

## *Against Conditional Obligation\**

DANIEL BONEVAC  
The University of Texas at Austin

Lennart Aqvist (1984) explains why most contemporary deontic logics use a primitive binary conditional obligation operator. He then points to an account of *prima facie* obligation as the primary outstanding problem facing these logics. Solving that problem, I hope to show, also solves the puzzles that motivated such theories in the first place.

The crucial feature of obligation sentences to which the puzzles point is that such sentences, and evaluative sentences more generally, are *defeasible*. They may be warranted, given some information, only to be defeated by further information. A theory that recognizes this no longer needs to see conditional obligation as anything more than a simple combination of unary obligation and the conditional.

My title may thus be overly provocative. I do not mean to deny that some obligations hold only if some condition is fulfilled. Nor do I mean to deny that sentences Kant would have called hypothetical imperatives, such as *If you like Mexican food, you should try Jorge's*, express such obligations. Finally, I do not mean to deny that there is anything wrong with the common strategy of using a primitive conditional obligation operator such as  $O(q/p)$  in a logical or philosophical theory of deontic concepts. It usefully avoids commitments about the relative roles of obligation and the conditional.

What I do want to attack is the idea that a theory with a primitive conditional obligation operator—a *binary* theory, as I shall call it—can be a *final* logical or philosophical theory about deontic concepts. At the very least, a theorist using a conditional obligation operator owes us an explanation of how the semantics of the operator depends on the semantics for obligation and the conditional *simpliciter*. Sentences expressing conditional obligations are intelligible to anyone understanding *should* (or *ought to*) and *if*. The combination of these words is no idiom. The meanings of such sentences, therefore, should be explicable in terms

of the meanings of *if* and *should* construed independently. In Richmond Thomason's words (1981a),

A proper theory of conditional obligation... will be the product of two separate components: a theory of the conditional and a theory of obligation.

### 1. Motivations for Binary Theories

The simplest way to achieve this is to analyze conditional obligation sentences as having the form  $p \rightarrow Oq$  or  $O(p \rightarrow q)$ , where  $\rightarrow$  is an appropriate conditional connective.<sup>1</sup> Initially, deontic logicians did just this. (See, for example, Mally 1926 and von Wright 1951.) Before long, however, von Wright (1956), Rescher (1958, 1962), and Chisholm (1964) began using primitive conditional obligation operators. Ever since, most current deontic theories have been binary.<sup>2</sup> Why?

According to the consensus view, only binary theories can explain some central properties of conditional obligation sentences.<sup>3</sup>

#### *Detachment*

Notoriously, this argument is not deductively valid:

(1) If you promised, then you should keep your promise	$p \rightarrow Oq$
You promised	$p$
You should keep your promise	$Oq$

(Throughout these examples, I use the most obvious unary symbolization.) Other circumstances and obligations may intervene. But this confronts unary accounts with a dilemma. One can avoid making (1) valid only by surrendering modus ponens. That threatens to cripple practical reasoning; one could never infer an unconditional obligation from a conditional one. Yet the only alternative seems to be accepting invalid forms of reasoning.

The premises of argument (1) do not guarantee the truth of its conclusion, but they do lend it some support. In particular, when other circumstances and obligations do not intervene, (1) appears to be an exemplar of practical reasoning. To account for this, a theory must allow arguments such as (1) to be successful without being deductively valid. In short, it must allow for defeasible success—a notion classical logic lacks.

#### *Puzzles of Derived Obligation*

Arthur Prior's (1954) paradoxes of derived obligation point out that a unary account forces conditional obligation to suffer the problems of the associated conditional. I will not rehearse Prior's version, but simply appeal to the resulting adequacy criteria for representations of conditional obligation, as spelled out in Aqvist (1984). None of the following should be valid:

- (2)  $\frac{\text{You didn't promise}}{\text{If you promised, you should die}} \quad \frac{\neg p}{p \rightarrow Oq}$
- (3)  $\frac{\text{You should keep your promise}}{\text{You should keep your promise if it kills you}} \quad \frac{Oq}{p \rightarrow Oq}$
- (4)  $\frac{\text{You shouldn't promise}}{\text{If you promise, you should die}} \quad \frac{O\neg p}{p \rightarrow Oq}$
- (5)  $\frac{\text{If you promised, you should keep your promise}}{\text{If you promised and died, you should keep your promise}} \quad \frac{p \rightarrow Oq}{(p \ \& \ r) \rightarrow Oq}$

The first two are instances of the paradoxes of material implication. The above symbolizations, with  $\rightarrow$  interpreted materially, make (2), (3), and (5) valid. Symbolizing the conditional obligations on the pattern of  $O(p \rightarrow q)$  instead validates (4) and (5).<sup>4</sup> Using a strict conditional can solve the problems posed by (2)–(4), but (1) and (5) remain. Using a counterfactual conditional solves (2)–(5), but (1) remains. What (1)–(5) require is a variable conditional that, in David Lewis’s (1973) terms, is not centered.

*The Robber, Gentle Murder, and Knower Puzzles*

The robber and victim paradoxes lead from seemingly innocuous premises to troubling conclusions:

- (6)  $\frac{\text{John is a robber.} \quad r}{\text{If John is a robber, he should repent.} \quad r \rightarrow Op}$   
 $\frac{\text{One can repent only if one has done wrong.} \quad p \rightarrow q}{\text{John should have done wrong.} \quad Oq}$

I use  $\rightarrow$  for strict implication; the paradoxes rely on the logical connection between repenting and doing wrong. The more recent gentle murder paradox (Forrester 1984, Castañeda 1985, 1986, Sinnott-Armstrong 1985, Belzer and Loewer 1986, Goble 1990a, b, c) has a similar form:

- (7)  $\frac{\text{If you murder Jones, you should murder him gently.} \quad q \rightarrow Op}{\text{You murder Jones.} \quad q}$   
 $\frac{\text{If you murder Jones gently, you murder him.} \quad p \rightarrow q}{\text{You should murder Jones.} \quad Oq}$

The gentle murder paradox is just the special case of the robber paradox in which  $r = q$ .

Also similar in form is Aqvist’s knower puzzle, discussed in von Wright (1983, 155–156):

(8) Bob is unreliable.	$p$
If so, you should have known him to be unreliable.	$p \rightarrow \text{OK}p$
You can know that Bob is unreliable only if he	
is unreliable.	$\text{K}p \rightarrow p$
Bob ought to be unreliable.	$\text{O}p$

This is simply the robber paradox, with  $r = p$ ,  $q = p$ , and  $p = \text{K}p$ .

These paradoxes challenge a common feature of traditional deontic logics: closure under logical consequence. Many writers have consequently rejected closure. In a nonmonotonic logic, as I shall argue, that is not the only option.

### *Chisholm's Contrary-to-Duty Puzzle*

The final and most-cited obstacle to unary theories is Roderick Chisholm's (1963) contrary-to-duty paradox:

(9) Ann ought to visit her grandmother.	$\text{O}p$
It ought to be the case that, if she visits, she calls.	$\text{O}(p \rightarrow q)$
Ann doesn't visit her grandmother.	$\neg p$
If she doesn't visit, she shouldn't call.	$\neg p \rightarrow \text{O}\neg q$

The English sentences in (9) seem to be consistent and mutually independent. In standard deontic logic, however, these assertions are inconsistent. The first two, in an instance of *deontic* detachment, imply  $\text{O}q$ , while the last two, in an instance of *factual* detachment, imply  $\text{O}\neg q$  (Greenspan 1975, Loewer and Belzer 1983).<sup>5</sup> Yet the English sentences in (9) are not only consistent but commonplace. Many have drawn the moral that one must choose between these two modes of detachment.

This would be unfortunate, however, for both are important to moral reasoning. Factual detachment is obviously important in allowing us to draw practical conclusions from hypothetical imperatives, as I observed earlier. Without it, it is hard to see how conditional obligations could have any force in practical reasoning. Deontic detachment, while arguably less central, is still important: it allows us to reason about the combined force of obligations. It underlies, for example, the force of

(10) Ann ought to visit her grandmother or call her.	$\text{O}(p \vee q)$
It ought to be the case that, if Ann visits, she calls.	$\text{O}(p \rightarrow q)$
Ann ought to call her grandmother.	$\text{O}q$

Admittedly, sometimes obligations seem to conflict. In such cases, their combined force is not clear. When there is no conflict, however, we need some way of computing their combined force. Deontic detachment provides that.

Fortunately, the choice between these two modes of detachment is unnecessary in a nonmonotonic system. One can have both deontic and factual detachment, so

long as both are defeasible. They are defeated, moreover, just when the above argument leads us to expect their defeat—namely, when obligations conflict.

## 2. Commonsense Entailment

I hope to show that a unary theory adequate for *prima facie* obligation, in which the underlying logic and corresponding account of the conditional are nonmonotonic, can explain the properties of conditional obligation sentences that have motivated binary views. John Horty (1993, 1994) and Asher and Bonevac (1996, 1997) have called for a nonmonotonic approach to deontic logic. But they have failed to notice that, once the notion of entailment is defeasible—allowing for conclusions to be overridden by further information—there is no longer any reason to take a conditional obligation operator as primitive. Detachment failures and deontic paradoxes become easily explicable, even when the only primitive deontic notion is unary, and even without the temporal machinery of, for example, Sellars (1967), Greenspan (1975), van Eck (1981), Thomason (1981a, b), or Belzer and Loewer (1983, 1986). The nonmonotonic system I adopt here is Asher and Morreau's (1995) *commonsense entailment*, supplemented with deontic and modal operators.

Classical deductive logic is *monotonic* in the sense that conclusions, once established, *stay* established. Adding premises to a valid argument always produces another valid argument. (That is, if  $\Gamma \vdash \varphi$ ,  $\Gamma \cup \Sigma \vdash \varphi$ .) In nonmonotonic logic, however, adding a premise may make a valid argument invalid. (Where  $\approx$  symbolizes nonmonotonic entailment,  $\Gamma \approx \varphi$  does not imply  $\Gamma \cup \Sigma \approx \varphi$ .) Conclusions drawn in a nonmonotonic system are defeasible. They may have to be surrendered in light of further information.

The key ideas of nonmonotonic logic, as I shall develop it, are those of (a) assuming only the information available in the premises of an argument, and (b) assuming that the world is otherwise as normal as possible. Ordinarily, one assumes, for purposes of evaluating an argument, that the premises are true. In a nonmonotonic logic, one assumes further that the premises exhaust the relevant information and, in particular, the relevant abnormalities. This idea is vague, and may be developed precisely in various ways. But it already has an important consequence: that nonmonotonic entailment is superclassical, that is, that  $\Gamma \approx \varphi$  whenever  $\Gamma \vdash \varphi$ .

To make use of the additional power of nonmonotonic entailment, one needs a rather weak conditional which, classically, enters into few deductive relationships. As I argued above, what one needs for deontic logic, at least, is a variable, decentered conditional. I follow Asher and Morreau (1995) in symbolizing this *generic* conditional (so-called because it is designed to capture English generics such as *Birds fly*) with  $>$  and specifying the truth condition:

$p > q$  is true at a world  $w$  iff  $q$  holds in all  $p$ -normal worlds relative to  $w$ .

This is weaker than a Lewis counterfactual, for  $w$  need not be a  $p$ -normal world relative to itself. (The  $p$ -normal worlds relative to  $w$  thus must not be construed as the closest  $p$ -worlds to  $w$ .) Asher and Morreau do require that:

- (a)  $p$  is true in all  $p$ -normal worlds; and
- (b)  $(p \vee q)$ -normal worlds are either  $p$ -normal or  $q$ -normal.

The first constraint guarantees the truth of  $p > p$ ; the second, of  $((p > r) \& (q > r)) \rightarrow ((p \vee q) > r)$ .

Together, these constraints imply that more specific conditionals take precedence over less specific conditionals. They validate an argument similar to that commonly known as the Penguin Principle:

- |   |                   |
|---|-------------------|
| (11) Tweety is a bird.  | $q$               |
| Birds fly.  | $q > r$           |
| Penguins are birds.   | $p \rightarrow q$ |
| Penguins do not fly.  | $p > \neg r$      |
| Tweety is a penguin.  | $p$               |
| <hr style="width: 100%; border: 0.5px solid black; margin: 0;"/> Tweety does not fly. | $\neg r$          |

(The symbolization here should more properly contain quantifiers. Since the argument of this paper depends solely on sentential issues, however, I use a simpler, in effect instantiated, symbolization.) Applied to this case directly, the above constraints validate the following inference:

- (12) Penguins are birds.
- No normal penguins are normal birds.
- Birds are normally nonpenguins.

This is enough to give the result.<sup>6</sup>

The generic conditional  $p > q$  has few monotonic consequences. In particular, it does not support modus ponens:  $p \& (p > q) \not\models q$ . Its real power comes from the assumption that things are as normal as possible, that is, as normal as the premises allow them to be. The assumption justifies *defeasible modus ponens*:

$$p \& (p > q) \vDash q.$$

Thus, modus ponens on  $>$  is valid nonmonotonically, in the absence of contrary information. If, however, applying modus ponens would yield a conclusion contradicting information already available, the application fails:  $p \& \neg q \& (p > q)$  does not imply  $q$ . To simplify somewhat, we can supplement a set of premises containing  $p > q$  with the corresponding material conditional  $p \rightarrow q$  unless the supplementation produces inconsistency.

To treat the puzzles of section 1 above, I need the modal connectives  $O$  and  $\rightarrow$ , the latter easily definable in terms of the material conditional and  $\Box$ . To keep

matters as simple as possible, suppose that the background modal logic is S5, with all worlds accessible from all others, and that each world has associated with it a nonempty set of ideal worlds. Then, in the standard way,

$\Box p$  is true at world  $w$  iff  $p$  is true in all worlds.

$Op$  is true at world  $w$  iff  $p$  is true in all  $w$ 's ideal worlds.

Note that the logic of  $O$  is normal, respecting necessitation, distribution, and, thus, closure under logical consequence;  $Op$  and  $p \rightarrow q$  entail  $Oq$ . Also,  $O$  is classical in that true conflicts of obligation cannot arise;  $Op$  and  $O\neg p$  are contradictory.  $O$  is thus an actual obligation operator; it expresses what is obligatory all things considered.

The generic conditional  $p > Oq$  is ideally suited to representing *prima facie* obligation. It is true iff  $Oq$  holds in all  $p$ -normal worlds. It means, in other words, that if  $p$  is true, then  $q$  is normally obligatory. Given the premise that  $p$ , we can infer  $Oq$  in the absence of any contradictory information. Additional information, however, might force us to withhold that conclusion. In particular, we might face a situation of moral conflict in which  $p, p > Oq, r$ , and  $r > O\neg q$  all hold. In such a circumstance, we could infer neither  $Oq$  nor  $O\neg q$ . The system thus allows us to represent *prima facie* obligation sentences and circumstances of moral conflict while having no deontic operator but a classical, unary  $O$ .

More specifically, representing *prima facie* obligation sentences as generic conditionals in this system satisfies a number of desiderata:

(I) *Default Detachment*. Inferences such as (1) count as acceptable if no other moral considerations apply. They are not deductively valid, but they are legitimate default inferences; we may draw the conclusion if no other rules intervene.

(II) *Conditional Conflict*. In cases where conditional *prima facie* principles conflict, we can generally draw no conclusions:  $\{p > Or, q > O\neg r, p, q\}$  implies neither  $Or$  nor  $O\neg r$ . Consider Plato's puzzle: If you promised to return the knife, you should return it; if returning the knife will cause mayhem, you should not return it. You did promise, but returning the knife will cause mayhem. Should you return the knife? This is a substantive moral question; no obligation should follow as a matter of logic.

(III) *Deontic Specificity*. More specific *prima facie* obligations take precedence over less specific ones.  $\{p > Or, (p \& q) > O\neg r, p, q\}$  implies  $O\neg r$ . To return to Plato's puzzle: Add the premise that, if you promised to return the knife but doing so will cause mayhem, you should not return it. Then it follows that you should not return the knife.

(V) *Unconditional Actual Obligation*. Unconditional statements of actual obligation are expressible and cannot conflict. This pair is inconsistent:

- (13)  $a$  has an actual obligation to  $F$   
 $a$  has an actual obligation not to  $F$

You cannot be obliged, actually, all things considered, both to return the knife and not to return it.

(VI) *Unconditional Prima Facie Obligation*. It is possible to express unconditional *prima facie* obligations, which can conflict. This pair of statements is consistent:

- (14) *a* has a *prima facie* obligation to *F*  
*a* has a *prima facie* obligation not to *F*

In Plato's puzzle, for example, you have a *prima facie* obligation to return the knife and another not to return it. In general, *q* is *prima facie* obligatory if and only if there is a *p* such that  $p \& (p > Oq)$  (Chisholm 1964).

For the technical development of this conception of nonmonotonic consequence, I will refer the reader to Asher (1995), Asher and Morreau (1995), and Asher and Bonevac (1996, 1997). Here I will give only a simple but equivalent formulation that captures the central idea of assuming that things are as normal as possible by supplementing sets of premises containing generic conditionals with corresponding material conditionals unless that results in inconsistency.

Let  $\Gamma$  be a finite set of sentences—for simplicity, without nested deontic operators. To define the set  $NCon(\Gamma)$  of nonmonotonic consequences of  $\Gamma$ , begin by enumerating the members of  $Prop(\Gamma)$ , the set of antecedents of generic conditionals in formulas (or subformulas thereof) in  $\Gamma$ . Because of constraint (c) above, the only enumerations that matter are those that respect specificity—in which, in other words, *p* precedes *r* (and  $Op$  precedes  $Or$ ) whenever  $p \models r$  and  $r \not\models p$ . Then, for each enumeration  $\nu$ , upon reaching antecedent *p* at stage *i*, add material conditionals corresponding to generic conditionals with *p* as antecedent, if the result is consistent. If not, do nothing. Cycle through until a fixed point is reached. Then, repeat this process for nested generic conditionals. Iterate for generics, then nested generics, etc., until reaching a fixed point for the entire recursion. A sentence  $\varphi \in NCon(\Gamma)$  iff it is a monotonic consequence of the fixed point set for every enumeration.<sup>7</sup>

It is easy to see that defeasible modus ponens holds, but only defeasibly. To the set  $\{p, p > q\}$  we can add  $p \rightarrow q$  without contradiction, so  $\{p, p > q\} \models q$ . Adding  $p \rightarrow q$  to  $\{p, \neg q, p > q\}$ , however, would result in inconsistency, so  $\{p, \neg q, p > q\}$  does not imply *q*. Similarly, in a case of moral conflict, conclusions that would follow in the absence of conflict are suspended.  $\{p, p > Oq\} \models Oq$  and  $\{r, r > O\neg q\} \models O\neg q$ , but  $\{p, p > Oq, r, r > O\neg q\}$  nonmonotonically entails neither  $Oq$  nor  $O\neg q$ . Some enumerations yield  $Oq$ , but others yield  $O\neg q$ . Nevertheless, because enumerations that consider weaker antecedents before stronger ones violate constraints (a) and (b), deontic specificity holds:  $\{p > Or, (p \& q) > O\neg r, p, q\} \models O\neg r$ .

### 3. Puzzle Solutions

The nonmonotonic system sketched rather informally in the previous section solves the deontic puzzles that have motivated binary theories. The system was



designed, of course, to solve the detachment and derived obligation puzzles. But it is worth seeing in more detail how it does so.

*Detachment*

The detachment puzzle is that this argument is not deductively valid:

$$\begin{array}{r}
 (1) \text{ If you promised, then you should keep your promise} \quad p > Oq \\
 \hline
 \text{You promised} \quad p \\
 \hline
 \text{You should keep your promise} \quad Oq
 \end{array}$$

(Symbolizations now use the generic conditional for conditional obligation.) The dilemma was that, if we accept modus ponens, (1) is valid; if not, conditional obligation sentences are incapable of playing any role in moral or practical deliberation. In a nonmonotonic system, however, we escape the dilemma. Argument (1) is an instance of defeasible modus ponens. As such, it is valid nonmonotonically. In face of further information that implies  $\neg Oq$ , however, the conclusion would be withdrawn, for, although  $\{p, p > Oq\} \vDash Oq$ ,  $\{p, p > Oq, \neg Oq\}$  does not imply  $Oq$ .

*Puzzles of Derived Obligation*

These arguments should not be valid:

$$\begin{array}{r}
 (2) \text{ You didn't promise} \quad \neg p \\
 \hline
 \text{If you promised, you should die} \quad p > Oq \\
 \\
 (3) \text{ You should keep your promise} \quad Oq \\
 \hline
 \text{You should keep your promise if it kills you} \quad p > Oq \\
 \\
 (4) \text{ You shouldn't promise} \quad O\neg p \\
 \hline
 \text{If you promise, you should die} \quad p > Oq \\
 \\
 (5) \text{ If you promised, you should keep your promise} \quad p > Oq \\
 \hline
 \text{If you promised and died, you should keep} \\
 \text{your promise} \quad (p \ \& \ r) > Oq
 \end{array}$$

Indeed, on my account, they are not, monotonically or nonmonotonically, for  $>$  is a variable decentered conditional. Arguments (2) and (3) fail because there is no connection between the truth of  $\neg p$  or  $Oq$  in a world and what happens in  $p$ -normal worlds relative to that world. (4) fails because of a similar lack of connection between ideals and  $p$ -normal worlds. Finally, (5) fails because there is no connection between  $p$ -normality and  $(p \ \& \ q)$ -normality.

*The Robber, Gentle Murder, and Knower Puzzles*

Belzer and Loewer (1983, 1986) contend that the robber and gentle murder paradoxes revolve around issues of judgment and deliberation, rather than around

closure of logical consequence. Thomason (1981a, 1981b) 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 another 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, gentle murder, and knower paradoxes have similar forms.

(15) Robber	Gentle murder	Knower
$r$	$q$	$p$
$p \rightarrow q$	$p \rightarrow q$	$Kp \rightarrow p$
$r > Op$	$q > Op$	$p > OKp$
$Oq$	$Oq$	$Op$

The conclusions follow nonmonotonically. Adding the material conditionals  $r \rightarrow Op$  and  $q \rightarrow Op$ , respectively, produces no inconsistency. 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 *given that information*. These conclusions, then, are of no interest in the context of deliberation.

In fact, explicit recognition of them propels us from the context of deliberation to the context of judgment, in which we consider the moral status of the facts. Making this explicit requires adding a moral premise to each argument form:

(16) Robber	Gentle murder	Knower
$O\neg q$	$O\neg q$	$O\neg p$
$r$	$q$	$p$
$p \rightarrow q$	$p \rightarrow q$	$Kp \rightarrow p$
$r > Op$	$q > Op$	$p > OKp$
$Oq$	$Oq$	$Op$

Now, the conclusions do not follow. Adding the material conditionals would yield inconsistency. In the context of deliberation, then, the robber should repent, you should murder Jones gently, and you should have known that Bob was unreliable. The questions of whether the robber should have become a robber, whether you should murder Jones, and whether Bob should be unreliable do not arise. Raising

them switches us to the context of judgment. There, the robber should never have been in a position to repent; you should not murder Jones, gently or otherwise; and Bob should not be unreliable. 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 or whether or not Bob is unreliable. It does not involve a thought experiment of going back in time to a point before these matters were settled. It simply depends on recognizing that, although they are true, they should not be.

*Chisholm's Contrary-to-Duty Paradox*

The contrary-to-duty paradox seems to bring factual and deontic detachment into conflict:

- |   |                    |
|---|--------------------|
| (9) Ann ought to visit her grandmother.                 | $Op$               |
| It ought to be the case that, if she visits, she calls. | $O(p > q)$         |
| Ann doesn't visit her grandmother.                      | $\neg p$           |
| If she doesn't visit, she shouldn't call.               | $\neg p > O\neg q$ |

But my system has defeasible versions of both:  $\{Op, O(p > q)\} \approx Oq$ , and  $\{\neg p, \neg p > O\neg q\} \approx O\neg q$ . Nevertheless, the assertions in (9) are not inconsistent. What follows nonmonotonically from  $\{Op, O(p > q), \neg p, \neg p > O\neg q\}$ ? The answer,  $O\neg q$ , is just what our intuitions indicate—Ann should not call, since she is not coming. Factual detachment takes precedence over deontic detachment.

**4. Objections**

Horatio Arlo-Costa has raised objections to the theory I have presented here based on two additional puzzles. The first, Fred Feldman's (1984) paradox of the second-best plan, is specifically deontic. The second is the drowning problem (Benferhat, Cayrol, Dubois, Lang, and Prade 1993).

*The Second-best Plan*

Dr. Denton has two treatment options with respect to a certain patient, one better than the other. Each involves giving the patient medication on each of two days. Mixing the medications is dangerous, and would do serious harm. Suppose the doctor gives the second-best medication on the first day. Which should he give on day 2? Intuitively, the answer seems clear: he should give the second-best medication again. It would have been better to give the better medication on both days, but, now that the second-best treatment has begun, it is better to continue it. To switch to the better treatment plan would endanger the patient.

If these facts are symbolized as follows, my theory seems to yield the wrong result:

(17) Denton should give <i>X</i> today and again tomorrow.	$O(p \ \& \ q)$
Denton does not give <i>X</i> today.	$\neg p$
If he does not give <i>X</i> today, he should not give	
<i>X</i> tomorrow.	$\neg p > O\neg q$
Denton should give <i>X</i> tomorrow?	$Oq?$

But the problem here is with the symbolization. *O* is an actual obligation operator.  $O(p \ \& \ q)$  thus means that, all things considered, Denton should give *X* both today and tomorrow. That implies that, all things considered, Denton should give *X* tomorrow. In the case at hand, however, Denton should do no such thing. In general, medication *X* is better; he should give it both days. But, embarking on the second-best treatment plan overrides that general advice. A better representation of the situation, then, is this:

(18) Denton should give <i>X</i> today and again tomorrow.	$\top > O(p \ \& \ q)$
Denton does not give <i>X</i> today.	$\neg p$
If he does not give <i>X</i> today, he should not give	
<i>X</i> tomorrow.	$\neg p > O\neg q$
Denton should not give <i>X</i> tomorrow.	$O\neg q$

The general, rather than all-things-considered, aspect of the first premise I represent with a generic conditional with a vacuous antecedent. This is generally how categorical (i.e., unconditional) imperatives should be interpreted within the theory. Because anything contingent is more specific than logical truth, the conditional  $\neg p > O\neg q$  takes precedence and yields the desired conclusion  $O\neg q$ .

### *The Drowning Problem*

My system falls prey to the drowning problem, illustrated by the following example from Arlo-Costa:

(19) Ann buys a plane ticket.	$p$
If Ann buys a plane ticket, she should call.	$p > Oq$
If Ann buys a plane ticket, it is from TWA.	$p > t$
She does not buy the ticket from TWA.	$\neg t$

Intuitively, we would like to derive the conclusion that Ann should call; which airline she flies has no bearing on her obligation. But commonsense entailment considers all generic conditionals with the same antecedent at once. It adds corresponding material conditionals for all or for none. Consequently, the conflict over *t* drowns out the desired conclusion  $Oq$ .

There are ways of solving the drowning problem in commonsense entailment. The simplest is to enumerate in the recursion not antecedents of conditionals but conditionals themselves. This allows for supplementing the premises of (19) with  $p \rightarrow Oq$  but not  $p \rightarrow t$ , thus obtaining the conclusion  $Oq$  without obtaining a

contradiction by deriving  $t$ . A more complex but satisfying solution is given in Asher (1995), which takes into account degrees of independence.

I think there is a case to be made, nevertheless, for *not* solving the drowning problem. Compare (19) to the apparently analogous (20):

(20) Ann visits her grandmother.	$p$
If Ann visits, she should call.	$p > Oq$
If Ann visits, her grandmother spends days preparing.	$p > t$
Her grandmother cannot spend days preparing.	$\neg \diamond t$

Courtesy demands that Ann call, but a call, say, would endanger her grandmother’s health or at least upset her. (Imagine that grandma has just been released from the hospital.) Should Ann call? It is not clear. Evidently deriving the conclusion  $Oq$  here would be inappropriate. Hence Asher’s attempt to build degrees of independence into commonsense entailment.

I am not convinced, however, that drowning effects are a matter of logic at all. Consider some nondeontic examples. Say that Tweety is a bird and that birds fly. Suppose further that Tweety is abnormal in some respect. What should we conclude about Tweety’s capacities for flight? Can we expect logic to provide an answer?

(21) Tweety is a bird.	$p$
Birds fly.	$p > q$
Birds	$p > r$
	look blue, brown, gray, or red
	have claws
	eat seeds and insects
	nest in trees
	have wings
Tweety does not	$\neg r$
	look blue, brown, gray, or red
	have claws
	eat seeds and insects
	nest in trees
	have wings
Tweety flies?	$q?$

As one goes down the list of typical properties in (21), one is less and less inclined, I think, to draw the conclusion that Tweety flies. If Tweety is green, the conclusion still seems acceptable; if Tweety has no wings, it is unacceptable; if Tweety does not eat seeds and insects, or does not nest in trees, it is not clear what to say.

My system in this paper captures this situation. Given just the premises  $p, p > q, p > r$ , and  $\neg r$ , one cannot conclude  $q$ . But one could reach that conclusion with the additional premise  $(p \ \& \ \neg r) > q$ . What we do, faced with (21), is to evaluate the acceptability of added premises of that form. We ask ourselves, in other words:

- (22) Do birds that        are unusually colored        fly?  
                                  are clawless  
                                  do not eat seeds or insects  
                                  do not nest in trees  
                                  do not have wings

Most of us, with limited knowledge of birds, would answer the first affirmatively, the last negatively, and the others uncertainly. The key point is that, in doing this, we are assessing the acceptability of an added premise that would yield the conclusion in question. We are not evaluating (21) as it stands.

If this is right, then the inferences that give rise to the drowning problem are not what they seem. They are not valid, even defeasibly. They are elliptical. They depend on the acceptability of additional premises that do nonmonotonically imply the conclusion. We do not draw the inference that Ann should call her grandmother in (19) from the premises given; we assess the additional premise

- (23) If Ann has bought a plane ticket from an airline other than TWA, she should call her grandmother,

find it acceptable, and draw the conclusion from the supplemented set of premises. Similarly, in (20), we assess the further premise

- (24) If Ann has bought a plane ticket but her grandmother cannot spend days preparing, Ann should call her grandmother

and find it dubious. We thus decline to infer that Ann should call.

## 5. Conclusion

I do not claim that my system in this paper solves all deontic problems. There are interactions between tense and deontic modality, for example, that I do not address. Nor do I address degrees of value, judging one option better than another, or the rich range of deontic notions connected to these ideas. (Paul McNamara's (1996a, b) work on supererogation and related notions, however, could fit comfortably into this framework.) Finally, I have said little about how one actually goes about assessing generic conditionals, or how they and the account of non-monotonic inference commonsense entailment provides relate to probabilistic inference.

I do hope to have shown, however, that, if our underlying logic allows for defeasible conclusions, it gives us not only a theory of defeasible or *prima facie* obligation but also a deontic logic that solves the puzzles motivating binary accounts. That logic employs nothing but a unary and entirely classical obligation operator. If that is right, deontic logicians have spent a great deal of time worrying about the peculiarities of specifically deontic notions when they should have been worrying about more general features of the conditional and entailment in general. The *deontic* part of deontic logic is easy; the *logic* part is hard.

## Notes

\* I presented earlier versions of this paper at a meeting of the Society for Exact Philosophy in October 1996 and at the Eastern Division meetings of the American Philosophical Association in December 1996. I am grateful to my audiences, especially to Michael Dunn, John Horty, Mark Lance, Isaac Levi, Bernard Linsky, Paul McNamara, Charles Morgan, Micheal Morreau, Donald Nute, Rohit Parikh, and Jeff Pelletier, for their comments and criticisms. I owe a special debt of thanks to Horatio Arlo-Costa, who commented on this paper at the APA and clarified my thinking on many points, and to Nicholas Asher, with whom I have worked out the theory that lies behind the argument of this paper.

1. One could accomplish the same goal by using a conditional obligation operator  $O(q/p)$  and then explaining how the semantics of the operator is a function of the semantics of the conditional  $\rightarrow$  and the unary  $O$ . One might contend, for example, that the conditional has a general meaning that becomes deontic, modal, or tensed when an appropriate auxiliary appears in the consequent, and otherwise takes a default value. This would be equivalent to treating conditional obligation sentences as having the form  $p \rightarrow_0 q$ . The conditional obligation operator  $\rightarrow_0$  in such a theory, however, would not be primitive, but a special case of a more general conditional operator  $\rightarrow_x$ .

2. Binary theories have since been advanced by Powers (1967), Danielsson (1968), Hansson (1969), Segerberg (1971), Follesdal and Hilpinen (1971), van Fraassen (1972), Lewis (1973, 1974), Chellas (1974, 1980), Thomason (1981a, b), Belzer and Loewer (1983, 1986), Aqvist (1984), Goble (1990a, b, c), Horty (1993, 1994), and many others.

3. Sometimes, elegance offers the chief motivation (Chisholm 1964); sometimes, a semantic insight or parallel with other constructions (Lewis 1973). I shall have nothing to say about these motivations in this paper.

4. It is worth noting that von Wright (1956) takes Prior's puzzle, all by itself, as implying the inadequacy of unary deontic theories:

I think that the proper conclusion to be drawn from Prior's objection is that  $O(A \rightarrow B)$  is *not* (contrary to my earlier opinion) an adequate expression in symbolic terms of the notion of commitment (or derived obligation). My belief is, moreover, that a formalization of this notion *cannot* be accomplished at all within the system developed in my paper. (509)

He immediately introduces a binary permission operator to solve the problem. If my argument here is correct, von Wright overreacts; the problem is not with the unary representation of deontic concepts but with the assumed theory of the conditional.

5. Symbolizing the conditional obligation sentences differently feels strange, even if suggested by the English phrasing I have used above. But it is necessary if the formulas are to be mutually independent. If we change the symbolization of the second sentence to  $p \rightarrow Oq$ , it follows from the third. If we change the symbolization of the last sentence to  $O(\neg p \rightarrow \neg q)$ , it follows from the first.

6. To see how the constraints validate (12): Assume that penguins are birds and that no normal penguins are normal birds. The set of penguins is the set of penguins that are birds, so the set of birds is the union of the set of penguins and the set of birds that are not penguins. By (b), then, the set of normal birds is a subset of the union of the set of normal penguins and the set of normal birds that are not penguins. Since no normal penguins are normal birds, the set of normal birds is a subset of the set of normal nonpenguin birds, which, by (a), is a subset of the set of nonpenguin birds, which is a subset of the set of nonpenguins. So, birds are normally nonpenguins.

To see how this validates (11), assume that the premises hold in  $w$ . Then, in particular,  $p$  and  $q$  hold in  $w$ . In all  $q$ -normal worlds relative to  $w$ ,  $r$  holds; in all such  $p$ -normal worlds,  $\neg r$  does. Thus, the  $q$ -normal and  $p$ -normal worlds are disjoint. By (12), then,  $\neg p$  holds in all  $q$ -normal worlds relative to  $w$ . Since  $p$  holds in  $w$ , however,  $w$  is not  $q$ -normal. Assuming as much normality as possible, then, we can assume  $w$  to be  $p$ -normal, but not  $q$ -normal. That is, we can assume that Tweety is a normal penguin, but not that Tweety is a normal bird. We thus conclude that Tweety does not fly.

The Penguin Principle itself is (11), but with the strict implication  $p \rightarrow q$  replaced with a generic  $p > q$ . That is not valid in the Asher and Morreau system.

7. More formally:

- (a)  $\Gamma_{1,0^\nu} = \Gamma$
- (b) If  $j$  is odd, then  $\Gamma_{j,i+1^\nu} = \Gamma_{j,i^\nu} \cup \{p \rightarrow q: \nu(p) = i + 1 \ \& \ p > q \in \text{Con}(\Gamma_{j,i^\nu})\}$ , if the result is consistent; if not,  $\Gamma_{j,i+1^\nu} = \Gamma_{j,i^\nu}$ .
- (c) If  $j$  is even,  $\Gamma_{j,i+1^\nu} = \Gamma_{j,i^\nu} \cup \{O(p \rightarrow q): \nu(p) = i + 1 \ \& \ O(p > q) \in \text{Con}(\Gamma_{j,i^\nu})\}$ , if the result is consistent; if not,  $\Gamma_{j,i+1^\nu} = \Gamma_{j,i^\nu}$ .
- (d) At limit ordinals  $\lambda$ ,  $\Gamma_{j,\lambda^\nu} = \bigcup_{i < \lambda} \Gamma_{j,i^\nu}$ .
- (e)  $\Gamma_{j+1,0^\nu} = \Gamma_{j,\mu^\nu}$ , where  $\mu$  is a fixed point.
- (f) At limit ordinals  $\lambda$ ,  $\Gamma_{\lambda,i^\nu} = \bigcup_{j < \lambda} \Gamma_{j,i^\nu}$ .

Eventually this entire process reaches a fixed point  $\Gamma^{*\nu}$ . A sentence  $\varphi \in \text{NCon}(\Gamma)$  iff  $\Gamma^{*\nu} \models \varphi$  for each enumeration  $\nu$ .

## References

- Aqvist, L.: 1984, "Deontic Logic," in D. Gabbay and F. Guentner (ed.), *Handbook of Philosophical Logic, Volume II: Extensions of Classical Logic* (Dordrecht: D. Reidel): 605–714.
- Asher, N.: 1995, "Commonsense Entailment: A Conditional Logic for Some Generics," in G. Crocco, L. Fariñas del Cerro, and A. Herzog (ed.), *Conditionals: from Philosophy to Computer Science*. Oxford: Clarendon Press, 103–146.
- Asher, N., and Bonevac, D.: 1996, "Prima Facie Obligation," *Studia Logica* 57 (1996): 19–45.
- Asher, N., and Bonevac, D.: 1997, "Common Sense Obligation," in D. Nute (ed.), *Defeasible Deontic Logic* (Dordrecht: Kluwer).
- Asher, N., and Morreau, M.: 1995, "What Some Generic Sentences Mean," in G. Carson and J. Pelletier (ed.), *The Generic Book* (Chicago: University of Chicago Press): 300–338.
- Belzer, M., and Loewer, B.: 1983, "Dyadic Deontic Detachment," *Synthese* 54, 295–318.
- Belzer, M., and Loewer, B.: 1986, "Help for the Good Samaritan Paradox," *Philosophical Studies* 50, 117–127.
- Benferhat, S., Cayrol, C., Dubois, D., Lang, J., and Prade, H.: 1993, "Inconsistency Management and Prioritized Syntax-based Entailment," *IJCAI 93*, Chambéry.
- Castañeda, H. N.: 1986, "Obligations, Aspectual Actions, and Circumstances," *Philosophical Papers* 15, 155–170.
- Chellas, B.: 1974, "Conditional Obligation," in S. Stenlund (ed.), *Logical Theory and Semantic Analysis*. Dordrecht: D. Reidel, 23–33.
- Chellas, B.: 1980, *Modal Logic*. Cambridge: Cambridge University Press.
- Chisholm, R.: 1963, "Contrary-to-duty Imperatives and Deontic Logic," *Analysis* 24: 33–36.
- Chisholm, R.: 1964, "The Ethics of Requirement," *American Philosophical Quarterly* 1, 147–153.
- Danielsson, S.: 1968, *Preference and Obligation: Studies in the Logic of Ethics* (Uppsala: Filosofiska föreningen).
- Feldman, F.: 1984, *Doing the Best We Can*. Dordrecht: D. Reidel.
- Follesdal, D., and Hilpinen, R.: 1971, "Deontic Logic: An Introduction," in R. Hilpinen (ed.), *Deontic Logic: Introductory and Systematic Readings* (Dordrecht: D. Reidel): 1–35.
- Forrester, J.: 1984, "Gentle Murder, or the Adverbial Samaritan," *The Journal of Philosophy* 81, 193–197.
- Goble, L.: 1990, "A Logic of Good, Should, and Would, Part I," *Journal of Philosophical Logic* 19, 169–200.
- Goble, L.: 1990, "A Logic of Good, Should, and Would, Part II," *Journal of Philosophical Logic* 19, 253–276.
- Goble, L.: 1990, "Murder Most Gentle: The Paradox Deepens," *Philosophical Studies* 64, 217–227.
- Greenspan, P.: 1975, "Conditional Ought and Hypothetical Imperatives," *Journal of Philosophy* 72, 259–276.
- Hansson, B.: 1969, "An Analysis of Some Deontic Logics," *Nous* 3: 373–398.



- Horty, J.: 1993, "Deontic Logic as Founded on Nonmonotonic Logic," *Annals of Mathematics and Artificial Intelligence* 9, 69–91.
- Horty, J.: 1994, "Moral Dilemmas and Nonmonotonic Logic," *Journal of Philosophical Logic* 23, 35–65.
- Lewis, D.: 1973, *Counterfactuals*. Cambridge: Harvard University Press.
- Lewis, D.: 1974, "Semantical Analysis for Dyadic Deontic Logic," in S. Stenlund (ed.), *Logical Theory and Semantic Analysis*. Dordrecht: D. Reidel, 1–14.
- Mally, E.: 1926, *Grundgesetze des Sollens: Elemente der Logik des Willens* (Graz: Lueschner & Lubensky).
- McNamara, P.: 1996, "Doing Well Enough: Toward a Logic for Common-Sense Morality," *Studia Logica* 57, 167–192 .
- McNamara, P.: 1996, "Making Room for Going Beyond the Call," *Mind* 105, 415–450 .
- Powers, L.: 1967, "Some Deontic Logicians," *Nous* 1, 381–400.
- Prior, A.: 1954, "The Paradoxes of Derived Obligation," *Mind* 63, 64–65.
- Rescher, N.: 1958, "An Axiom System for Deontic Logic," *Philosophical Studies* 9, 24–30.
- Rescher, N.: 1962, "Conditional Permission in Deontic Logic," *Philosophical Studies* 13, 1–6.
- Seegerberg, K.: 1971, "Some Logics of Commitment and Obligation," in R. Hilpinen (ed.), *Deontic Logic: Introductory and Systematic Readings* (Dordrecht: D. Reidel): 148–158.
- Sellars, W.: 1967, "Some Reflections on Contrary-to-Duty Imperatives," *Nous* 1, 303–344.
- Sinnott-Armstrong, W.: 1985, "A Solution to Forrester's Paradox of Gentle Murder," *Journal of Philosophy* 82.
- Thomason, R.: 1981, "Deontic Logic as Founded on Tense Logic," in R. Hilpinen (ed.), *New Studies in Deontic Logic*. Dordrecht: D. Reidel.
- Thomason, R.: 1981, "Deontic Logic and the Role of Freedom in Moral Deliberation," in R. Hilpinen (ed.), *New Studies in Deontic Logic*. Dordrecht: D. Reidel.
- van Eck, J. E.: 1981, *A System of Temporally Relative Modal and Deontic Predicate Logic and its Philosophical Applications* (Gröningen: University of Gröningen).
- van Fraassen, B. C.: 1972, "The Logic of Conditional Obligation," *Journal of Philosophical Logic* 1, 417–438.
- van Fraassen, B. C.: 1973, "Values and the Heart's Command," *Journal of Philosophy* 70, 5–19.
- von Wright, G. H.: 1951, "Deontic Logic," *Mind* 60, 1–15.
- von Wright, G. H.: 1956, "A Note on Deontic Logic and Derived Obligation," *Mind* 65: 507–509.
- von Wright, G. H.: 1983, "Norms, Truth, and Logic," in *Practical Reason* (Ithaca: Cornell University Press).