THE JOURNAL OF PHILOSOPHY
VOLUME CVII, NO. 3, MARCH 2010

IFS AND OUGHTS*

Ten miners are trapped either in shaft *A* or in shaft *B*, but we do not know which. Flood waters threaten to flood the shafts. We have enough sandbags to block one shaft, but not both. If we block one shaft, all the water will go into the other shaft, killing any miners inside it. If we block neither shaft, both shafts will fill halfway with water, and just one miner, the lowest in the shaft, will be killed.[1]

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

We take it as obvious that the outcome of our deliberation should be

(1) We ought to block neither shaft.[2]

* This paper and its companion paper, "Ought: Between Objective and Subjective," emerged out of discussions that began in NK's Spring 2007 graduate seminar and continued in NK's and Jay Wallace's Spring 2008 seminar and JM's Spring 2008 seminar. We would like to thank participants in all of these seminars for their feedback. We are also grateful for the feedback we received from colleagues at the Workshop on Context-dependence, Perspective and Relativity in Language and Thought at the Institut Jean Nicod in Paris, the Arizona Ontology Conference, the University of Nevada at Las Vegas, the University of Minnesota, the University of North Carolina at Chapel Hill, the University of Toronto, the University of Pittsburgh, Texas Tech University, Williams College, the New York Institute of Philosophy Disagreement Workshop, Berkeley's Group in Logic and the Methodology of Science, the Workshop on Contextualism and Relativism at the Arché Center in St. Andrews, and Krister Bykvist and Jonas Olson's Trinity 2008 seminar at Oxford.

[1] We take the example from Derek Parfit ("What We Together Do," unpublished) who credits Donald Regan, *Utilitarianism and Cooperation* (New York: Oxford, 1980), p. 265n1.

[2] We take this conclusion to be largely independent of one's background moral views. Although it is obviously ratified by consequentialist norms, which advise us to act so as to maximize expected utility, it seems to us that most reasonable deontological and virtue theories will also ratify it. We acknowledge, however, that there may be some extreme moral views that would reject it.

0022-362X/10/0703/115–143

Still, in deliberating about what to do, it seems natural to accept:

(2) If the miners are in shaft *A*, we ought to block shaft *A*.
(3) If the miners are in shaft *B*, we ought to block shaft *B*.

We also accept:

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

But (2), (3), and (4) seem to entail

(5) Either we ought to block shaft *A* or we ought to block shaft *B*.

And this is incompatible with (1). So we have a paradox.[3]

A paradox demands a solution. Here are the ones that most obviously come to mind:

I. Reject one or more of the premises.

(a) Reject (1).
(b) Reject (2) or (3).

II. Distinguish objective and subjective senses of 'ought' (or take 'oughts' to be context sensitive), so that (1) and (5) are compatible.
III. Take the argument to be invalid by taking it to have a nonobvious logical form:

(a) Take 'ought' in (2) and (3) to have wide scope over the conditional.
(b) Analyze (2) and (3) using a dyadic conditional obligation operator.

All of these are represented somewhere in the literature on 'ought's and conditionals. We will argue that none of them work. The best way to resolve the paradox, we will argue, is to give a semantics for deontic modals and indicative conditionals that lets us see how the argument can be invalid even with its obvious logical form. This requires rejecting the general validity of at least one classical deduction rule:

IV. Take the argument to be invalid even with its obvious logical form.

(a) Reject disjunction introduction.
(b) Reject disjunction elimination.
(c) Reject modus ponens for the indicative conditional.

[3] This is not the first paradox involving conditional obligation to have been discussed by philosophers. There is a healthy literature on other paradoxes of conditional obligation, such as the gentle murder paradox and other paradoxes involving "contrary to duty obligations." We think, though, that this paradox raises issues that are not raised by the others, and avoids other issues that they raise. James Dreier presents a similar paradox involving 'better' rather than 'ought' in "Practical Conditionals," in David Sobel and Steven Wall, eds., *Reasons for Action* (New York: Cambridge, 2009), pp. 116–33. He too surmises, as we go on to claim, that modus ponens must be invalid for the relevant conditionals.

We plump for IV(c). At first glance this might seem like no solution at all—a bit like killing the baby to save the bathwater. We will argue, to the contrary, that there are good reasons, independent of 'ought', for rejecting modus ponens for the indicative conditional. And we will show that rejecting modus ponens is not as revisionary as it sounds, because most ordinary reasoning using modus ponens can be vindicated.

I. REJECTING A PREMISE

Those who want to solve the paradox by rejecting (or at least refusing to accept) a premise have two options. They can reject (1), or they can reject the two conditionals (2) and (3). We will consider these options in turn.

I.1. Rejecting (1): Objectivism. One clear motivation for rejecting (1) would be the position we call

> *Objectivism*
>
> *S* ought to φ iff φ-ing is the best choice available to *S* in light of all the facts, known and unknown.

According to objectivism, (1) is false, since in light of *all* the facts, the best course of action is to block whichever shaft the miners are in. (As a heuristic for the objective 'ought', consider what an omniscient being would advise us to do.)

The obvious worry about objectivism is that, in deciding what we ought to do, we always have limited information, and are in no position to determine what is the best course of action in light of *all* the facts. Thus the objectivist's 'ought' seems useless in deliberation.[4]

Objectivists reply by noting that we may be justified in judging or asserting that we ought to φ, despite our limited knowledge, provided it is probable on our evidence that φ-ing is the best course of action in light of all the facts. As Moore puts it, "we may be justified in saying

[4] For this general line of objection, see W. D. Ross, *Foundations of Ethics* (Oxford: Clarendon, 1939), pp. 146–67; A. C. Ewing, *The Definition of Good* (New York: MacMillan, 1947), p. 128; H. A. Prichard, "Duty and Ignorance of Fact," in Ross, ed., *Moral Obligation* (Oxford: Clarendon, 1949), pp. 18–39; Frank Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," *Ethics*, CI (1991): 461–82, at pp. 466–67; John Broome, *Weighing Goods* (Cambridge: Blackwell, 1991), p. 128; T. M. Scanlon, "Thomson on Self-Defense," in Alex Byrne, Robert Stalnaker, and Ralph Wedgwood, eds., *Fact and Value: Essays on Ethics and Metaphysics for Judith Jarvis Thomson* (Cambridge: MIT, 2001), pp. 199–214; Scanlon, *Moral Dimensions: Permissibility, Meaning, and Blame* (Cambridge: Harvard, 2008); Parfit, *Reasons and Persons* (New York: Oxford, 1984), p. 25, "What We Together Do" and "On What Matters," unpublished; Ralph Wedgwood, "Choosing Rationally and Choosing Correctly," in Sarah Stroud and Christine Toppolet, eds., *Weakness of Will and Practical Irrationality* (New York: Oxford, 2003), pp. 201–230, at p. 204; and Allan Gibbard, "Truth and Correct Belief," *Philosophical Issues*, XV (2005): 338–50, at pp. 343–44.

many things, which we do not know to be true, and which are in fact not so, provided there is a strong probability that they are."[5] But this reply cannot help us with the miners case, since we know with certainty that leaving both shafts open is *not* the best course of action in light of all the facts. Moore's gambit is dubious anyway. One would *not* be justified in saying that one ought to speed through a blind intersection on a country road, even though the probability is very high that there is no car coming, and hence that what one ought to do in light of all the facts is speed through the intersection. Allan Gibbard diagnoses the problem well: "from objective *oughts* we can glean only an ordinal utility scale for the sure alternatives. What one ought to do subjectively depends not only on this, but on the cardinal utilities involved."[6] If the hazard were a mud puddle rather than a vehicular collision, what one ought to do would be different, even if the likelihoods of the outcomes and the ranking of them from best to worst were the same. Clearly the objective 'ought' is not the 'ought' that matters when we are deliberating about what to do.

I.2. Rejecting (2) and (3): Subjectivism. An appreciation of the problems with objectivism might incline one to accept

Subjectivism

S ought (at t) to φ iff φ-ing is the best choice available to S in light of what S knows at t.

Subjectivism, in conjunction with an account of the indicative conditional that licenses modus ponens, implies that at least one of (2) and (3) is false, since it has a true antecedent and a false consequent. Subjectivism would validate only the weaker pair of conditionals:

(6) If we know that the miners are in shaft A, we ought to block shaft A.
(7) If we know that the miners are in shaft B, we ought to block shaft B.

It seems to us that the loss of (2) and (3) is already a significant cost. These conditionals naturally occur to one in the course of deliberation, and they seem perfectly acceptable—until one starts thinking about the paradoxes. It would be preferable, we think, to have an account of 'ought' that allowed these conditionals to be true, on some construal. By offering such an account, we hope to undercut any motivation for retreating from (2) and (3) to (6) and (7).

[5] *Ethics* (Oxford: UP, 1912), pp. 100–01. See also Judith Jarvis Thomson, "Imposing Risks," in William Parent, ed., *Rights, Restitution, and Risk* (Cambridge: Harvard, 1986), pp. 173–91, at p. 178; Richard B. Brandt, *Ethical Theory* (Englewood Cliffs, NJ: Prentice Hall, 1959), p. 367.

[6] Gibbard, "Truth and Correct Belief," p. 345.

In addition, we think that there are strong independent reasons for rejecting subjectivism. Although subjectivism seems well suited to make sense of the use of 'ought' in deliberation, it cannot make good sense of the use of 'ought' in *advice*. Suppose the deliberator in the miners case is confronted by an adviser who knows where the miners are:

Dialogue 1

Agent: I ought to leave both shafts open, guaranteeing that nine survive.

Adviser: No, you ought to block shaft *A*. Doing so will save all ten of the miners.

If we suppose, with the subjectivists, that Agent is making a claim about the best choice available to her in light of her evidence at that time, we can make good sense of her assertion. But then how do we understand Adviser's reply? On the subjectivist construal, Adviser is making a claim about the best choice available to Agent in light of her (Agent's) evidence. But that is pretty clearly *not* what he is doing. Indeed, he presumably knows that Agent has already got the right answer to *that* question. As Judith Thomson puts the point:

> On those rare occasions on which someone conceives the idea of asking for my advice on a moral matter, I do not take my field work to be limited to a study of what he believes is the case: I take it to be incumbent on me to find out what is the case.[7]

A subjectivist might be tempted to respond that just by hearing Adviser's reply, Agent acquires evidence that the miners are in shaft *A*—so that Adviser's claim becomes true, on the subjectivist construal, partly as a result of its being made.[8] But this response is inadequate in two ways. First, it does not capture the sense in which Adviser is *disagreeing* with Agent ("No, ..."). For on this interpretation Adviser's and Agent's claims would be compatible claims about what the Agent ought to do at different times, or relative to different bodies of evidence. Second, it will work only when conditions are right for the testimonial transfer of knowledge. In a case where Agent has good reason to think Adviser is ill informed or malevolently disposed, Agent will *not* acquire knowledge of the miners' location from Adviser's assertion. Agent might even take Adviser's assertion to support the

[7] Thomson, *op. cit.*, p. 179.

[8] See Ross, *Foundations of Ethics*, pp. 152–53, paraphrasing Prichard in "Duty and Ignorance of Fact."

view that the miners *are not* in Shaft *A*. Subjectivists will have to concede that if Adviser knows that Agent has these doubts, both of them will know that his advice is false.

Thus the subjectivist is committed to explaining why Adviser should give advice she knows to be false. Granted, an adviser might have good reason to get an agent to have a false belief. In the case under discussion, the reason might be that if Agent acted on the false belief, it would lead to all ten miners' being saved. But in a case where *both* Agent and Adviser know that the advice is false, Adviser will have no reason to suppose that Agent will believe what he says. So we are left with no reason for Adviser to give the response she does.

More fundamentally, this sort of strategic consideration cannot possibly explain why Adviser would not only say, but *believe* that Agent ought to block shaft *A*. But surely Adviser would be quite rational to believe this, given what he knows.

II. MAKING (1) AND (5) COMPATIBLE

In light of our discussions of objectivism and subjectivism, it is tempting to think that there is something right about both views. Perhaps each is correct, but about different senses—or different uses—of 'ought'. If that is right, it opens up the possibility of resolving our paradox by saying that (1) and (5) are compatible.

II.1. Disambiguation. In the philosophical literature on ethics, it is commonly assumed that 'ought' is ambiguous between an objective and a subjective sense. Here is a representative statement:

> We can ask what one ought to do in light of all the facts. Alternatively, we can ask what one ought to do in light of available information... Standardly in moral theory, we distinguish what a person ought to do in the *objective* sense and what she ought to do in the *subjective* sense.[9]

If that is right, then we can defuse our paradox by disambiguating:

(1a) We $\text{ought}_{\text{subj}}$ to block neither shaft.
(2a) If the miners are in shaft *A*, we $\text{ought}_{\text{obj}}$ to block shaft *A*.
(3a) If the miners are in shaft *B*, we $\text{ought}_{\text{obj}}$ to block shaft *B*.
(4) Either the miners are in shaft *A* or they are in shaft *B*.
(5a) Either we $\text{ought}_{\text{obj}}$ to block shaft *A* or we $\text{ought}_{\text{obj}}$ to block shaft *B*.

(1a) is perfectly compatible with (5a).

We see two major problems with this approach to the paradox. First, the disambiguator still cannot make sense of advice. She can secure the truth of Adviser's statement (in Dialogue 1) by interpreting its

[9] Gibbard, *op. cit.*, p. 340.

'ought' in the objective sense, but only at the cost of having Adviser "talking past" Agent. If Agent has made a claim about what she ought, subjectively, to do, this claim is in no way contradicted by Adviser's claim about what she ought, *objectively*, to do. Yet Advisor takes herself—rightly, we think—to be disagreeing with Agent. His rejoinder can felicitously be prefaced by "No, ...," "I disagree, ...," or even "False!" The disambiguator cannot explain why that should be appropriate.

Moreover, two senses of 'ought' are not going to be enough. To see this, consider a slight variant of our miners case. Here, Adviser does not know where the miners are but knows more than Agent about hydrology. Adviser can see that the water will come more forcefully at shaft *A* than at shaft *B*. He knows that if both shafts are left open, the first rapid flows of water down shaft *A* will cause a thick section of clay wall to collapse, sealing off *A* from further incursion of water and causing *B* to be flooded. On the other hand, if shaft *A* is sandbagged, the sandbag wall will eventually collapse, and half the water will go into each shaft. Finally, if shaft *B* is sandbagged, all the water will go into shaft *A*. Summing up what Adviser knows:

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

In this case, the following dialogue would be natural:

Dialogue 2

Agent: I ought to block neither shaft, guaranteeing that nine are saved.

Adviser: No, you ought to block shaft *A*. That is what will guarantee that nine are saved.

What kind of 'ought' is Adviser using? It cannot be the objective 'ought', because the best choice in light of *all* the facts is either to block shaft *B* (if the miners are in *B*) or to leave both shafts open (if they are in *A*). It cannot be the subjective 'ought', either, because the best choice in light of *Agent's* evidence—which does not include the Adviser's hydrological knowledge—is to leave both shafts open. So we will need a third sense of 'ought'. By constructing more cases of this kind, we can motivate what Frank Jackson calls an "annoying profusion of 'oughts'."[10]

[10] Jackson, "Decision-Theoretic Consequentialism," p. 471.

II.2. Contextualism. Once this point has been seen, it begins to look more attractive to take 'ought' to be univocal but context sensitive. For example, we might say that

Context-sensitive ought—simple

An occurrence of 'S ought to φ' at a context c is true iff φ-ing is the best course of action available to S in light of the evidence available to the agent of c (that is, the speaker, and not, in general, S).[11]

This proposal helps with the profusion problem; however, it still has advisers talking past deliberators. If each speaker's 'ought' is contextually sensitive to that speaker's evidence, Adviser is no more contradicting Agent than he would be in the following dialogue:

Dialogue 3

Agent: [in Miami] It is warm here.

Adviser: [in Anchorage] No, it isn't warm here.

This problem might be addressed by moving to a more flexible form of contextualism:

Context-sensitive ought—flexible

An occurrence of 'S ought to φ' at a context c is true iff φ-ing is the best course of action available to S in light of the evidence relevant at c.

On this view, 'ought' can be used, depending on the context, in relation to any number of relevant bodies of evidence—including the speaker's, the audience's, or some combination of these, and possibly even evidence which has not yet been gathered. We can thus solve the "talking past" problem by taking both Agent and Adviser to be using 'ought' in relation to the *group's* collective evidence, or perhaps in relation to all the evidence that will be gathered by a particular time.

Technically, this kind of contextualism allows that a use of (1) and a use of (5) can both be true—provided they are used in contexts where different bodies of evidence are relevant. But this is not a very convincing resolution to the paradox if, as it seems to us, (1), (2), and (3) will naturally occur in a single episode of deliberation. Why should it be that, in our paradoxical argument, (1) is used relative to the agent's

[11] We note that, whereas the idea that 'ought' is ambiguous between subjective and objective senses is dominant in the philosophical literature, the idea that 'ought' is context sensitive is a commonplace in the linguistics literature. See, for example, Angelika Kratzer, "The Notional Category of Modality," in Hans-Jürgen Eikmeyer and Hannes Rieser, eds., *Words, Worlds, and Context* (New York: Walter de Gruyter, 1981), pp. 38–74.

current evidence, while (2) and (3) are used relative to a more informed body of evidence? The contextualist owes an explanation of why in such cases there should always be a shift in the contextually relevant evidence.

It seems best, then, to think of the contextualist as pursuing a different resolution to the paradox than the disambiguator: not taking (1) and (5) to be consistent, but instead either joining subjectivists in rejecting (5), if the contextually relevant body of evidence recommends blocking neither shaft, or joining objectivists in rejecting (1), if the contextually relevant body of evidence recommends blocking one of the shafts. Either way, however, the contextualist will face a version of the problems scouted before for these views.

If the contextualist takes the former route, she will still face a beefed-up argument from the possibility of advice. If Thomson is right that we do not limit our advice to what is recommended by the advisee's own evidence, it also seems right that in giving advice we are not making predictions about what might be recommended by the group's evidence, or even by the evidence that will eventually be gathered. Moreover, as we have argued elsewhere,[12] the appropriateness of a "corrective" response on the part of an adviser—that is, of saying "I disagree" or "No, that is wrong"—does not depend on whether the adviser's evidence is "contextually relevant." It persists even when the adviser is a completely unexpected source of knowledge. In order to avoid the "talking past" problem, then, the contextualist must broaden the contextually relevant sources of evidence to include any possible sources of advice, no matter how unexpected (even, say, physicists who happen to have been working on a neutrino experiment in a neighboring shaft and heard sounds coming from shaft *A*). This amounts to taking the second route—joining the objectivist in rejecting, or at least refusing to accept, (1)—since our deliberators do not have good grounds for holding that blocking neither shaft is the thing to do in light of this expanded and largely unknown body of evidence. And 'ought' judgments now seem too remote from available evidence to play a role in guiding deliberation.

III. PLAYING WITH THE LOGICAL FORM

If we do not reject a premise or construe (5) so that it is compatible with (1), then the only remaining way to resolve the paradox is to deny that (5) follows from the premises. We could do that by rejecting

[12] "Ought: Between Objective and Subjective." See also the similar arguments in MacFarlane, "Epistemic Modals are Assessment-Sensitive," in Andy Egan and Brian Weatherson, eds., *Epistemic Modals* (New York: Oxford, forthcoming).

the validity of one of the standard rules one would use in deriving (5) from these premises. Less radically, we could argue that the surface form of the argument is a misleading guide to its logical form, and that its logical form is invalid even given the standard rules.

III.1. Wide-Scoping. Perhaps the most natural suggestion along these lines is that 'ought' in (2) and (3) has wide scope over the conditional. A perspicuous representation of the argument's logical form, taking 'ought' as a propositional operator, would then be

(2w) Ought(If the miners are in shaft *A*, we block shaft *A*).
(3w) Ought(If the miners are in shaft *B*, we block shaft *B*).
(4) Either the miners are in shaft *A* or they are in shaft *B*.
(5) ∴ Either Ought(we block shaft *A*) or Ought(we block shaft *B*).

Clearly this is not a valid form—or, if it is valid, it is because of special features of 'ought', not 'if' and 'or'.

This solution has the advantage of familiarity: the idea that 'if... must' exhibits a scope ambiguity goes back to the medieval distinction between *necessitas consequentiae* and *necessitas consequentis*, and John Broome has made use of a comparable scope distinction for 'ought' in distinguishing between reasons and normative requirements.[13] However, it is not a fully general solution to our paradox. For although it blocks the paradox in its original form, it does not help with a slightly enhanced version of the paradoxical argument, presented here with wide-scope readings of the conditionals:

(2w) If the miners are in shaft *A*, we ought to block shaft *A*.
Ought(If the miners are in shaft *A*, we block shaft *A*).
(3w) If the miners are in shaft *B*, we ought to block shaft *B*.
Ought(If the miners are in shaft *B*, we block shaft *B*).
(4w) The miners must be either in shaft *A* or in shaft *B*.
Must(The miners are in shaft *A* or they are in shaft *B*).
(8) Necessarily, if we block shaft *A*, we block one shaft.
Must(If we block shaft *A*, we block one shaft).
(9) Necessarily, if we block shaft *B*, we block one shaft.
Must(If we block shaft *B*, we block one shaft).
(10) ∴ We ought to block one shaft.
Ought(we block one shaft).

(Here an operator 'must' is used for epistemic necessity; it is given wide scope in the conditionals (8) and (9), as seems plausible.)

This argument comes out valid, provided the following assumptions hold:

(A1) Modus ponens is valid for the conditional in question. Thus, if φ and $\ulcorner$if φ, $\psi\urcorner$ are true at a world *w*, then ψ is true at *w*.

[13] "Normative Requirements," *Ratio*, XII (1999): 398–419.

(A2) The ‘ought’ operator quantifies over “ideal worlds.” That is, ⌜Ought(φ)⌝ is true at a world w just in case φ is true at all the “most ideal” worlds relative to w. (This is a standard assumption when ‘ought’ is treated as a propositional operator.)

(A3) The ideal worlds relative to w are all epistemically possible relative to w. (That is: if it ought to be that φ, then it is possible that φ. This assumption is also standardly made in deontic logic.)

For, given (A1) and (A2), premise (2w) says that all the ideal worlds that are miners-in-*A* worlds are we-block-*A* worlds, and (3w) says that all the ideal worlds that are miners-in-*B* worlds are we-block-*B* worlds. But (4w) says that all the epistemically possible worlds, and hence (given A3) all the ideal worlds, are either miners-in-*A* worlds or miners-in-*B* worlds. It follows that all the ideal worlds are either we-block-*A* worlds or we-block-*B* worlds, and thus, given (8) and (9), that all the ideal worlds are we-block-one worlds. The conclusion (10) follows immediately, given (A2).

Thus the wide-scope approach can handle only *some* of the paradoxical cases. (In section IV.2, below, we will see another class of related cases, involving nested conditionals, that cannot be handled using a wide-scope strategy.)

A further strike against the wide-scope approach is that it requires us to think of conditionals as sentential connectives. Most linguists now think of conditional antecedents as modifiers of an implicit or explicit modal in the consequent, for good syntactic and semantic reasons.[14] If conditionals are modifiers of modals, then the modals they modify cannot take wide scope over them.

III.2. Dyadic Operators. Another approach, born out of the recognition that wide-scoping will not always make good sense of conditional obligation statements,[15] is to represent these statements using an irreducible dyadic conditional obligation operator. On this view, ‘if... ought’ is really an idiom, whose meaning cannot be captured by the interaction of separate components ‘if’ and ‘ought’. ⌜Ought($\psi \mid \varphi$)⌝, read ⌜it ought to be that ψ conditional on φ⌝, is true just in case ψ holds at all the worlds that are most ideal given φ. Thus, for example, ‘If Sam hits his sister, he ought to apologize’ is true, because the worlds that are most ideal given that Sam hits his sister are worlds where he also apologizes.

[14] See Kratzer, “Blurred Conditionals,” in W. Klein and W. Levelt, eds., *Crossing the Boundaries in Linguistics* (Dordrecht: Reidel, 1981), pp. 201–09, and her “Conditionals,” unpublished typescript; William Lycan, *Real Conditionals* (New York: Oxford, 2001), ch. 1.

[15] G. H. von Wright, “A Note on Deontic Logic and Derived Obligation,” *Mind*, LXV (1956): 507–09, at pp. 508–09.

If the conditionals in our paradox are represented with the dyadic conditional obligation operator, as

(2d) Ought(we block shaft A | the miners are in shaft A).
(3d) Ought(we block shaft B | the miners are in shaft B).

then (5) cannot be derived from them together with (4).[16] The enhanced argument considered in the last section is also blocked. It does follow from the premises that the worlds that are most ideal given that the miners are in A are worlds where we block one shaft, and that the worlds that are most ideal given that the miners are in B are worlds where we block on shaft. But from this we cannot conclude that the worlds that are most ideal given that the miners are either in A or in B are all worlds where we block one shaft.[17]

However, we ought to be skeptical of the idea that 'if…ought' is an idiom. Idioms tend to be idiosyncratic to languages. It would be nothing short of miraculous if all known languages just *happened* to express conditional obligation using a combination of a conditional and a word expressing obligation. The most obvious explanation of why they do is that the meanings of conditional obligation statements are determined compositionally by the meanings of these more basic constituents. If that explanation is rejected, another is needed, and as far as we know none has been offered.

Moreover, it would be surprising if 'if…ought' were linguistically much different from 'if…must', where 'must' is an epistemic modal. Deontic and epistemic modals have so much in common, both syntactically and semantically, that one would not expect deep differences in logical form. But nobody to our knowledge has proposed a dyadic analysis of

(11) If it is raining, the streets must be wet.

Finally, as Richmond Thomason points out, the dyadic approach founders on *mixed* cases, like

(12) If John has promised to give up smoking then either he ought to give up smoking or he will be released from his promise.[18]

[16] Assuming the semantics of (A2) for the monadic 'ought' in (5).

[17] This would follow given the additional assumption that, if w is among the most ideal worlds given φ, then w is among the most ideal worlds given ψ, for any ψ that entails φ and is true at w. Although some proponents of dyadic accounts seem committed to this assumption, it is not obligatory. (It is tantamount to the denial that the deontic selection function is *seriously information-dependent*, in the sense of §IV.3, below.)

[18] "Deontic Logic as Founded on Tense Logic," in Risto Hilpinen, ed., *New Studies in Deontic Logic: Norms, Actions, and the Foundations of Ethics* (Dordrecht: Reidel, 1981), pp. 165–76.

This is partly a conditional obligation statement, but partly just an ordinary indicative conditional. So it cannot be represented using a dyadic conditional obligation operator; we will need independent accounts of 'if' and 'ought'. We therefore echo Thomason's conclusion that

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

IV. REJECTING THE ARGUMENT AS INVALID

Suppose we let the paradoxical argument have the logical form it appears to have, so that (2) and (3) are indicative conditionals with 'oughts' in their consequents. Then the argument can be shown to be valid using just three basic logical rules: disjunction elimination, disjunction introduction, and modus ponens.

1	$inA \vee inB$	
2	if inA, $O(blA)$	
3	if inB, $O(blB)$	
4	$\quad inA$	
5	$\quad O(blA)$	2, 4, MP
6	$\quad O(blA) \vee O(blB)$	5, $\vee$ intro
7	$\quad inB$	
8	$\quad O(blB)$	3, 7, MP
9	$\quad O(blA) \vee O(blB)$	8, $\vee$ intro
10	$O(blA) \vee O(blB)$	1–9, $\vee$ elim

So if we are to reject the argument as invalid, we must reject one of these rules.

IV.1. Rejecting Disjunction Introduction or Elimination. Rejecting disjunction introduction and elimination would be difficult to motivate independently, and it is easy to see that these moves will not get to the bottom of the problem.

If we reject disjunction introduction, we can block steps 6 and 9 in the above proof. But the paradox can be reinstated by adding two new premises that can hardly be rejected:

(13) If we ought to block shaft *A*, then we ought to block at least one shaft.

(14) If we ought to block shaft *B*, then we ought to block at least one shaft.

Using these premises together with our old ones, we can derive 'we ought to block at least one shaft' without using disjunction introduction at all. This conclusion is just as paradoxical as the old one. So rejecting disjunction introduction will not help.

Rejecting disjunction elimination will block both of these proofs. But it will not help with a simpler paradoxical argument that uses only one conditional premise:

(1) We ought to block neither shaft.
(15) ∴ It is not the case that we ought to block shaft *A*.
(2) If the miners are in shaft *A*, we ought to block shaft *A*.
(16) ∴ The miners are not in shaft *A*.

Clearly, the premises here do not support the conclusion. We cannot deduce the location of the miners simply by reflecting on our moral predicament. In this case the disjunction rules cannot be blamed. We have, however, relied on modus tollens, and hence indirectly on modus ponens, since modus tollens can be proved using reductio and modus ponens:

$$
\begin{array}{r|ll}
1 & \text{if } \phi, \psi & \\
2 & \neg\psi & \\ \hline
3 & \quad \phi & \\
4 & \quad \psi & 1, 3, \text{MP} \\
5 & \quad \bot & 2, 4, \bot \text{ intro} \\
6 & \neg\phi & 3\text{–}5, \text{reductio}
\end{array}
$$

Modus ponens is the only common factor between this paradox and the original one. Thus, we point the finger at modus ponens. If we are to resolve the paradox without rejecting a premise, we must reject the widely held view that modus ponens is a valid argument form.[19]

[19] Meg Wallace questions whether modus ponens can really be the heart of the problem, suggesting that a similar paradox could be constructed using the disjunctions

(17) Either the miners are not in *A*, or we ought to block off *A*.
(18) Either the miners are not in *B*, or we ought to block off *B*.

instead of the conditionals (2) and (3). The objection is only a serious one if the 'or' in (17) and (18) is construed as a truth-functional disjunction, for if it is read intensionally—such that $\ulcorner\varphi \text{ or } \psi\urcorner$ is equivalent to $\ulcorner\text{if}\neg\varphi, \psi\urcorner$—then rejecting modus ponens is relevant after all. (In favor of the intensional reading, we note that transposing the disjuncts in (17) and (18) seems to make a difference to their acceptability.) Suppose, then, that (17)

IV.2. Rejecting Modus Ponens. We doubt that our readers will be willing to give up modus ponens *just* to deal with our paradox. So, before offering a semantics that invalidates modus ponens, we want to note that there are good reasons for thinking modus ponens invalid, quite independently of inferences involving 'ought'.

Here is an analogue of our paradox using epistemic 'must':

(19) The murder might have occurred in the morning, and it might have occurred in the evening. [We do not know which.]
(20) If the butler did it, the murder must have occurred in the morning.
(21) If the nephew did it, the murder must have occurred in the evening.
(22) Either the butler did it or the nephew did it [but we do not know which].
(23) ∴ Either it must have occurred in the morning or it must have occurred in the evening.

The conclusion of the argument, (23), is inconsistent with (19). And we have the same options as before. Here, of course, most philosophers will be inclined to go for a wide-scope solution. And in this case, wide-scoping *will* work. But, given the close kinship of epistemic and deontic modals, it would be odd to deal with these very similar paradoxes in very different ways. If wide-scoping will not help with the version using deontic modals, that gives us a reason not to use it here either. But then, unless we are going to reject the premises, it seems we must reject modus ponens.[20]

There is, in addition, Vann McGee's famous counterexample to modus ponens:[21]

(24) If a Republican wins the election, then if it is not Reagan who wins it will be Anderson.
(25) A Republican will win the election.
(26) ∴ If it is not Reagan who wins, it will be Anderson.

The context is just before the 1980 US presidential election, in which (Republican) Ronald Reagan was running against (Democrat) Jimmy

and (18) are truth-functional disjunctions. Why should they seem acceptable to a reasoner who is ignorant of the location of the miners and thus rejects their second disjuncts? Surely, because they seem to follow from the conditionals (2) and (3). Note, however, that if modus ponens is rejected, so is the deduction of (17) from (2), and of (18) from (3). So by rejecting modus ponens, we can explain not only why the original paradox seems valid even though it is not, but also why someone might suppose that (17) and (18) are warranted even though they are not.

[20] Note that if we reject modus ponens, we also remove one of the main motivations for a wide-scope reading of (20) and (21), which is to find a reading of these sentences on which they can be true.

[21] "A Counterexample to Modus Ponens," this JOURNAL, LXXXII (1985): 462–71.

Carter, with (Republican) John Anderson as a third candidate. The fact that there were only two Republicans in the race made (24) unassailable. And (25) was, we now know, true. But the conclusion (26) was presumably false, since Anderson had virtually no chance of getting more votes than Carter. So again we have a counterexample to the validity of modus ponens, and in this case wide-scoping does not even seem to be an option.

IV.3. Semantics for Informational Modals. We do not propose to reject modus ponens solely on the basis of the counterexamples. We would like to have some account of *why* modus ponens fails when it does, and also of why it seems to work fine in most cases. To discharge these tasks, we will need a semantic account of epistemic and deontic modals and indicative conditionals.

Our semantics will take the form of a recursive definition of truth at a point of evaluation. A point of evaluation will normally consist of a context and an index, the latter consisting of a possible world-state, an assignment of values to the variables, and perhaps more. For our purposes here, however, we can make do with a very simple representation of points of evaluation:

Point of evaluation

A *point of evaluation* is a pair $\langle w, i \rangle$, where w is a possible world-state (representing epistemic possibilities), and i is an information state (a set of possible world-states).

Our *possible world-states* can be thought of as assignments of extensions to all the basic predicates and terms of the language.[22] They are meant to represent epistemic possibilities—ways the world might actually be—and not alethic possibilities—ways the world could have been. So there can be a world-state that assigns Falsity to 'Hesperus is Phosphorus', for example. We model an *information state* as a set of possible world-states: intuitively, the set of state descriptions that might, given what is known, depict the actual world.[23]

[22] Predicates like 'is obligatory' must be excluded here, since their extensions will be definitionally connected to complex sentences involving deontic modals.

[23] This is an *epistemic* and *nonprobabilistic* model of information states; it takes information states to be sets of known facts. We have chosen this model because we think that what one ought to do (relative to an information state) supervenes on what is known: mere differences in beliefs (or partial beliefs) or perceptual states, unaccompanied by differences in what is known, cannot make a difference to what an agent ought to do. This is, of course, a substantive assumption. Much of what we say in what follows about the semantics of deontic operators can be modified to work with nonepistemic or probabilistic models of information states, for example, a model of an information state as an assignment of probabilities to sets of worlds. We will flag points where we assume epistemic information states.

We think of epistemic and deontic modals as specifications of generic *informational* modal operators. What distinguishes informational modals from other kinds of modals is that they are sensitive to an information state—a set of epistemically possible worlds. The generic informational modals have the following semantics:[24]

$\Box_f$ *and* $\Diamond_f$

$\ulcorner \Box_f \varphi \urcorner$ is true at $\langle w, i \rangle$ iff for all $w' \in f(i)$, φ is true at $\langle w', i \rangle$.

$\ulcorner \Diamond_f \varphi \urcorner$ is true at $\langle w, i \rangle$ iff for some $w' \in f(i)$, φ is true at $\langle w', i \rangle$.

Here f is a *selection function*, generally supplied by context. Depending on f, '$\Box_f$' will be an epistemic necessity operator ('it must be the case that') or one of many different sorts of deontic necessity operators ('it ought [legally/morally/according to the rules of my club] to be the case that').[25]

An *epistemic selection function*, e, maps an information state to the set of worlds that might, as far as this state knows, be actual. In our framework we can assume $e(i) = i$ for all i.[26]

A *deontic selection function*, d, maps an information state to the set of worlds that are *as deontically ideal as possible*, given that information.[27] Deontic ideality is a special kind of ideality. A world can be much more ideal than another in other ways (for example, in how fortunate

[24] This approach to modals differs from the usual approach in one important respect: the set of worlds over which the modal operators quantify is provided by a separate parameter (the information state) rather than being determined by the world of evaluation and an accessibility relation. See Seth Yalcin, "Epistemic Modals," *Mind*, CXVI (2007): 983–1026, and MacFarlane, "Epistemic Modals," for arguments for such an approach to epistemic modals.

[25] There is some evidence that 'ought' is a weaker necessity operator than (deontic) 'must', but we will ignore this distinction in what follows.

[26] If we represent information states probabilistically, as functions from sets of worlds to probabilities, things get more interesting. If the set of worlds is finite, we can define $e(i) = \{w \mid i(\{w\}) > 0\}$. If there are infinitely many worlds, this definition will not work, since a set of possible worlds may be assigned probability 0. For example, the probability that a randomly selected point on the globe will be on the equator is 0, but it is not *impossible* that such a point will be on the equator. For many purposes, though, it is harmless to assume that the set of world-states is finite. If this assumption is not made, we will need a more complex representation of i and a different definition of e. For an example of such a framework, see Yalcin, "Epistemic Modals."

[27] In assuming that there is such a set, we presuppose that it will not be the case that for every world w in an information state i, there is another world w' in i that is more deontically ideal than w relative to i. This is a safe assumption if (a) there can be only finitely many agents, (b) each agent can have only finitely many possible choices, and (c) no two worlds where agents make the same choices differ in respect of deontic ideality (relative to i). If the assumption were relaxed, a more complex account of the informational modals would be needed (cf. David Lewis, *Counterfactuals* (Cambridge: Harvard, 1973)).

people are) without being deontically more ideal. We will not try to characterize deontic ideality generically. It is natural to think that the species of deontic ideality relevant to our efforts to save the miners depends somehow on choice.[28] But this does not seem true of all species of deontic ideality. For, in addition to talking of what agents ought to do, we talk of what thinkers ought to believe, and even of how engines ought to work. Thinkers do not generally *choose* what to believe, and engines certainly do not choose to function properly. In the semantics itself, we want to remain neutral about how one should think of deontic ideality. Consequentialists may want to think of it in terms of maximization of expected utility (in light of an information state), while deontologists may want to think of it in terms of satisfaction of principles. In addition, different kinds of deontic ideality—moral, legal, prudential, role-based, and so on—may be at issue in different uses of deontic modals. Context will determine how the modal is to be interpreted by supplying a selection function.

We will assume that deontic selection functions are *realistic*:

Realistic

A deontic selection function d is *realistic* iff for all information states i, $d(i) \subseteq i$.

Suppose it is known that Sam has insulted Jane. Then it will be the case that worlds in which Sam apologizes to Jane after insulting her count as deontically ideal relative to our information state even though, speaking absolutely, it would have been more ideal had Sam not insulted Jane in the first place.

[28] One way of making this dependence explicit would be:

Ought implies can choose

For all φ, $\langle w, i\rangle$, if $\ulcorner\Box_d\varphi\urcorner$ is true at $\langle w, i\rangle$ and $d(i)$ is nonempty, then φ is choosable relative to $\langle w, i\rangle$.

Choosable

φ is choosable relative to $\langle w, i\rangle$ iff there is some action specification Δ such $\ulcorner\Diamond_e(\Delta$ is done by agents who know they are doing $\Delta)\urcorner$ and $\ulcorner\Box_e(\Delta$ is done $\supset \varphi)\urcorner$ are both true at $\langle w, i\rangle$.

This would explain why '$\Box_d$(We save all ten)' is not true relative to our incomplete information, even though it brings about the best outcome. Saving all ten is not choosable relative to our information state: though we can knowingly block shaft A, it is not epistemically necessary (given our information) that this act will save all ten miners. It also explains why '$\Box_d$(We save all ten)' *is* true relative to an informed observer's more complete information: from the observer's perspective, saving all miners is choosable. That is, there is a specific action we can (knowingly) perform that will guarantee the miners' safety (blocking shaft A).

It is important to keep in mind that it is not just the set of ideal worlds that varies as the information state is shifted, but also the *ranking* of worlds as more or less ideal. A world may be more ideal than another relative to one information state and less ideal than it relative to another. For example, a world in which both shafts are left open may be more ideal than one in which shaft A is closed relative to a less informed state, but less ideal relative to a more informed state. Deontic selection functions therefore can be *seriously information-dependent*:

Seriously information-dependent

A deontic selection function d is *seriously information-dependent* iff for some information states i_1, $i_2 \subseteq i_1$, there is a world $w \in i_2$ such that $w \in d(i_1)$ but $w \notin d(i_2)$.

Intuitively: an ideal world can be nonideal relative to a contracted information state that contains it. Because of this, worlds cannot be ranked for ideality independent of an information state.[29]

We acknowledge, finally, that our decision to treat 'ought' as a deontic necessity operator brings some problems in its wake. First, one might worry about regimenting sentences of the form $\ulcorner S$ ought to $\varphi\urcorner$ as $\ulcorner$Ought(S φs)$\urcorner$. Syntactically, 'ought' takes a subject and an infinitival phrase as its complement; a deontic necessity operator, by contrast, takes a sentential complement. So, although we are in good company in analyzing 'ought' as a modal box, we want to flag some discomfort with this strategy. Second, by treating 'ought' as a necessity operator and assuming that it is realistic, in the sense defined above, we commit ourselves to the validity of the following inference forms:

(27) $\Box_d\varphi, \Box_e(\varphi \supset \psi)$ / $\therefore \Box_d\psi$
(28) $\Box_e\psi$ / $\therefore \Box_d\psi$

Both inference forms lead to paradoxical-sounding conclusions. The first leads to Ross's paradox: if you ought to post the letter, it follows that you ought to either post the letter or burn it. The second implies that it ought to be the case that $2 + 2 = 4$, and that it ought to be the case that Lincoln was assassinated (since it is now epistemically necessary that this was so).[30]

[29] Here our view contrasts with that of Lewis (*op. cit.*, p. 96), who assumes a fixed ranking of worlds (relative to each world of evaluation), and Kratzer (*op. cit.*), who takes the ranking of worlds to be determined by a contextually supplied "ordering source" and the world of evaluation.

[30] One response to this second problem is to revise the definition of $\ulcorner\Box_d\varphi\urcorner$ to require that φ be not only true at all worlds in $d(i)$, but also not true at all worlds in i. However,

There is considerable controversy over whether these problems require an alternative semantic treatment of 'ought' or some other treatment. We think these problems are orthogonal to the issues we are dealing with here and so propose to lay them aside for now. Even if it is not the final story, treating 'ought' as a modal operator can yield genuine illumination about the paradox we set out to solve.

IV.3.1. Semantics for Indicative Conditionals. We follow Kratzer in taking conditional antecedents to be modifiers of modals, rather than sentential connectives. We will represent $\ulcorner$if $\varphi\urcorner$ as an operator $\ulcorner[\mathit{if}\ \varphi]\urcorner$, and impose the syntactic constraint that this kind of operator may occur only in front of an informational modal. In indicative conditionals, the modal is normally an epistemic modal, so when ψ lacks an explicit modal, the indicative $\ulcorner$if φ, $\psi\urcorner$ gets analyzed as $\ulcorner[\mathit{if}\ \varphi]\Box_e\psi\urcorner$. But 'if' can modify explicit informational modals of all kinds. For example, 'if it rains, the game might be canceled' will have the form $\ulcorner[\mathit{if}\ \varphi]\Diamond_e\psi\urcorner$. And 'if it rains, then you ought to take an umbrella' will have the form $\ulcorner[\mathit{if}\ \varphi]\Box_d\psi\urcorner$.[31]

As a first approximation, we can think of $\ulcorner[\mathit{if}\ \varphi]\urcorner$ as contracting the information state by ruling out worlds at which φ is false:

$[\mathit{if}\ \varphi]$ (first approximation)

$\ulcorner[\mathit{if}\ \varphi]\psi\urcorner$ is true at $\langle w, i\rangle$ iff ψ is true at $\langle w, i'\rangle$,
where $i' = \{w' \in i \mid \varphi \text{ is true at } \langle w', i\rangle\}$.

This is intuitively plausible: to evaluate $\ulcorner$if φ, it must be that $\psi\urcorner$, we ask whether the truth of ψ is guaranteed by our existing stock of information together with the truth of φ.

However, this account is problematic when the antecedent itself contains informational modals. Consider

by allowing '$\Box_d$' to differ from '$\Box_e$' in more than just the selection function, this would strike against the unity of the informational modals. A less drastic response is just to say that when $\ulcorner\Box_e\varphi\urcorner$ is true, $\ulcorner\Box_d\varphi\urcorner$, while true, is deliberatively irrelevant: pointless to consider in decision-making, or to offer as advice. Suppose that we expect 'ought' propositions to be deliberatively relevant. Then we may tend to try to evaluate them relative to information states at which they are deliberatively relevant. This might explain why 'It ought to be the case that $2 + 2 = 4$' strikes us as bizarre, whereas 'It ought to be the case that Lincoln was assassinated' strikes us as straightforwardly false. There is no information state at which the former is deliberatively relevant. By contrast, there is an information state, such as that of a concerned American on the morning of April 14, 1865, at which the latter is deliberatively relevant, and relative to that information state, it is false.

[31] It actually would not make a difference if all these conditionals were taken to have an implicit epistemic necessity operator in front of the explicit modal, since in our system, '$\Box_e\Diamond_e$' is equivalent to '$\Diamond_e$', and '$\Box_e\Box_d$' to '$\Box_d$'.

(29) If we ought to block shaft A, then we ought to start moving sandbags.

$[if\ \Box_d blA]\Box_d M$

On the account above, whether 'we ought to block shaft A' is true at a point $\langle w, i\rangle$ depends only on i, not on w. So i' will be either i (when the antecedent is true) or the empty set (when the antecedent is false). In the former case, the conditional will have the same truth value as its consequent, and in the latter case it will be trivially true. So the conditional will behave like a material conditional.

Things are even worse for the first approximation account when the truth of the antecedent depends on both the world and the information state, as in

(30) If the miners are in shaft B but it is possible that they are not, ...[32]

If we start out with an information state i containing both miners-in-A worlds and miners-in-B worlds, and remove all the worlds w such that the antecedent of (30) is false at $\langle w, i\rangle$, we are left with a state i' containing only the miners-in-B worlds. Note, however, that the antecedent is false relative to $\langle w', i'\rangle$ for every world $w' \in i'$ (because of its second conjunct). So, bizarrely, the first approximation account tests such conditionals for truth by seeing whether their consequents are true throughout an information state where the antecedent is false. That makes little intuitive sense.

Moreover, as Yalcin notes, indicative conditionals beginning $\ulcorner$if φ and possibly $\neg\varphi\urcorner$ seem incoherent in much the same way as do conditionals with antecedents known to be false. An attractive explanation for this is that when we contract down to a state containing only φ worlds, the second conjunct of the antecedent is no longer true; it is impossible to find an information state such that both φ and $\ulcorner\Diamond_e\neg\varphi\urcorner$ are true throughout the state. But the (first approximation) account above cannot explain the incoherence of these conditionals in this way, since it does not require that the antecedent be true relative to the contracted information state. On that account, $\ulcorner$if φ and possibly $\neg\varphi\urcorner$ has essentially the same effect on the information state as $\ulcorner$if $\varphi\urcorner$.

The key to a solution is to find a contracted information state relative to which the antecedent is true. More precisely: a subset i' of the original information state i such that the antecedent is *true throughout* i':

True throughout

φ is *true throughout* an information state i iff for all $w \in i$, φ is true at $\langle w, i\rangle$.

[32] We owe this point to Yalcin, *op. cit.*

Yalcin defines i' as the *largest* subset of i such that the antecedent is true throughout i'.[33] Though this idea seems to us to be on the right track, one cannot assume that there is a unique largest such subset. Consider, for example,

(31) If we ought to close just one shaft, then the miners are in shaft *A*.
(32) If we ought to close just one shaft, then the miners are in shaft *B*.

Here there are two subsets of the original (ignorant) information state at which the antecedent is true: one containing only worlds at which the miners are in *A*, one containing only worlds at which the miners are in *B*. Both are maximal in the sense that matters:

Maximal φ-subset

i' is a maximal φ-subset of i iff (a) φ is true throughout i', and (b) there is no i'' such that $i' \subset i'' \subseteq i$ and φ is true throughout i''.

Given the symmetry of the epistemic situation, it would certainly be odd to say that one of these conditionals is true and the other false. We think that neither conditional is true (relative to the original state of ignorance about the miners' location). This suggests that the truth of a conditional requires truth at *all* of the maximal contracted information states at which the antecedent is true. More precisely:

[*if* φ] *(revised)*

$\ulcorner[\mathit{if}\ \varphi]\psi\urcorner$ is true at $\langle w, i\rangle$ iff ψ is true at $\langle w, i'\rangle$ for every maximal φ-subset i' of i.

This semantics predicts the truth of (2) and (3) in our paradoxical inference. For, if we remove all the worlds from our original (ignorant) information state in which the miners are not in shaft *A*, we are left with an information state that "knows" the miners are in *A*; and relative to this state, we ought to block shaft *A*. Similarly, if we remove all the worlds from our original information state in which the miners are not in shaft *B*, we are left with a state that "knows" the miners are in *B*; and relative to this state, we ought to block shaft *B*.[34]

[33] *Idem*, p. 998.

[34] For a similar approach to the interaction of modals and conditionals, see John Cantwell, "Changing the Modal Context," *Theoria*, LXXIV (2008): 331–51. Like us, Cantwell takes conditional antecedents to restrict the class of worlds over which their modal consequents quantify; like us, he notes that modus ponens, modus tollens, and reasoning by cases are invalid on such a semantics. One important difference is that,

IV.4. Why Modus Ponens Is Invalid. We are now in a position to see why modus ponens should be invalid for a conditional with this semantics. First, though, we need to say what validity is:

Validity

An argument is *valid* iff there is no information state i and world $w \in i$ such that the premises are all true at $\langle w, i \rangle$ and the conclusion is false at $\langle w, i \rangle$.

The restriction to points $\langle w, i \rangle$ where $w \in i$ needs some motivation. The thought here is that, in defining validity, we should restrict ourselves to "proper" points of evaluation—points that could correspond to the actual situation and information of a reasoner.[35] Since we are assuming that information is knowledge, and that nothing false can be known, the "actual world" of a reasoner must belong to the set of epistemically open worlds for that reasoner.[36]

The reason that modus ponens is invalid is then simple to state. On our semantics, ⌜If φ, ψ⌝ is true iff ψ is true relative to the φ-shifted information state(s). But this can be so even if φ is true and ψ false relative to the original, *non*shifted information state.

The point can be illustrated using McGee's counterexample (section IV.2). (24) is true because its consequent (26) is true throughout the information state that results when all the Republican-losing worlds are removed. For (26) to be true simpliciter, however, it would have to be true throughout the *original* information state. So, in order for the argument to be valid, the remaining premise (25) would have to make up the difference, ensuring that (26) is true at all the Republican-losing worlds in the contextually relevant information state. Of course, it cannot do this, since its truth does not depend on what goes on in any nonactual worlds.[37]

---

on Cantwell's view, deontic modals are not "seriously information-dependent" in the sense defined above. This is so because the set of worlds over which such modals quantify is generated by an information-independent ranking of worlds (p. 346; cf. note 29, above). Thus, although Cantwell's view helps with the gentle murder paradox, it does not help with our miners case.

[35] This is one key idea of David Kaplan's "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals," in Joseph Almog, John Perry, and Howard Wettstein, eds., *Themes from Kaplan* (New York: Oxford, 1989), pp. 481–566.

[36] Validity so defined amounts to preservation of truth at every context of use, given the contextualist definition of truth at a context (section V), and to preservation of the property of being true as used at and assessed from the same context, given the relativist definition. This latter notion might well be called "diagonal validity."

[37] Lycan gives a similar analysis (*op. cit.*, pp. 66–69).

Next, consider the miners case. As noted above, (2) is true because, relative to a shifted information state including only worlds where the miners are in shaft *A*, we ought to block shaft *A*. Now it may in fact be the case that the miners are in shaft *A*. But that would not make it the case that

(33) We ought to block shaft *A*.

is true relative to our original information state—the one that includes both worlds where the miners are in shaft *A* and worlds where they are in shaft *B*.

IV.5. Life without Modus Ponens. It may seem insane to deny the validity of modus ponens. This is an inference form we rely on all the time. Some philosophers have even taken it to be constitutive of the meaning of the conditional.[38] So how can we reject it? Isn't the fact that our semantics for the conditional does not validate it just a refutation of our semantics?

We think not. Here are some considerations that should help make rejecting modus ponens seem less outrageous.

First, we are in no way questioning the validity of modus ponens for the *material* conditional used in first-order logic:

MP0

$\varphi \supset \psi, \varphi \; / \; \therefore \; \psi$

We are only questioning the validity of modus ponens for the natural-language *indicative* conditional. To be more precise, we are rejecting the inference forms

MP1

$[\mathit{if}\ \varphi]\Box_e\psi, \varphi \; / \; \therefore \; \psi$

MP2

$[\mathit{if}\ \varphi]\psi, \varphi \; / \; \therefore \; \psi$

(We give both forms, since when the conditional premise of a modus ponens inference contains an implicit epistemic necessity operator, the conclusion of the inference is usually given without the operator.)

[38] For example, R. M. Hare, "Meaning and Speech Acts," *Philosophical Review*, LXXIX (1970): 3–24, at p. 16; Paul Boghossian, *Fear of Knowledge: Against Relativism and Constructivism* (New York: Oxford, 2006), p. 107.

Second, the use of modus ponens in most ordinary reasoning can be vindicated. For, although MP1 and MP2 are not *valid*, they are truth-preserving under (roughly) the following conditions:

(a) when the antecedent is already known (because then the information state does not shift), and
(b) when the consequent is not sensitive to the information state (because then the shifting does not matter).

To make (a) more precise, we define the notion of *quasi-validity*:

Quasi-valid

An inference from premises $\varphi_1, \varphi_2, ..., \varphi_n$ to conclusion ψ is *quasi-valid* iff the inference from $\ulcorner \Box_e \varphi_1 \urcorner, \ulcorner \Box_e \varphi_2 \urcorner, ..., \ulcorner \Box_e \varphi_n \urcorner$ to ψ is valid.

Quasi-validity is related to the following informal property of inferences, which (as Daniel Nolan notes) is easily confused with validity: the conclusion must be true if the premises are *known*.[39] Although MP1 and MP2 are not valid, they are quasi-valid:

Theorem 1

MP2 is quasi-valid.

Proof: Suppose $\ulcorner \Box_e \varphi \urcorner$ and $\ulcorner \Box_e [\mathit{if}\ \varphi]\psi \urcorner$ are true at $\langle w, i \rangle$, where $w \in i$. Then, since $w \in i$, $\ulcorner [\mathit{if}\ \varphi]\psi \urcorner$ is true at $\langle w, i \rangle$. Since $\ulcorner \Box_e \varphi \urcorner$ is true at $\langle w, i \rangle$, φ is true throughout i, and i is itself a maximal φ-subset of i. So, by the semantics for the conditional, ψ is true at $\langle w, i \rangle$.

Corollary 2

MP1 is quasi-valid.

Proof: This follows immediately from Theorem 1 and the fact that the argument from $\ulcorner \Box_e \psi \urcorner$ to ψ is valid.

Quasi-validity is a good standard for inferences in categorical contexts, where one is drawing new conclusions from what one takes to be known facts. So, we should expect that modus ponens inferences should seem unobjectionable in categorical contexts, and that is what we find. When you *know* that it is raining, there is nothing wrong with inferring as follows:

(34) If it is raining, the streets must be wet.

[39] Nolan, "Defending a Possible-Worlds Account of Indicative Conditionals," *Philosophical Studies*, CXVI (2003): 215–69, at p. 231; Robert Stalnaker, in "Indicative Conditionals," *Philosophia*, V (1975): 269–86, invokes a similar property to explain the plausibility of the "direct argument" from $\ulcorner P$ or $Q \urcorner$ to $\ulcorner$if not-P then $Q \urcorner$.

(35) It is raining.
(36) So, the streets must be wet.

Similarly, if you *know* that a Republican will win the race—perhaps you have inside information that the race is fixed—and that, if a Republican wins, if it is not Reagan it will be Anderson, then you can safely infer that if Reagan does not win, Anderson will.[40]

It is when the premises are not asserted as known, but rather supposed hypothetically, that modus ponens can lead one astray. Suppose you are in your office with the blinds down. You have not been outside for a while, and you remark,

(37) The streets might not be wet.
(38) If it is raining, the streets must be wet.

By using modus ponens inside a hypothetical context, you could then conclude, without any evidence at all, that it is not raining:

(39) Suppose (for reductio) that it is raining.
(40) Then the streets must be wet. (modus ponens, 38, 39)
(41) But it is not the case that the streets must be wet. (from 37)
(42) So, by reductio, it is not raining.

The same move can be used to construct a more powerful variant of McGee's counterexample, in which the modus ponens step is forced inside a subproof:

(43) If a Republican wins, then if Reagan does not win, Anderson will. (premise)
(44) It is not the case that if Reagan does not win, Anderson will. (premise)
(45) Suppose (for reductio) that a Republican will win.
(46) Then, if Reagan does not win, Anderson will. (modus ponens, 43 and 45)
(47) But this contradicts (44).
(48) So, by reductio, a Republican will not win.

[40] Bernard D. Katz, in "On a Supposed Counterexample to Modus Ponens," this JOURNAL, XCVI (1999): 404–15, at p. 414, seems to be thinking of McGee's counterexample in a categorical context, where the premises are accepted and not merely hypothesized: "In order to evaluate (24) ... we must first look at the consequent of (25), that is, (26), in light of our initial stock of beliefs adjusted to include the antecedent of (24), that is, (25); of course, *since we already accept (25), our adjusted stock of beliefs will be exactly the same as our initial stock of beliefs,* which is why (24) and (26) have the same truth value" (emphasis added and numbering changed). As noted above, it is easy to dismiss the counterexample if one thinks of it in this kind of context, since the argument is at least quasi-valid. See below for a version of McGee's argument that is not quasi-valid, and thus not even tempting in categorical contexts.

Unlike McGee's original counterexample, the inference from (43) and (44) to (48) is not even quasi-valid.

Of course, we often do use modus ponens without running into trouble, even in hypothetical contexts. Consider, for example, the following inference:

(49) If the miners are in shaft *A*, they have a jackhammer.
(50) If the miners are in shaft *B*, they have a blowtorch.
(51) Either the miners are in shaft *A* or they are in shaft *B*.
(52) So, either they have a jackhammer or they have a blowtorch.

This seems unobjectionable, even though formally it is like our paradoxical inference, which is not even quasi-valid. Fortunately, this inference can be vindicated. It differs relevantly from the paradoxical inference in having an information-invariant consequent:

Information-invariant

A formula φ is *information-invariant* just in case, for all information states i and i' and worlds w, φ is true at $\langle w, i\rangle$ iff φ is true at $\langle w, i'\rangle$.

World-invariant

A formula φ is *world-invariant* just in case, for all worlds w and w' and information states i, φ is true at $\langle w, i\rangle$ iff φ is true at $\langle w', i\rangle$.

Theorem 3 (Restricted modus ponens)

If φ is either world-invariant or information-invariant[41] *and ψ is information-invariant, then the inference from φ and $\ulcorner[if\ \varphi]\Box_e\psi\urcorner$ to ψ is valid.*

Proof: Suppose φ and $\ulcorner[if\ \varphi]\Box_e\psi\urcorner$ are true at $\langle w, i\rangle$, where $w \in i$. By assumption φ is either world-invariant or information-invariant.

- If φ is world-invariant, then i itself is a maximal φ-subset of i. Since by assumption $w \in i$, w is in a maximal φ-subset of i.
- If φ is information-invariant, then φ is true throughout $\langle w, \{w\}\rangle$. Since $w \in i$, $\{w\} \subseteq i$. If $\{w\}$ is not a maximal φ-subset of i, this can only be because $\{w\}$ is a subset of a maximal φ-subset of i. So w is in a maximal φ-subset of i.

[41] To see why this restriction on the antecedent is needed, let φ = 'we ought to block neither shaft and the miners are in shaft *A*' and ψ = 'the miners are not in shaft *A*'. (Note that φ is neither world-invariant nor information-invariant; the truth of its second conjunct varies with the world, while the truth of its first conjunct varies with the information state.) The conditional $\ulcorner[if\ \varphi]\Box_e\psi\urcorner$ is vacuously true at $\langle w, i\rangle$, since the only maximal φ-subset of i is $\varnothing$. Choose a point $\langle w, i\rangle$ such that the miners are in shaft *A* at w and i is ignorant about the location of the miners. Then φ is true and ψ false at $\langle w, i\rangle$, and we have a counterexample to the unrestricted theorem.

Either way, there is some maximal φ-subset of i—call it i'—such that $w \in i'$. Since $\ulcorner[if\ \varphi]\Box_e\psi\urcorner$ is true at $\langle w, i\rangle$, $\ulcorner\Box_e\psi\urcorner$ is true at $\langle w, i'\rangle$. So for all $w' \in i'$, ψ is true at $\langle w', i'\rangle$. Since $w \in i'$, ψ is true at $\langle w, i'\rangle$. Since ψ is information-invariant, it follows that ψ is true at $\langle w, i\rangle$.

In sum: although modus ponens is not valid, its use in categorical contexts can be vindicated across the board (because it is quasi-valid), and its use in hypothetical contexts can be defended in a restricted range of cases—where the consequent is information-invariant and the antecedent is either world-invariant or information-invariant. Outside of these restricted bounds, modus ponens can fail to preserve truth, and indeed we can find intuitive counterexamples.[42]

V. CONTEXTUALISM OR RELATIVISM?

So far we have only discussed truth at a point of evaluation. We have not said anything about how truth at a point of evaluation relates to truth at a context. A natural thought would be to embed this view in a contextualist framework:

Contextualist version

An occurrence of a sentence S at a context c is true iff S is true at $\langle w_c, i_c\rangle$, where w_c is the world of c and i_c is the information state relevant at c.

This would leave us with something like the flexible contextualist view discussed in section II.2, above. When we considered contextualism earlier, it was as a way to resolve the paradox by making (1) and (5) consistent. We complained that the contextualist had no good explanation of why the context should shift in just the way required to make these consistent. The present proposal, by contrast, can solve the paradox in another way (by rejecting modus ponens), even in contexts where (1) and (5) are inconsistent.

However, our other criticisms of flexible contextualism would still apply to this version of it. Contextualism does not yield the right predictions about the appropriateness of responses like "I disagree" and "No, that is wrong." It can explain these to an extent, by appealing to the flexibility of "relevant," but if this flexibility is pressed too far, it becomes difficult to understand how speakers ever take themselves to be warranted in asserting that they ought to do something.

[42] Of course, conditional proof will have to go, too, although we can no doubt recover restricted forms of it. Without restrictions, we could use conditional proof to derive 'if the miners are in shaft *A*, we ought to leave both shafts open' from 'we ought to leave both shafts open'. Restrictions on reiteration into conditional proof contexts are standard for modal conditionals.

Hence, we prefer a relativist version of the idea:[43]

Relativist version

An occurrence of a sentence S at a context c_1 is true as assessed from a context c_2 iff S is true at $\langle w_{c_1}, i_{c_2} \rangle$, where w_{c_1} is the world of c_1 and i_{c_2} is the information state relevant at c_2.

Here it is the context in which a use of a sentence is *assessed* that determines which informational state is relevant, not the context of use.

This is not the place to argue further for the relativist version.[44] Most of the arguments in this paper have been neutral between the two versions. Here we just want to note one thing. Because we are taking epistemic and deontic modals to be sensitive to the same "information state" parameter, the decision must go the same way for both. So, given the semantics proposed above, arguments for a relativist treatment of epistemic modals[45] and arguments for a relativist treatment of deontic modals are mutually supporting.

NIKO KOLODNY
JOHN MACFARLANE

University of California, Berkeley

[43] For background, see MacFarlane, "The Assessment Sensitivity of Knowledge Attributions," *Oxford Studies in Epistemology* I (New York: Oxford, 2005), pp. 197–233, and "Relativism and Disagreement," *Philosophical Studies*, CXXXII (2007): 17–31.

[44] We defend the relativist version further in a companion paper, "Ought: Between Objective and Subjective."

[45] MacFarlane, "Epistemic Modals."