suppressing references to models and assignments) $\bullet(w, [A]) \subseteq [B]$. A generic $A > B$ is true in $w$ if $B$ is true in all normal worlds relative to $A$ and $w$; if, in other words, $(w, [A]) \subseteq [B]$. A unary obligation statement $OA$ is true in $w$ if $A$ holds in all ideal worlds, i.e., if $\oplus(w) \subseteq [A]$. It follows that an epistemic principle $A > OB$ is true in $w$ if $\ominus*(w, [A]) \subseteq [B]$.

**Definition 1.1.** An $L$ frame $\mathfrak{S}$ is a quintuple $(W, D, \ominus, \star, \bullet)$, where $W$ is a nonempty set of worlds, $D$ is a nonempty set of individuals, $\ominus$ is a function from $W$ to $\wp(W)$, $\star$ is a function from $W \times \wp(W)$ to $\wp(W)$, and $\bullet$ is a function from $W \times \wp(W)$ to $\wp(W)$.

**Definition 1.2.** A world $w$ is normal (written $Nw$) iff $\exists w' w \in \star(w', W)$.

**Definition 1.3.** Proper $L$ frames are frames obeying:

(a) the **Seriality** constraint: $\ominus(w) \neq \emptyset$;

(b) the **Facticity** constraint: $\star(w, p) \subseteq p$; and

(c) the **Disjunction** constraint: $\star(w, p \cup q) \subseteq \star(w, p) \cup \star(w, q)$.

(d) the **Ideals Good-and-Simple** constraint: $Nw \rightarrow \ominus \star(w, p) \subseteq \bullet(w, p)$

(e) the **Deontic Disjunction** constraint: $\bullet(w, p \cup q) \subseteq \bullet(w, p) \cup \bullet(w, q)$.

The seriality constraint guarantees that $O$ is an actual obligation operator, admitting no moral dilemmas. The facticity constraint stipulates that $p$ is one of the things that normally hold when $p$ holds, thus making sentences such as *Criminals are criminals* valid. The Disjunction and Deontic Disjunction constraints guarantee the validity of inferences such as

(11) If Dudley loves Nell, he should marry Nell
    If Dudley loves his horse, he should marry Nell
    If Dudley loves either Nell or his horse, he should marry Nell

and analogous nondeontic forms. The "ideals good and simple" constraint allows constitutive principles to imply epistemic principles nonmonotonically.

Facticity and the Disjunction schema entail both generic specificity, validating Penguin Principle inferences such as

(12) Birds fly
    Penguins are birds
    Penguins don't fly
    Tweety is a penguin
    Tweety doesn't fly
and the closely related deontic specificity, validating arguments such as (3) above and satisfying the Deontic Specificity criterion (III).

**FACT 1:** \[ p \subseteq q \& * (w, p) \cap * (w, q) = \emptyset \Rightarrow * (w, q) \cap p = \emptyset. \]

**Proof.** Assume Facticity and Disjunction. Assume further that \( p \subseteq q \& * (w, p) \cap * (w, q) = \emptyset. \) Since \( p \subseteq q, p \cap q = p. \) So, \( q = p \cup (q - p). \) By Disjunction, \( * (s, q) \subseteq * (s, p) \cup * (s, q - p). \) Since \( * (s, q) \cap * (s, p) = \emptyset, * (s, q) \subseteq * (s, q - p). \) But, by facticity, \( * (s, q - p) \subseteq q - p \subseteq -p. \) It follows that \( * (s, q) \cap p = \emptyset. \)

We cannot yet show formally that fact 1 implies the nonmonotonic validity of (12), for we have not yet defined nonmonotonic validity. But the central idea is easy to explain. Since birds fly, but penguins don't, normal penguins are not normal birds. But all penguins are birds. It follows, by fact 1, that no penguins are normal birds. Thus, Tweety is not a normal bird. In evaluating (12) for nonmonotonic validity, to anticipate, we assume as much normality as possible, given the premises. We cannot consistently assume Tweety to be a normal bird. But we can assume Tweety to be a normal penguin, and normal penguins do not fly. So, we can conclude, nonmonotonically, that Tweety does not fly.

**FACT 2:** \( \oplus \) and \( \bullet \) obey the following:

a. \( p \subseteq q \rightarrow \oplus p \subseteq \oplus q \)

b. \( \oplus (p \cup q) = \oplus p \cup \oplus q \)

c. \( \oplus (p \cap q) \subseteq \oplus p \cap \oplus q \)

d. \( p \neq \emptyset \rightarrow \oplus p \neq \emptyset \)

e. \( Nw \rightarrow (* (w, p) \neq \emptyset \rightarrow \bullet (w, p) \neq \emptyset) \)

Fact 2 is chiefly of instrumental value. But 2a is crucial to the closure of obligation under logical consequence. 2d is a form of seriality, guaranteeing that nonempty sets of worlds have nonempty sets of ideals. It is important to Deontic Specificity, as we shall see from the proof of Fact 3. 2e says that, in normal worlds, nonempty sets of normal worlds are accompanied by nonempty sets of good-and-simple worlds. This is important for deducing just about anything from constitutive deontic principles.

**FACT 3:** \( Nw \rightarrow [(p \subseteq q \& \bullet (w, p) \cap \bullet (w, q) = \emptyset) \Rightarrow * (w, q) \cap p = \emptyset]. \)
Prima facie obligation

Proof. Suppose \( M, w, p \subseteq q \) and \( \bullet (w, p) \cap \bullet (w, q) = \emptyset \). By the "ideals good-and-simple" constraint, \( \odot \star (w, p) \cap \odot \star (w, q) = \emptyset \). Fact 2c allows us to deduce \( \odot (\star (w, p) \cap \star (w, q)) = \emptyset \). Applying fact 2d, we obtain \( \star (w, p) \cap \star (w, q) = \emptyset \). This, and fact 1, yield \( \star (w, q) \cap p = \emptyset \).

Fact 3 says that, if \( p \) is more specific than \( q \), but the two give rise to conflicting obligations, then \( p \) is not a normal \( q \). To see how this yields Deontic Specificity, assume that commands should be obeyed, unjust commands should not be obeyed, and Odysseus' command is unjust. Odysseus' command gives rise to conflicting obligations, so it is not a normal command. To obtain nonmonotonic consequences of the premises, we assume as much normality as is consistent with the premises. We cannot assume that Odysseus' command is a normal command, but we can assume that it is a normal unjust command. Since unjust commands should not be obeyed, we conclude that Odysseus' command should not be obeyed.

1.2. Models

Definition 1.4. A base model is a tuple \( \langle W, D, \odot, \star, \bullet, \square \rangle \), where \( \langle W, D, \odot, \star, \bullet \rangle \) is a proper \( L \) frame and \( \square \) is a function from nonlogical constants of \( L \) to intensions (which, in turn, are functions from worlds to extensions).

A variable assignment \( \alpha \) is a function from variables into \( D \), the domain of a base model; all worlds have the same domain. Similarly, for the sake of simplicity, we assume that all objects in \( D \) have names in \( L \), and we assume bivalence. We define satisfaction on everything but the connectives \( > \) and \( >_O \) in the standard way. Truth is satisfaction under all assignments: \( M, w \models A \iff \forall \alpha M, w, \alpha \models A \). Where \( \Gamma \) is a set of formulas, \( M, w, \alpha \models \Gamma \iff M, w, \alpha \models A \) for all \( A \in \Gamma \). \( \llbracket A \rrbracket_{M, \alpha} \), the proposition \( A \) expresses under \( \alpha \) in \( M \), is the set of worlds in which \( \alpha \) satisfies \( A \): \( \{ w \in W : M, w, \alpha \models A \} \). If \( A \) is a sentence, and \( \Gamma \) a set thereof, we write \( \llbracket A \rrbracket_M \) for \( \{ w \in W : M, w \models A \} \) and \( \llbracket \Gamma \rrbracket_M \) for \( \cap \{ \llbracket A \rrbracket_M : A \in \Gamma \} \).

Definition 1.5.

\[
\delta_w(t) = [t]_w, \text{ if } t \text{ is a constant, and } \alpha(t), \text{ if } t \text{ is a variable}
\]

\[
M, w, \alpha \models R t_1 \ldots t_n \iff \langle \delta_w(t_1), \ldots, \delta_w(t_n) \rangle \in \llbracket R \rrbracket_w
\]

\[
M, w, \alpha \models \neg A \iff \text{not } M, w, \alpha \models A
\]

\[
M, w, \alpha \models A \land B \iff M, w, \alpha \models A \text{ and } M, w, \alpha \models B
\]

\[
M, w, \alpha \models \forall x A \iff M, w, \alpha^t \models A \text{ for all } \alpha^t \text{ such that, for all } v \neq x, \alpha(v) = \alpha^t(v)
\]
If $A$ is an $L_O$ formula, and $\Sigma$ a set thereof, then $\Sigma \models A$ iff, for all $M, w, \alpha$, if $M, w, \alpha \models \Sigma$, then $M, w, \alpha \models A$. $A$ is valid, $\models A$, iff $\emptyset \models A$.

### 1.3. Axiomatization

We define derivability using the axioms and rules:

- (A1) Truth-functional $L$-tautologies
- (A2) $\forall x A \rightarrow A(t/x)$ for any term $t$
- (A3) $\forall x A \leftrightarrow \exists x \neg A x$
- (A4) $\forall x (A \rightarrow B) \rightarrow (\exists x A \rightarrow B)$ for $x$ not free in $B$.
- (A5) $\forall x (A > B) \rightarrow (A \forall x B)$, for $x$ not free in $A$.
- (A6) $A > A$
- (A7) $((A > C) \& (B > C)) \rightarrow ((A \lor B) > C)$
- (A8) $\neg O \perp$
- (A9) $O(A \rightarrow B) \rightarrow (OA \rightarrow OB)$
- (A10) $A >_O \top$
- (A11) $\forall x (A >_0 B) \rightarrow (A >_0 \forall x B)$, for $x$ not free in $A$.
- (A12) $((A >_0 C) \& (B >_0 C)) \rightarrow ((A \lor B) >_0 C)$
- (A13) $(A >_0 B) \rightarrow (\top > (A >_0 OB))$

**Rules:**

- (R1) $\vdash A$ and $\vdash A \rightarrow B \Rightarrow \vdash B$
- (R2) $\vdash (B_1 \& \ldots \& B_i) \rightarrow B \Rightarrow \vdash (A >_0 B_1 \& \ldots \& A >_0 B_i) \rightarrow A > B$
- (R3) $\vdash A \rightarrow B(t/x) \Rightarrow \vdash A \rightarrow \forall x B$, where $t$ is a constant not in $A$ or $B$
- (R4) $\vdash A \Rightarrow \vdash A(t/x)$ where $t$ is a term not in $A$
- (R5) $\vdash A \leftrightarrow B$ and $A$ a subformula of $C \Rightarrow \vdash C \leftrightarrow C(B/A)$
- (R6) $\vdash A \Rightarrow \vdash OA$
- (R7) $\vdash (B_1 \& \ldots \& B_i) \rightarrow B \Rightarrow \vdash (A >_0 B_1 \& \ldots \& A >_0 B_i) \rightarrow A >_0 B$
- (R8) $\vdash A \leftrightarrow B \Rightarrow \vdash A > C \leftrightarrow B > C$
- (R9) $\vdash A \leftrightarrow B \Rightarrow \vdash A >_0 C \leftrightarrow B >_0 C$

**Completeness Theorem:** $\Sigma \models A$ iff $\Sigma \vdash A$. 

$$M, w, \alpha \models OA \iff \oplus(w) \subseteq \llbracket A \rrbracket_{M, \alpha}$$

$$M, w, \alpha \models A > B \iff \star(w, \llbracket A \rrbracket_{M, \alpha}) \subseteq \llbracket B \rrbracket_{M, \alpha}.$$
2. Nonmonotonic consequence

To determine the nonmonotonic consequences of a set of premises,

Assume everything to be as normal as possible, given the premises (and only the premises); ask what follows.

Assume, initially; that we have no information. If we think of our information state as a set of worlds compatible with the information we have, then the information state is simply \( W \). Each premise gives us information. In monotonic logic, the first premise, say \( A_1 \), restricts the information state to \([A_1]\); each \( A_{n+1} \) changes the state \( s_i \) obtained by processing the first \( n \) premises to \( s_i \cap [A_{n+1}] \). In nonmonotonic inference, each premise also restricts \( s_i \) so that everything is as doxastically or deontically normal as possible. That is, it restricts the state to the subset of \( w - [A_{n+1}] \)-normal (or \( w - [A_{n+1}] \)-good-and-simple) worlds, unless the result of this further restriction would be an empty set of worlds.

The function * specifies what is generically normal in each world. We use it to characterize a normalization function on information states.

**Definition 2.1.** \(* (s, p) = \cup_{w \in s} * (w, p); \bullet (s, p) = \cup_{w \in s} \bullet (w, p)\**

**Definition 2.2.**

\[
N(s, p) = (s \cap W - p) \cup (s \cap * (s, p)) \text{ if } s \cap * (s, p) \neq \emptyset;
\]

\[
s \quad \text{otherwise.}
\]

Normalizing information state \( s \) with respect to proposition \( p \) restricts \( s \) to worlds that are not \( p \)-abnormal. If \( s \subseteq p \), normalization yields \( s \cap * (s, p) \), so long as this set is nonempty. If the set would be empty, then normalization does nothing. Intuitively, this corresponds to a situation we already know to be abnormal.

2.1. Deontic normalization

Practical reasoning requires a complication of information states and, correspondingly, of normalization. Because we frequently reason using both doxastic generics (such as *People are selfish*) and deontic generics (such as *People ought to be altruistic*), we must keep careful track of two different kinds of information. An agent will typically have information about the way the world is and about the way it ought to be. We begin by defining deontic normalization and nonmonotonic consequence for arguments without nested deontic operators, to keep the intuitive picture clearer; a small change extends the scheme to arguments with nesting of arbitrary finite depth.
We represent information states as pairs of sets of worlds. One is factual, the other normative. Updating an information state involves adding to its information about the way the world is, or to its information about the way the world ought to be, or both. Normalizing an information state, similarly, involves

(a) assuming that the world is as doxastically normal as possible, given the state's picture of how the world is,

(b) assuming that everything is as deontically normal as possible in ideal worlds, given the state's picture of how the world is and ought to be, and

(c) assuming that everything is as doxastically normal as possible in ideal worlds.

Part (a) proceeds just as above. To effect (b), we assume that circumstances are good-and-simple whenever possible. The set \( t \) is the set of "normative" worlds—the worlds, that is, that determine obligations—and \( s \) is the set of "doxastic" worlds of the information state in which the nonmoral statements are true:

**Definition 2.3.**

\[
\Omega(s, t, p) = \{w \in t : w \in \bullet(s, p)\} \quad \text{if} \quad \bullet(s, p) \cap t \neq \emptyset,
\]

\[
*\bullet(s, p) \cap s \neq \emptyset, \text{ and } p \subseteq s
\]

= \( t \) \quad \text{otherwise.}

The effect of updating a set \( t \) of normative worlds deontically with respect to \( p \), given an information state \( s \), then, is to restrict \( t \) to worlds that are morally normal with respect to \( p \). Like doxastic normalization, the update occurs only if certain conditions are satisfied. First, the update must not result in absurdity. This makes the obligation *prima facie*. Second, it must be doxastically possible that we are dealing with a normal \( p \) situation. If we are not, then *prima facie* obligations conditional on \( p \) are moot. Finally, the information state \( s \) must contain \( p \). Conditional obligations apply only when their antecedents hold.

To effect (c), we apply doxastic normalization, as defined above, to the result of (b). This step is necessary to capture the nonmonotonic validity, for example, of the inference from \( OA \) and \( O(A > B) \) to \( OB \).
2.2. Information models

**Definition 2.4.** The update function $+$: $\wp(W) \times \wp(L) \rightarrow \wp(W)$ for $M = \langle W, D, \oplus, , \circ, \rangle$ is such that, for all information states $s \in \wp(W)$, and for all sets $\Gamma$ of $L$ sentences: $s + \Gamma = s \cap [\Gamma]_M$.

**Definition 2.5.** A model $M$ and information state $s$ support an $L$-sentence $A$ (written $M, s \models A$) if and only if $s \subseteq [A]_M$.

**Definition 2.6.** A total assignment $v$ of truth values to $L$-sentences conforms to $M$ and $(s, t)$ iff

If $A$ is atomic or of the form $B > C$ or of the form $B >o C$:

\[
\begin{align*}
   v(A) &= 1 \text{ if } M, s \models A \\
   v(A) &= 0 \text{ if } M, s \models \neg A
\end{align*}
\]

If $A$ is of the form $OB$:

\[
\begin{align*}
   v(A) &= 1 \text{ iff } t \models B \\
   v(A) &= 0 \text{ iff } not \ t \models B
\end{align*}
\]

If $A$ is of the form $B \& C$:

\[
\begin{align*}
   v(A) &= 1 \text{ iff } v(A) = v(B) = 1 \\
   v(A) &= 0 \text{ iff } v(A) = 0 \text{ or } v(B) = 0
\end{align*}
\]

If $A$ is of the form $\neg B$

\[
\begin{align*}
   v(A) &= 1 \text{ iff } v(B) = 0 \\
   v(A) &= 0 \text{ iff } v(B) = 1
\end{align*}
\]

If $A$ is of the form $\forall x B$:

\[
\begin{align*}
   v(A) &= 1 \text{ iff } v(Ac/x) = 1 \text{ for all } c \in L: \\
   v(A) &= 0 \text{ iff } v(Ac/x) = 0 \text{ for some } c \in L:
\end{align*}
\]

**Definition 2.7.** A model $M$ and information state $(s, t)$ support an $L$-sentence $A$ $(M, (s, t) \models A)$ iff $A$ is true on every total assignment of truth values to $L$-sentences conforming to $M$ and $(s, t)$. 
The idea of assuming nothing but the premises of an argument underlies nonmonotonic reasoning. To model this, begin with a state of complete ignorance, in other words, with the set $W^*$ of all possible worlds in the canonical model $M^*$. It supports only logical validities. To evaluate an argument, update it with the argument’s premises. Then, normalize with a unified procedure, assuming as much normality as is consistent with the premises. Let $\Gamma$ be a set of sentences, $P$ an enumerable set of propositions, and $\nu$ any one-to-one correspondence between $P$ and the natural numbers, or some initial segment thereof. The $P$-normalization chain with respect to $\nu$ that begins from a state that assumes nothing but the premises $\Gamma$ is the sequence with $(W^* + \Gamma, W^*)$ as its first element. Here $W^* + \Gamma$ is the factual set of worlds; we know nothing about the actual world except what $\Gamma$ has told us. $W^*$, the second element of the pair, is the normative set of worlds. Prior to processing the information in $\Gamma$, we know nothing about how things ought to be.

The recursion we describe cycles through $P$ as enumerated by $\nu$, starting over again when $P$ is exhausted and after each limit ordinal. Let $p^\alpha$ be the $\alpha$th proposition in the cycle determined by $\nu$, and let 1st and 2nd map ordered pairs into their first and second constituents, respectively.

**Definition 2.8.** The $P$-normalization chain with respect to $\nu$ that begins from $s$ is the sequence:

$$
\begin{align*}
  s^0_\nu &= (s, W^*) \\
  s^{\alpha+1}_\nu &= (N(1st(s^\alpha_\nu), p^\alpha), N(\Omega(1st(s^\alpha_\nu), 2nd(s^\alpha_\nu), p^\alpha), p^\alpha)) \\
  s^\lambda_\nu &= (\cap_{\mu \in \lambda}, 1st(s^\mu_\nu), \cap_{\mu \in \lambda}, 2nd(s^\mu_\nu))
\end{align*}
$$

Each state has a successor obtained by normalizing with the next proposition. At limit ordinals, we take the intersection of all preceding states.

Each state and enumeration $\nu$ of $P$ determines a $P$-normalization chain beginning with $(W^* + \Gamma, W^*)$ for $\Gamma$. Since normalization monotonically depletes sets of possible worlds, every chain $C$ reaches a fixed point $C^*$: for each $s$ and $\nu$, there is an ordinal $\alpha$ such that, for all larger ordinals $\beta$, $s^{\beta}_\nu = s^{\alpha}_\nu$.

We define nonmonotonic consequence:

**Definition 2.9.** $\Gamma \models_P A$ iff, for any $P$-normalization chain $C$ beginning from $W^* + \Gamma, M^*, C^* \models A$.

We are interested, specifically, in the case where $P$ is the set of all instantiations of antecedents of $\succ$- and $\succ_0$-conditionals in $\Gamma$ to those individual constants appearing in the premises and new constants serving as witnesses.
for existential premises. Call this set of instantiations \( \text{Prop}(\Gamma) \). Only the elements of \( \text{Prop}(\Gamma) \) lead to normalizations that affect conclusions to be derived from \( \Gamma \). So, only they need to be considered in defining nonmonotonic entailment:

**Definition 2.10.** \( \Gamma \models A \) iff \( \Gamma \models_{\text{Prop}(\Gamma)} A \).

### 3. Nonmonotonic inferences

Despite the complexity of its definition, the nonmonotonic validity of most arguments can be decided by a simple finite procedure. To illustrate, we show that our theory satisfies the criteria we developed earlier.

First, the theory satisfies criterion (I). Default detachment is nonmonotonically valid.

(13) Neoptolemus is a person

People should be kind

Neoptolemus should be kind

(14) Neoptolemus is a person

Neoptolemus should not be kind

People should be kind

Neoptolemus should be kind*

(Recall that '*' indicates a conclusion that does not follow.) (13) is acceptable, but (14) fails.

We predict this correctly. For both arguments, \( \text{Prop}(\Gamma) = \{ \text{Neoptolemus is a person} \} \). So, begin with the premise that Neoptolemus is a person. Restrict the factual set to worlds in which this holds. Process the premise that people should be kind. This does not affect the factual set of worlds, but it restricts the normative set to those in which everyone who is a person, according to the factual set, is kind. In particular, Neoptolemus is kind in all such worlds. (In this and the following examples, for convenience, we represent pairs of information states that are the fixed points of normalization chains by writing the relevant sentences true in the factual and normative information states.)

<table>
<thead>
<tr>
<th>Factual Set</th>
<th>Normative Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neoptolemus is a person</td>
<td>People (in the factual set) are kind</td>
</tr>
<tr>
<td></td>
<td>Neoptolemus is kind</td>
</tr>
</tbody>
</table>

So, the result of normalizing is a state that supports the conclusion, *Neoptolemus should be kind*—Neoptolemus is kind in every world in the state’s normative set.
To see why (14) fails, again assume that we begin with a state of ignorance. The premise that Neoptolemus is a person restricts the factual set to worlds in which Neoptolemus is a person. The premise that Neoptolemus should not be kind restricts the normative set to worlds in which Neoptolemus is not kind.

<table>
<thead>
<tr>
<th>Factual Set</th>
<th>Normative Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neoptolemus is a person</td>
<td>Neoptolemus is not kind</td>
</tr>
</tbody>
</table>

Now, the premise that people should be kind restricts the normative set, for each entity that is a person according to the factual set, to worlds in which that person is kind—except where the result of the restriction would be absurdity, that is, the empty set of worlds. Neoptolemus is such an exception.

We also satisfy criterion (II), Conditional Conflict. The reasoning showing the invalidity of (14) extends easily to show the invalidity of the following arguments, which are deontic versions of the Nixon diamond:

(15) Judges should be partial to no one
     Fathers should be partial to their children
     Sam is a judge and a father
     Sam should be partial to no one* (to his children*)

Neither conclusion gains support in every fixed point, so both arguments are invalid. Normalization chains that take the first premise before the second yield fixed points in which Sam is partial to no one. Normalization chains that do the reverse yield fixed points in which Sam is partial to his children. In each fixed point, Sam is either partial to no one or partial to his children. This establishes the nonmonotonic validity of:

(16) Judges should be partial to no one
     Fathers should be partial to their children
     Sam is a judge and a father
     Sam should be partial to his children or to no one

We are now in a position to see more precisely that our theory satisfies criterion (III), Deontic Specificity. Consider:

(17) Commands should be obeyed
     Unjust commands are commands
     Unjust commands should not be obeyed
     Odysseus’ command is unjust
     Odysseus’ command should not be obeyed
We begin in ignorance and update with the premises, restricting the factual set to worlds in which Odysseus' command is unjust. Because commands should be obeyed, but unjust commands should not be obeyed, we encounter a conflict. By Fact 3 above, however, we can conclude from this that normal commands are not unjust; this in turn implies that Odysseus' command is not a normal command. We then normalize, assuming things to be as normal, generically and deontically, as possible. Although Odysseus' command is an abnormal command, we may still assume that it is a normal unjust command. So, we restrict the normative set to worlds in which unjust commands are not obeyed.

<table>
<thead>
<tr>
<th>Factual Set</th>
<th>Normative Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>O's command is unjust</td>
<td>Unjust commands (in factual set) are not obeyed</td>
</tr>
<tr>
<td>O's command is not obeyed</td>
<td></td>
</tr>
</tbody>
</table>

In no world of the normative set, therefore, is Odysseus' command obeyed.

Finally, we satisfy criterion (IV), Unwanted Implications. We follow Marvin Belzer and Barry Loewer ([17, 18]) in holding that the robber and gentle murder paradoxes revolve around issues of judgment and deliberation, rather than around closure of logical consequence (as in [10, 11, 12]). But we have a rather different view of that contrast. Richmond Thomason ([25, 26]) distinguishes the context of deliberation, in which one must decide what to do and takes the facts simply as given, from the context of judgment, in which one can reflect on the moral status of the facts themselves. According to Thomason, Belzer, and Loewer, the paradoxes reveal that the context of deliberation presupposes that whatever is, is right: robbers should rob and murderers should murder. This happens because, in deliberation, we restrict our attention to possible courses of events that share our history. It is harmless because we are not deliberating about the present or matters otherwise already settled; the context of judgment is available for doing that. The essential contrast between judgment and deliberation is, on this view, a difference in what we take as settled.

From our perspective, however, the contrast is one of including or suppressing information about the moral status of certain facts, whether they are in the present or the past, whether they are settled or unsettled. The robber and gentle murder paradoxes have similar forms. For both, let B monotonically imply A:

\[
\begin{align*}
\text{Robber} & : & C & \quad \text{Gentle murder} & : & A \\
& & C >_O B & & A >_O B \\
& & OA & & OA
\end{align*}
\]

The gentle murder paradox is just the special case of the robber paradox in
which \( C = A \). The conclusions follow nonmonotonically. To take just the robber paradox:

\[
\begin{array}{c|c}
\text{Factual Set} & \text{Normative Set} \\
C & B, A \\
\end{array}
\]

The first premise restricts the factual set to worlds in which \( C \) is true. The second restricts the normative set to those in which \( B \) is true. But, since \( B \) monotonically implies \( A \), \( A \) is also true in every world in the normative set. This certainly seems counterintuitive. In the context of deliberation, however, we ignore the moral status of the facts as they are now; we simply take them for granted and ask what ought to be done about them. We assume that the robber has robbed, and you will murder Jones. We ask what ought to be done \textit{given that information}.

In the context of judgment, however, we consider the moral status of the facts. Making this explicit requires adding a moral premise to each argument form (where, again, \( B \) monotonically implies \( A \)):

\[
(19) \quad \begin{array}{c|c}
\text{Factual Set} & \text{Normative Set} \\
C & \neg A \\
C >_O B & A >_O B \\
OA* & OA* \\
\end{array}
\]

Now, the conclusions do not follow, even nonmonotonically. Normalization with \( A \) fails to yield anything new, because every world in the normative set already verifies something incompatible with the result of normalizing. For the robber paradox:

\[
\begin{array}{c|c}
\text{Factual Set} & \text{Normative Set} \\
C & \neg A \\
\end{array}
\]

In the context of deliberation, then, the robber should repent, and you should murder Jones gently. The question of whether the robber should have become a robber and whether you should murder Jones does not arise. In the context of judgment, however, the robber should not have been in a position to repent, and you should not murder Jones, gently or otherwise. This conclusion does not depend on ignoring the fact that the robber is a robber, or treating it as unsettled whether or not you will murder Jones. It does not involve a thought experiment of going back in time to a point before the truth of \( C \) or \( A \) above. It simply depends on recognizing that, although they are true, they should not be.

The most distinctive features of our approach emerge in its solution to Roderick Chisholm's contrary-to-duty paradox. As we represent (5) above:

\[
(20) \quad OA \\
O(A > B)
\]
\[ \neg A \]
\[ \neg A >_O \neg B \quad \text{(or } \neg A > O \neg B) \]

In standard deontic logic, these assertions are inconsistent. The first two, in an instance of deontic detachment, imply \( OB \), while the last two, in an instance of factual detachment, imply \( O \neg B \) ([13, 17]). Many have drawn the moral that one must choose between these two modes of detachment.

Our nonmonotonic logic allows us to have both factual and deontic detachment without inconsistency. It is easy to verify that both are nonmonotonically valid:

(21)  

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>OA</td>
<td>( O(A &gt; B) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( OB )</td>
</tr>
<tr>
<td>b.</td>
<td>( \neg A )</td>
<td>( \neg A &gt;_O \neg B )</td>
</tr>
</tbody>
</table>

(21)b is analogous to (13) above. The premises of (21)a restrict the normative set to worlds in which \( A \), \( A > B \), and– because we normalize the normative set as well as the factual set– \( B \) are true.

Nevertheless, the assertions in (20) are not inconsistent. Neither of the following is valid:

(22)  

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>OA</td>
<td>( O(A &gt; B) )</td>
</tr>
<tr>
<td></td>
<td>( \neg A )</td>
<td>( \neg A &gt;_O \neg B )</td>
</tr>
<tr>
<td></td>
<td>( OB^* )</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>OA</td>
<td>( O(A &gt; B) )</td>
</tr>
<tr>
<td></td>
<td>( \neg A )</td>
<td>( \neg A &gt;_O \neg B )</td>
</tr>
<tr>
<td></td>
<td>( O \neg B^* )</td>
<td></td>
</tr>
</tbody>
</table>

If we begin by restricting the normative set to worlds in which \( A \) is true, as \( OA \) demands, and then process the second premise, we derive a normative set that supports \( B \). We may restrict the factual set to worlds in which \( \neg A \) holds by processing the third premise, but we cannot then change anything by processing the fourth premise; restricting the normative set to worlds in which \( \neg B \) holds would yield the empty set of worlds. So, deontic normalization returns the original set of worlds in which \( B \) holds. Starting with the
third and fourth premises produces the opposite result, a normative information state supporting \( \neg B \). Neither \( OB \) nor \( O \neg B \), therefore, finds support in every fixed point.

Our theory, then, extends standard deontic logic, employing both deontic and factual detachment, without encountering contradiction. Notably, we do not need tense or even an irreducible conditional obligation operator to accomplish this. Our resolution of the paradox depends only on making detachment defeasible.

There are, of course, many other deontic paradoxes; we do not have space to consider them all. To discuss one more paradox: our embrace of closure under logical consequence leads us to validate Ross's paradox, (23)a:

\[
(23) \quad \begin{align*}
\text{a. } & \text{Mary should mail the letter.} & O A \\
& \text{Mary should mail the letter or burn it.} & O (A \lor B) \\
\text{b. } & O A \\
& \neg A \\
& OB^* 
\end{align*}
\]

The conclusion of (23)a suggests that, if Mary doesn't mail the letter, she should burn it. But this plainly does not follow: \( O (A \lor B) \) does not entail \( \neg A \rightarrow OB \), monotonically or nonmonotonically. Applying our theory to (23)b yields a factual set supporting \( \neg A \) and a normative set supporting \( A \) and, thereby, \( A \lor B \) but not \( B \) itself.

In conclusion, consider inferences (8) and (9) above:

\[
(8) \quad \begin{align*}
\text{a. } & \text{I should never do anything immoral.} \\
& \text{If I never do anything immoral, I should be revered.} \\
& \text{I should be revered.}^* \\
\text{b. } & \text{Those who remain alert should be rewarded} \\
& \text{Police officers should remain alert} \\
& \text{Pat is a police officer} \\
& \text{Pat should be rewarded}^* \\
\text{c. } & \text{Those who remain alert don't work the night shift} \\
& \text{Police officers should remain alert} \\
& \text{Pat is a police officer} \\
& \text{Pat shouldn't work the night shift}^* 
\end{align*}
\]

\[
(9) \quad \begin{align*}
\text{Midas is a person} \\
& \text{People should be generous} \\
& \text{Generous people should be thanked} \\
& \text{Midas should be thanked}^* 
\end{align*}
\]
All fail in our system. We consider only (8)a to show why. The reasoning is similar to that shown above for Ross's paradox. The first member of the normalization sequence will be a pair of states \((s, t)\) — \(s\) being the factual set, and \(t\) the normative set— where

\[ t \models \text{I never do anything immoral}, \]
while \(s\) supports the conditional second premise. We now normalize on the antecedent of this conditional, and since it is consistent with the premises that I never do anything immoral, we will get after normalization (where \(s^*\) represents the normalized state):

\[ s^* \models \text{I never do anything immoral} \rightarrow \text{I should be revered}. \]

But there is no way of using the information in \(t\), the normative set, to detach the conclusion of the material conditional true in \(s\), the factual set, on this or any subsequent normalization. The absurd conclusion of (8)a does not follow on our account.

On the other hand, a similar inference pattern, an example of which is given below, does follow:

(24) Those who remain alert should be rewarded
- Good police officers remain alert
  - Pat is a good police officer
  - Pat should be rewarded

In our theory, this inference holds. Updating with its third premise, we restrict the factual set to worlds in which Pat is a good police officer. The second premise restricts the factual set to one in which good police officers and, so, Pat remain alert. The first premise restricts the normative set to worlds in which those who remain alert in the factual set are rewarded. Thus, Pat is rewarded in every world in the normative set.

<table>
<thead>
<tr>
<th>Factual Set ((F))</th>
<th>Normative Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pat is a good police officer</td>
<td>All who remain-alert-in-(F) are rewarded</td>
</tr>
<tr>
<td>Good police officers remain alert</td>
<td>Pat is rewarded</td>
</tr>
<tr>
<td>Pat remains alert</td>
<td></td>
</tr>
</tbody>
</table>

4. Unconditional obligations

Unconditional prima facie obligations should be able to conflict, while unconditional actual obligations should not. We would go further. Say that a logic of unconditional obligation is *classical* iff it satisfies the following:
A logic of actual obligation should be classical, we maintain, while a logic of *prima facie* obligation should not be.

We satisfy criterion (V), Unconditional Actual Obligation, for seriality guarantees that $OA$ and $O \neg A$ are inconsistent. Our theory of $O$ is classical. To define a unary operator expressing *prima facie* obligation that fulfills criterion (VI), recall that (13) is only nonmonotonically valid:

(13) Neoptolemus is a person
People should be kind
Neoptolemus should be kind

This seems right if we read the conclusion as a statement of actual obligation. We might, however, read it as *prima facie*, taking the premises to imply monotonically that Neoptolemus has a *prima facie* obligation to be kind. Reading the conclusion as a *prima facie* ought, the inference appears monotonically valid.

We may understand this sort of unconditional *prima facie* obligation by defining it as an obligation following from an applicable rule. There is a *prima facie* obligation to $A$, then, if and only if there is a true $B$ such that $B >_O A$:

**Definition 4.1.** $O_{cp}A \iff$ There is a $B$ such that $B >_O A$, and $B$.

With our nonmonotonically weaker notion of conditional obligation, we can generate a similar concept of prima facie obligation based on moral generalizations:

**Definition 4.2.** $O_gA \iff$ There is a $B$ such that $B > O A$, and $B$,

and can identify *prima facie* obligation with one or the other of these concepts:

**Definition 4.3.** $O_{pf}A \iff (O_{cp}A \lor O_gA)$. 
As we might expect, *prima facie* obligation in this sense satisfies only some of the classical deontic axioms. Closure under logical consequence holds, because it holds for $>_O$ and $>$. And we have a *prima facie* obligation to the laws of logic. Agglomeration, in contrast, plainly fails; that there is a $B$ such that $B >_O A$ and a $D$ such that $D >_O C$, where both $B$ and $D$ hold, does nothing to show that there is a single $E$ such that $E >_O (A \& C)$. (The facts are similar for $>$ or a mixture of the two connectives.) The status of *ought-implies-can* depends on whether one adopts *generic seriality*: $(w,p) = \emptyset \rightarrow p = \emptyset$. Almost certainly one should not accept it; it is possible, after all, to promise the impossible. There is no such thing as a normal promise to do what cannot be done.

*Prima facie* obligations so defined can conflict. Our theory thus satisfies criterion (VI). This situation, for example, gives rise to moral conflict:

(25) Odysseus ordered Neoptolemus to trick Philoctetes.
    Doing so would be cruel.
    If Odysseus ordered Neoptolemus to trick Philoctetes, he should.
    If tricking Philoctetes would be cruel, Neoptolemus shouldn’t.
    Neoptolemus has a *prima facie* obligation to trick Philoctetes, and a
    *prima facie* obligation not to trick him.

When the conclusion has *prima facie* force, it follows monotonically from the premises. When the obligations alleged are actual, in contrast, the argument is a deontic Nixon diamond, like (2) above; nothing of interest follows, monotonically or nonmonotonically.

References

    of Nonmonotonic Reasoning’, *Proceedings of the 12th IJCAI*. Palo Alto: Morgan
    Kaufmann.


    sophical Studies* 50, 117–127.

    24, 33–36.

    1, 147–153.


Prima facie obligation


Nicholas Asher
Department of Philosophy
The University of Texas at Austin
Austin, Texas 78712 USA
nasher@bertie.la.utexas.edu

Daniel Bonevac
Department of Philosophy
The University of Texas at Austin
Austin, Texas 78712 USA
pudab@utxdp.dp.utexas.edu