Modus Ponens and the Logic of Dominance Reasoning

Nate Charlow

Abstract

If modus ponens is valid, you should take up smoking.

1 Introduction

Some recent work has challenged two principles thought to govern the logic of the indicative conditional: modus ponens (Kolodny & MacFarlane 2010) and modus tollens (Yalcin 2012). There is a fairly broad consensus in the literature that Kolodny and MacFarlane's challenge can be avoided, if the notion of logical consequence is understood aright (Willer 2012; Yalcin 2012; Bledin 2014). The viability of Yalcin's counterexample to modus tollens has meanwhile been challenged on the grounds that it fails to take proper account of context-sensitivity (Stojnić forthcoming).

This paper describes a new counterexample to modus ponens, and shows that strategies developed for handling extant challenges to modus ponens and modus tollens fail for it. It diagnoses the apparent source of the counterexample: there are bona fide instances of modus ponens that *fail to represent deductively reasonable modes of reasoning*.

The diagnosis might seem trivial—what else, after all, could a counterexample to modus ponens consist in?[1]—but, appropriately understood, it suggests something like a *method* for generating a family of counterexamples to modus ponens. We observe that a family of uncertainty-reducing "Dominance" principles must implicitly forbid deducing a conclusion (that would otherwise be sanctioned by modus ponens) when certain conditions on a background object (e.g., a background decision problem or representation of relevant information) are not met. There seems to be, in other words, a more general phenomenon at work—something that readers should bear in mind before attempting to develop any sort of one-off treatment of the paper's central example.

While it is not the primary aim of this paper to develop an understanding of this phenomenon, it will informally outline some possibilities. On the option I defend, one cannot generally choose a logic—a schematic representation of good deductive reasoning—independent of substantive normative considerations (in particular, considerations bearing on how to reason under conditions of subjective uncertainty); we can choose a logic that validates modus ponens, provided that certain "irrational" or "unreasonable" ways of organizing information are ruled out as possible relata of the consequence relation for natural language. While this way of thinking about the logic of natural language has some surprising features, it allows us to isolate a sense in which modus ponens is a rule of good deductive reasoning, and has other explanatory attractions besides.

1. Kolodny & MacFarlane (2010) attempt to adduce a counterexample to modus ponens that is not of this form: reasoning by modus ponens is, on their account, always deductively reasonable, though modus ponens is not, on their account, always truth-preserving. For a powerful critique of this mode of argument against modus ponens, see Bledin (2014).

2 The Counterexample

You (and I) love to smoke, but would hate to get cancer. You prefer the outcome in which you smoke but do not get cancer to the outcome in which you do not smoke and do not get cancer; you prefer the latter outcome to the outcome in which you smoke and do get cancer; worst of all outcomes is the one in which you do not smoke and get cancer. (This is the situation for most actual smokers.)

	CANCER	¬CANCER
SMOKE	☹	☺☺
DON'T	☹☹	☺

In this general sort of scenario, it is (obviously) better not to smoke: smoking greatly increases your risk of developing cancer, and this is a risk it would be wise to avoid.

Note that (1a) and (1b) are true in this scenario: they respectively summarize the information in the CANCER and ¬CANCER columns of the decision table. (1c) is true as well, and follows from (1a) and (1b). But, by application of modus ponens, we arrive at something unacceptable: (1d).[2]

(1)
   a. If you get cancer it's better to smoke.
   b. If you don't it's better to smoke.
   c. So, if you get cancer or you don't, it's better to smoke.
   d. #So, it's better to smoke.

	$p \Rightarrow \triangle q$	
	$\neg p \Rightarrow \triangle q$	
	$(p \vee \neg p) \Rightarrow \triangle q$	CA
	$(p \vee \neg p)$	$\top$
#	$\triangle q$	$\Rightarrow$E

Conditionals (1a) and (1b) are true, and I do not recommend trying to wiggle your way out of the problem by denying them. But, if you are doubtful, you may apply the Ramsey Test. Suppose you have preferences as described in the above decision table. Suppose that you get cancer. Is it better to smoke or not? Clearly it is better to smoke. Bearing this in mind, (1a) should command immediate assent. Suppose that you do not get cancer. Is it better to smoke or not? Clearly it is better to smoke. Bearing this in mind, (1b) should command immediate assent. And (1c)—more precisely, an intended reading of (1c)—is true, and follows immediately from (1a) and (1b).

Decision theorists, please hold your fire. I suspect I know what you want to say—and, of course, I agree: the decision problem I have described is ill-formed, by any metric of well-formedness you might choose. That, alas, does not alter the logical facts. If modus ponens is valid and (1c) is true, it's better to smoke.

2. A note on notation. I use '⇒' to abbreviate the indicative conditional. All of the counterxamples to modus ponens and modus tollens discussed here involve the use of modal, preferential, or probabilistic operators. Throughout I use a '△' to abbreviate such operators.

Linguists, please hold your fire. I know that the logical representations I have chosen for (1a)–(1c) ignore matters like tense and aspect—matters of special importance for evaluating deliberative conditionals like these. This is deliberate. Unless the representation of tense and aspect renders the move from (1c) to (1d) a non-instance of modus ponens or renders $(p \vee \neg p)$ a non-instance of $\top$—and I cannot see how to make out a story on which either would be the case[3]—tense and aspect are irrelevant in the analysis of the problem, and we do well to abstract away from the complications would accompany their introduction.

3 Logical Preliminaries

The reasoning in (1) makes use of two inference rules: (CA), a standard principle of conditional logic, and modus ponens (henceforth $\Rightarrow$E).

$$\begin{array}{|ll} p \Rightarrow r & \\ q \Rightarrow r & \\ (p \vee q) \Rightarrow r & \quad \text{CA} \end{array}$$

(CA) is assumed by the preponderance of theorists (including both Lewis and Stalnaker), but is rejected in certain context-shifting/dynamic frameworks (a point to which we will return below). I stress, however, that the matter of its general validity is irrelevant here. The use of (CA) is well-founded *in (1)*: given (1a) and (1b), (1c) is reliably heard as true; on the intended reading, it *summarizes* the information contained in (1a) and (1b).[4] Nor does anything rest on whether (1c) has a false reading (as it surely does). It seems clearly to have a true reading, on which it is summarizing the information expressed in (1a) and (1b). That reading—hereafter the "Intended Reading"—is the reading at issue here.

Notice that, given ($\Rightarrow$E), (CA) allows us to derive a standard elimination rule for disjunction, abbreviated ($\vee$E).

$$\begin{array}{|ll} p \vee q & \\ p \Rightarrow r & \\ q \Rightarrow r & \\ r & \quad \vee\text{E} \end{array} \qquad\qquad \begin{array}{|ll} p \vee q & \\ p \Rightarrow r & \\ q \Rightarrow r & \\ (p \vee q) \Rightarrow r & \quad \text{CA} \\ r & \quad \Rightarrow\text{E} \end{array}$$

3. If you deny Excluded Middle for future contingents, note that the formulation of counterexamples to ($\Rightarrow$E) of the form in (1) need not rely on the use of future contingents. See example (25). If you think $(p \vee \neg p)$ is ambiguous between an alternative (or inquisitive) semantic value and a classical Boolean semantic value, see §6.1.

4. Though I do not want to rest anything on this claim, the following is a plausible constraint on the logic of the conditional: $(p \Rightarrow r) \wedge (q \Rightarrow r) \dashv\vdash (p \vee q) \Rightarrow r$. The left-to-right direction encodes (CA). The right-to-left direction encodes Simplication of Disjunctive Antecedents (SDA). For more on SDA, see §6.1.

The availability of (∨E) would allow us to run an alternative argument for smoking:

(2) a. If you get cancer it's better to smoke.
b. If you don't, it's better to smoke.
c. #So it's better to smoke.

Our (very tentative) hypothesis is that both (1) and (2) rest ultimately on applications of (⇒E) that are, for an as-yet unidentified reason, illicit. We may say that (2), unlike (1), relies implicitly on (⇒E), if (as seems plausible) we regard (∨E) as a derived, rather than basic, logical rule.[5] (CA), whether or not it is a generally appropriate rule governing the logic of the conditional, appears to preserve truth in the smoking scenario: on the Intended Reading, (1c) is acceptable, given (1a) and (1b). (⇒E) is, it would seem, the only remaining culprit.

4 Stalnaker on Fatalism

Stalnaker (1975: §5) discusses a similar example of "fatalistic" reasoning.[6]

> The setting of the example is wartime Britain during an air raid. I reason as follows: "Either I will be killed in this raid or I will not be killed. Suppose that I will. Then even if I take precautions I will be killed, so any precautions I take will be ineffective. But suppose I am not going to be killed. Then I won't be killed even if I neglect all precautions; so, on this assumption, no precautions are necessary to avoid being killed. Either way, any precautions I take will be either ineffective or unnecessary, and so pointless." (280)

Stalnaker diagnoses this reasoning as follows:

> [T]he conclusion follows *validly* from the premiss, provided that the sub-arguments are *valid*. But it is not correct that the conclusion is a *reasonable inference* from the premiss, provided that the sub-arguments are *reasonable inferences* [...T]he sub-arguments are reasonable, but not valid, and this is why the argument fails [i.e., to be valid or reasonable]. (281)

I take Stalnaker to be suggesting that conditionals (3) and (4) must be *established by conditional proof* (i.e. derivation of the consequent under supposition of the antecedent).

(3) If I am killed in the air raid, then precautions are pointless.

---

5. I am suggesting, in other words , that the status of (CA) as a rule within the proof theory of conditional logic might be seen as dependent on the status of (⇒E) as a rule in that theory, so that giving up (⇒E) would undermine the status of (∨E) as a rule in that theory.

6. Thanks to [xxx] for the reminder. Dreier (2009), like Stalnaker, discusses fatalistic arguments of the form of (∨E), and concurs with me that they illustrate failures of (⇒E), on the grounds that "In some contexts [∨E] is suspect, but not here, I take it" (127–8). (Thanks to [xxx] for this reminder.) As far as I am aware, Dreier and I are the only two to diagnose fatalistic reasoning in this way (though of course Kolodny & MacFarlane (2010) diagnose apparent failures of Constructive Dilemma of the sort apparently attested in their Miner Case as failures of (⇒E) as well). I take it, however, that the move that Dreier makes in diagnosing this case, though correct, is not presently dialectically available, for two reasons. First, (∨E) *is* generally regarded as the suspect principle in cases of this general type (e.g., the Miner Case) (Willer 2012; Yalcin 2012; Bledin 2014). Second, if indeed the logical status of (∨E) depends on the logical status of (⇒E) (as suggested in §3), giving up (⇒E) means giving up (∨E), even for arguments in which (⇒E) is not explicitly invoked, like (2). Indeed, for this reason, and in view of cases like (2), we should say that (∨E) is *not* a generally valid rule of inference.

(4) If I am not killed in the air raid, then precautions are pointless.

Here, then, is a natural representation of the reasoning at issue:

I am killed in the air raid or I am not.

Suppose I am killed in the air raid.

Then, precautions are pointless.

So, if I am killed in the air raid, then precautions are pointless.

Suppose I am not killed in the air raid.

Then, precautions are pointless.

So, if I am not killed in the air raid, then precautions are pointless.

So, precautions are pointless.

The difficulty with this reasoning, Stalnaker observes, is that, in each sub-derivation, the consequent cannot strictly be *derived* under the relevant supposition: it is certainly not a *logical* consequence of your being killed in the air raid that precautions are pointless. It is, Stalnaker allows, *reasonable to infer* that precautions are pointless, on the supposition that you are killed in the air raid. But ($\vee$E) is not a rule that characterizes reasonable inference, as this example well-illustrates: from (3) and (4) it is not reasonable to infer that precautions are pointless. The fatalist's "proof" is, therefore, both *invalid* (since the sub-arguments establishing (3) and (4) are invalid) and *unreasonable* (since ($\vee$E) is not a rule that characterizes reasonable inference). It is a failure, twice over.

Alas, Stalnaker's diagnosis of fatalistic arguments does not help with (1). Notice that the conditionals in (1) were provided directly by context, not established by conditional proof: (1a) and (1b) are, very simply, *true*.[7] Since Stalnaker endorses (CA) and ($\Rightarrow$E), (1) is a valid, as well as sound, argument, by his own lights.

7. This is an important contrast with the "proof" of fatalism. Notice that (3) and (4) are not easily heard as true. Indeed, they are often rejected, on the grounds that reasonable precautions always have a point, namely, reduction of risk (see, e.g., Ayer (1964) and the statement from Hospers (1967) quoted at Buller (1995: 112)). Since the truth of (3) and (4) cannot be established directly (i.e., by appeal to semantic intuition), an indirect (e.g., logical) argument for their truth is required; but the obvious indirect argument (by conditional proof) is a logical failure, or so Stalnaker argues.

A reader might reasonably wonder if I my own argument is caught up in something like this dialectic: have I not also argued for (1a) and (1b) using something like conditional proof? In reply: I have, in fact, been careful to avoid this. Yes, I have tried to cajole the reader into accessing the relevant readings of (1a) and (1b) by suggesting the Ramsey Test. But this is not strictly intended as an *argument*, indirect or otherwise, in favor of accepting (1a) and (1b); it is rather an *instruction* for how to access the intended—and clearly true (I claim)—readings of (1a) and (1b), i.e., via application of the Ramsey Test.

A reader might also reasonably wonder about the prospects for reviving the strategy of Ayer and Hospers, i.e., about denying that (1a) and (1b) are true; perhaps it is *always* better to avoid the risk of developing cancer associated with smoking, in which case it is *always* better not to smoke, in which case, no conditional of the form 'if ϕ, then it is better to smoke' could be regarded as true. This is akin to the objection from Subjectivism discussed (and dispatched) in Kolodny & MacFarlane (2010). As they note, a clear difficulty with Subjectivism is empirical: it is a serious cost to render conditionals like (1a) and (1b) false, when a preponderance of competent speakers seem able to access a true reading. See Kolodny & MacFarlane (2010: §1.2) for another argument against Subjectivism.

5 Yalcin's Counterexample

Yalcin (2012) describes a counterexample to modus tollens (MT).

$$\begin{array}{|ll} p \Rightarrow r & \\ \neg r & \\ \neg p & \text{MT} \end{array}$$

There is a bag of marbles, varying in color and size as follows, from which one is drawn and concealed from you.

	BLUE	RED
BIG	10	30
SMALL	50	10

Sentences (5a) and (5b) are true in this scenario. But, by application of (MT), we arrive at something unacceptable: (5c).

(5) a. If it's big it's likely red.
b. It's not likely red.
c. #So, it's not big.

$$\begin{array}{r|ll} & p \Rightarrow \triangle q & \\ & \neg\triangle q & \\ \# & \neg p & \text{MT} \end{array}$$

This section describes two versions of a response to Yalcin's apparent counterexample to (MT). On both, the questionable inference is actually not an instance of (MT). On one, this is because the logical forms suggested for the premises of (5) are incorrect (because underdescribed). On another, this is because (MT) is claimed to be a rule governing preservation of truth with respect to a single context, and in (5) there is an illicit context-shift. I am not ultimately interested in whether these replies blunt the force of Yalcin's counterexample (though the second reply is, I will suggest, quite powerful). My main goal in this section is to establish that they do not extend to the counterexample to (⇒E) presented above.[8]

5.1 *Context-Sensitivity*

There are various ways of advancing an objection to Yalcin on grounds of context-sensitivity. Here I will present one (inspired by the analogy of modals to pronouns in Stojnić forthcoming); the specific details will be unimportant to the conclusion I draw.

5.1.1 Domain Restrictions

It is commonly thought that 'if'-clauses semantically function to introduce *domain restrictions* for downstream quantificational expressions (in particular: modal, preferen-

8. This section is, of necessity, technical. If you are satisfied of the truth of (1c) and uninterested in formal detail (or in issues surrounding the viability of Yalcin's apparent counterexample to (MT)), you can skip ahead to the next section. In summary, this section argues that none of the strategies for dealing with Yalcin's apparent counterexample to (MT) supply any reason for rejecting (1c).

tial, probabilistic, or epistemic operators).[9] Letting $\triangle$ be any such operator, the idea is that the relevant operators have restrictor and scope arguments:

$$\triangle_{\text{RESTRICTOR}}\text{SCOPE}$$

The antecedent of the conditional supplies the operator's restriction argument.

$$p \Rightarrow \triangle q \ := \ p \Rightarrow \triangle_p q$$

Once we disambiguate (5) along these lines we see it is not an instance of (MT):

$$\begin{array}{r|ll} & p \Rightarrow \triangle_p q & \\ & \neg\triangle_\top q & \\ \# & \neg p & ?? \end{array}$$

A similar treatment could be developed for (2): once we represent the relevant operator's domain restriction, we see that the inference is not an instance of ($\vee$E):

$$\begin{array}{r|ll} & p \vee \neg p & \\ & p \Rightarrow \triangle_p q & \\ & \neg p \Rightarrow \triangle_{\neg p} q & \\ \# & \triangle_\top q & ?? \end{array}$$

Similarly, perhaps, (1) can be understood as invalid by representing the reasoning as follows.[10] (RCM is a generally accepted principle of conditional logic, according to which the conditional is right upward monotone.)

$$\begin{array}{r|ll} & p \Rightarrow \triangle_p q & \\ & p \Rightarrow \triangle_p q \vee \triangle_{\neg p} q & \text{RCM} \\ & \neg p \Rightarrow \triangle_{\neg p} q & \\ & \neg p \Rightarrow \triangle_p q \vee \triangle_{\neg p} q & \text{RCM} \\ & (p \vee \neg p) \Rightarrow \triangle_p q \vee \triangle_{\neg p} q & \text{CA} \\ & \triangle_p q \vee \triangle_{\neg p} q & \Rightarrow\text{E} \\ \# & \triangle q & ?? \end{array}$$

9. The idea that 'if'-clauses are restrictors for downstream quantifiers owes originally to Kratzer (1981, 1991). The implementation here is not Kratzer's, since she treats the semantics of the 'if'-clause as *exhausted* by its restrictive function. I consider Kratzer's analysis of the relevant conditionals below.

10. An alternative representation might be offered, in which it is noted that $(p \vee \neg p) \Rightarrow \triangle_{p \vee \neg p} q$ does not follow from $p \Rightarrow \triangle_p q$ and $\neg p \Rightarrow \triangle_{\neg p} p$. This is akin to the strategy considered in §5.2.

But it seems clear that this simply misrepresents (1): a conditional of the form $(p \vee \neg p) \Rightarrow \triangle_p q \vee \triangle_{\neg p} q$ is not what is gotten by application of (CA) in (1).[11] The relevant reading of (1c), rather, appears to be given by:

$$(p \vee \neg p) \Rightarrow \triangle_{p \vee \neg p} q$$

If that is right, then in order to understand the derivation of (1c) as an application of (CA) (as it in fact seems to be) the argument must be rendered as follows:

$$\begin{array}{l|ll} & p \Rightarrow \triangle_{p\vee\neg p} q & \\ & \neg p \Rightarrow \triangle_{p\vee\neg p} q & \\ & (p \vee \neg p) \Rightarrow \triangle_{p\vee\neg p} q & \text{CA} \\ & (p \vee \neg p) & \top \\ & \triangle_{p\vee\neg p} q & \Rightarrow\text{E} \\ \# & \triangle q & ?? \end{array}$$

One difficulty[12] with this analysis is that, if modus ponens is indeed valid for indicative conditionals in natural language, we would expect $\triangle q$ to follow from $\triangle_{p\vee\neg p} q$: in other words, commitment to modus ponens for the indicative conditional would seem to bring in tow commitment to the principle that $\triangle q$ follows validly from $\triangle_{p\vee\neg p} q$. The next section will elaborate on this rough case.

5.1.2 Difficulties with Domain Restriction

Note first that, in the context that I have described, there is the strong sense that the restricted operators—(6) and (7)—are simply *true*.

(6) It is better to smoke, given CANCER. $\triangle_p q$

(7) It is better to smoke, given ¬CANCER. $\triangle_{\neg p} q$

That's hardly surprising—indeed, (6) and (7) are *precisely* the logical forms that Kratzer (1981, 1991) proposes for the conditionals (1a) and (1b): the compositional semantic function of the 'if'-clause is, on Kratzer's widely accepted account, *exhausted* by its restriction of some downstream quantifier.

This in mind, consider the following schematic derivation:

(8)
$$\begin{array}{|l} \triangle_p q \\ \triangle_{\neg p} q \\ \triangle_{p\vee\neg p} q \end{array}$$

11. Perhaps the consequent of (1c) is to be given a "sloppy" rather than "strict" reading, i.e., is to be rendered as $\triangle_p q \vee \triangle_{\neg p} q$? If the relevant reading were the sloppy reading, we would not blanch at drawing the conclusion that it is better to smoke (on the sloppy reading). And yet we do.

12. Another: there is no reason to suppose the logical forms of (1a) and (1b) are given by $p \Rightarrow \triangle_{p\vee\neg p} q$ and $\neg p \Rightarrow \triangle_{p\vee\neg p} q$. These logical forms are, it is true, the ones required to make the argument thus-represented go. But that is not a principled reason for thinking that the conditionals in question have these logical forms.

This derivation appears to preserve truth in context for (1): there is the strong sense that the following restricted operator is simply true.

(9) It is better to smoke, given CANCER ∨ ¬CANCER. $\triangle_{p\vee\neg p}q$

This is expected, from the vantage of the Kratzer analysis: (9) just is the logical form that Kratzer assigns (by default) to (1c). To put it a bit differently: given Kratzer's analysis, we may read the derivation in (8) as an apparently well-grounded instance of (CA).[13]

The problem is this. If restricted operators are conditional in interpretation—as they surely *are*—and (⇒E) is valid for sentences of natural language with a conditional interpretation, we can derive the problematic conclusion, i.e., that it is better to smoke, from (9), together with the fact that CANCER ∨ ¬CANCER is a tautology.

It would be natural to object to this reasoning, on the following grounds. There are model-theoretic reasons for thinking (⇒E) is clearly *not* valid for, e.g., a deontic-conditional or conditional-obligation reading of $\triangle_p q$ (≈it ought to be that q, given p). Weak Centering is the model-theoretic principle standardly used (in, e.g., the similarity-based semantics of Stalnaker and Lewis) to enforce the validity of (⇒E); it says that if w is a p-world, then w is among the closest-to-w p-worlds. But the domain determined by $\triangle_p$ is *not generally weakly centered*: if p is true at w it does not follow that w is in the domain of quantification for $\triangle_p$ at w: w, if a p-world, is not generally among the deontically-best-from-w (≈ideal-from-w) p-worlds.[14]

In reply, it suffices to note that (⇒E) does seem to be a principle governing the *vast majority* of restricted operators. One example, easily multiplied:

(10) a. You should turn off the A/C, given that the window is open.
b. The window is open
c. So, you should turn off the A/C

This is precisely what one would expect, if (i) (⇒E) is valid, (ii) Kratzer is correct in thinking that (10a) is paraphrasable with a vanilla indicative conditional. If (⇒E) is valid for natural language indicatives, there will *have to be* an understanding of its validity on which an inference like (10) can be rendered valid.

Such an understanding is not difficult to identify. Indeed, it has been independently suggested as a way of rescuing modus ponens from the challenge developed in Kolodny & MacFarlane (2010) (see esp. Willer 2012; Yalcin 2012; Bledin 2014). On this understanding of validity, the appropriate interpretation of logical consequence is *dynamic* or *epistemic* in nature. Valid arguments—what Bledin (2014) dubs "good deductive inferences"—are roughly, on this understanding, *knowledge-* or *information-preserving*; they do not generally track preservation of truth at a world of evaluation, since preservation of truth at a world of evaluation is not—as example (10) well-illustrates—required for good deductive inference.

---

13. It corresponds to the following (quite reasonable) model-theoretic constraint: the domain for $\triangle_{p\vee\neg p}$ is a subset of the domain for $\triangle_p$ unioned with the domain for $\triangle_{\neg p}$. On the reasonability of this constraint—a semantic analogue of the principle of Independence of Irrelevant Alternatives found in the social choice literature—see [xxx].

14. Weak centering is in fact beside the point here: the point of a modus ponens rule applied to $\triangle_p q$ is not to be able to detach $\underline{q}$, given p; it is rather to attach $\overline{\triangle q}$, given p. Nevertheless, the objection is correct in noting that, on an understanding of validity on which it amounts to preservation of truth at a world of evaluation, the truth of $\triangle_p q$ and p at w does not generally guarantee the truth of $\triangle q$ at w, and so the corresponding inference is invalid. For a countermodel, see Charlow (2013b: 2298).

A modus ponens rule for restricted operators may, then, be stated as follows:

$$\begin{array}{|ll} \triangle_p q & \\ p & \\ \triangle q & \Rightarrow\text{EO} \end{array}$$

Application of this rule will tend strongly—though, if I am right, imperfectly—to correspond to good deductive inference. The Bledin (2014) understanding of validity explains why: updating on the information expressed by the premises generally makes available the information expressed by the conclusion. This is, Bledin and I agree, the appropriate understanding of the relation of logical consequence for natural language (and for a ($\Rightarrow$E) rule for indicative conditionals—and sentences equivalent to indicative conditionals—in natural language). Most attractively, it rescues ($\Rightarrow$E) from the counterexamples of Kolodny & MacFarlane (2010), laying the blame instead with the elimination rule ($\vee$E). (As Bledin notes, information-preservation, like Stalnakerian Reasonable Inference, is not closed under a rule like ($\vee$E).)

Nevertheless, even by the lights of the most plausible understanding of the relation of logical consequence for natural language—i.e., Bledin's—modus ponens should apparently not be treated as generally valid. As example (9) well-illustrates, the following inference seems simply to fail.

$$\begin{array}{r|ll} & \triangle_{p\vee\neg p} q & \\ & p \vee \neg p & \top \\ \# & \triangle q & \Rightarrow\text{EO} \end{array}$$

The inference fails because it is, on the face of things, a *terrible* deductive inference to draw. The truth or acceptability of (9) simply is not sufficient to guarantee that it is better to smoke.[15]

5.2 *Context-Sensitivity Via Context-Shifting*

Whether or not context-sensitivity is the right response to Yalcin, it is the wrong response to our counterexample to modus ponens. A different, but related, strategy—pursued in Gillies (2009, 2010)—assigns the relevant operators their expected logical forms, but gives the conditional a special (i) information-sensitive, (ii) context-shifty interpretation.[16] To preview: on Gillies' analysis, in contrast with the context-sensitive analysis sketched in the prior section, (CA) is *not* generally valid.

I will present Gillies' analysis as he does, using epistemic modals (the epistemic necessity modal will be abbreviated '$\Box$', the epistemic possibility modal as '$\Diamond$'). We begin

15. Some readers may be scratching their heads: do I really mean to deny that $\triangle_{p\vee\neg p} q \vdash \triangle_{\top} q$? Yes, I do. More precisely, the data under consideration seems very strongly to suggest that the restrictor environment of quantificational operators (as well as the indicative conditional's 'if'-clause) is *hyperintensional*: it does not admit inter-substitution of classical logical equivalents. More later.

16. I will not discuss dynamic accounts separately, since the details are not sufficiently different to matter: dynamic accounts, like Gillies' account, tend to avoid the problematic conclusion of Kolodny & MacFarlane (2010) by voiding (CA) (see, e.g., Willer 2012).

with an example to make things concrete (Gillies 2010: 13). Suppose we know there is exactly one red or yellow marble concealed in an opaque box. Then:

(11) a. RED might be in the box and YELLOW might be in the box.
b. If it's not YELLOW, it must be RED.
c. If it's not RED, it must be YELLOW.

Gillies proves that the sentences in (11) must be (contrary to fact) *inconsistent*, if we assign them their obvious logical forms (i.e., treat the epistemic necessity modals in (11b) and (11c) as taking narrow scope in their respective conditionals), assume Weak Centering for the conditional, and apply an otherwise standard semantics for conditionals and epistemic modals. Here are the relevant formal details:

Definition 1. *An information state i. is a function $i : W \mapsto \wp(W)$ (where $w \in i_w$)*

Definition 2. *A selection function $f.(\cdot)$ is a function $f : W \mapsto (\wp(W) \mapsto \wp(W))$, where $f_w(p) \subseteq p$.*

Definition 3. *A state σ is a pair $\langle i, f \rangle$, with i an information state, f a selection function, with f weakly centered: $w \in p \rightarrow w \in f_w(p)$*

Definition 4. *Let $\sigma = \langle i, f \rangle$, and let $[\![p]\!]^\sigma = \{w : [\![p]\!]^{\sigma,w} = 1\}$. Then:*

i. $[\![p \Rightarrow q]\!]^{\sigma,w} = 1$ *iff* $f_w(i_w \cap [\![p]\!]^\sigma) \subseteq [\![q]\!]^\sigma$
ii. $[\![\Box_p q]\!]^{\sigma,w} = 1$ *iff* $i_w \cap [\![p]\!]^\sigma \subseteq [\![q]\!]^\sigma$
iii. $[\![\Diamond_p q]\!]^{\sigma,w} = 1$ *iff* $i_w \cap [\![p]\!]^\sigma \cap [\![q]\!]^\sigma \neq \emptyset$

The inconsistency proof is straightforward (Gillies 2010: 18). Here is a sketch. Suppose the sentences of (11) are true at $\sigma = \langle i, f \rangle$ and w. There are two possibilities: w is a $\neg$YELLOW-world, or w is a $\neg$RED-world. If w is a $\neg$YELLOW-world, then by weak centering w is a $\Box$RED-world, hence every world compatible with i_w is a RED-world. If w is a $\neg$RED-world, then by weak centering w is a $\Box$YELLOW-world, hence every world compatible with i_w is a YELLOW-world. So either no YELLOW-worlds or no RED-worlds are compatible with i_w. Either way, (11a) is false at σ and w.

Gillies' proposed solution is elegantly simple: to revise the semantics for the conditional in Definition 4(i). For Gillies, an indicative antecedent has a dual compositional function: it introduces a (universal) quantifier over worlds selected by applying the selection to a range of possibilities compatible with the antecedent, while also *shifting the context of interpretation* for context-sensitive material in the consequent (to a context that incorporates the information expressed by the antecedent).

Definition 5. *The update of $\sigma = \langle i, f \rangle$ with p, notation $\sigma[p]$, is $\langle \lambda w. i_w \cap [\![p]\!]^\sigma, f \rangle$*

Definition 6. *Let $\sigma = \langle i, f \rangle$. Then $[\![p \Rightarrow q]\!]^{\sigma,w} = 1$ iff $f_w(i_w \cap [\![p]\!]^\sigma) \subseteq [\![p]\!]^{\sigma[p]}$*

Here are some attractions of this analysis that are worth noting. First, it avoids predicting the sentences in (11) inconsistent: if $\neg$YELLOW is true at σ and w, then, by weak centering, it follows that $\Box$RED is true at $\sigma[\neg$YELLOW$]$ and w, but does not generally follow that $\Box$RED is true at σ and w. We may not, then, draw the conclusion that every world compatible with i_w is a RED-world. This is sufficient to disrupt the proof of inconsistency sketched above.

Second, like the context-sensitivity strategy, it has the resources to blunt the force of Yalcin's counterexample to (MT) (though I doubt Gillies would endorse the strategy I am about to describe). Gillies' semantics naturally suggests the following understanding of logical consequence: logical consequence is preservation of truth *in the absence of illicit context-shift*. 'Illicit' context shift occurs, roughly, when, in the course of an argument, a *single sentence* is evaluated as false relative to σ and w, and true relative to σ' ($\neq \sigma$) and w. According to Gillies' semantics, in Yalcin's counterexample to (MT), a sentence of the form $\triangle q$ is evaluated as true relative to an *enriched* information state—an information state enriched with the information expressed by the antecedent of (5a). Meanwhile $\triangle q$ is evaluated as false relative to an unenriched information state—the global information state appropriate to the distribution of marbles Yalcin describes. There is *no antecedent reason* to formulate the notion of logical consequence in a way such that inference-patterns superficially in the mold of (MT) that *depend on this sort of context-shift* would be counted as valid (a point also well-made in Stojnić forthcoming).

On this strategy, instances of (MT) should be individuated *semantically* rather than *syntactically*, so that only inferences whose semantic contents instantiate the following set-theoretically valid "inference schema"—that is to say, inferences that *keep proper track of the context against which q is evaluated for truth*—count as bona fide instances of (MT):

$$\begin{array}{ll} [\![p \Rightarrow q]\!]^{\sigma} & = \lambda w.f_w(i_w \cap [\![p]\!]^{\sigma}) \subseteq [\![q]\!]^{\sigma[p]} \\ [\![\neg q]\!]^{\sigma[p]} & \\ [\![\neg p]\!]^{\sigma} & \end{array}$$

Proof. Let w satisfy $\lambda w.f_w(i_w \cap [\![p]\!]^{\sigma}) \subseteq [\![q]\!]^{\sigma[p]}$ and $[\![\neg q]\!]^{\sigma[p]}$. Hence, $w \notin [\![q]\!]^{\sigma[p]}$. Hence, $w \notin f_w(i_w \cap [\![p]\!]^{\sigma})$. Given weak centering, $w \notin i_w \cap [\![p]\!]^{\sigma}$, hence (since $w \in i_w$) $w \notin [\![p]\!]^{\sigma}$, hence $w \in [\![\neg p]\!]^{\sigma}$. $\Box$

Inferences whose semantic contents instantiate this set-theoretic schema are valid instances of (MT). On Gillies' account, however, Yalcin's counterexample to (MT) instantiates the following *degraded* (because set-theoretically invalid) schema:

$$\begin{array}{lll} & [\![p \Rightarrow q]\!]^{\sigma} & = \lambda w.f_w(i_w \cap [\![p]\!]^{\sigma}) \subseteq [\![q]\!]^{\sigma[p]} \\ & [\![\neg q]\!]^{\sigma} & \\ \# & [\![\neg p]\!]^{\sigma} & \end{array}$$

The upshot is (i) a principled and relatively compelling explanation of why Yalcin's (5) is invalid, (ii) an independent and principled criterion according to which Yalcin's (5) is not, after all, an instance of (MT).

Now let us return to our counterexample to ($\Rightarrow$E), (1). Is this response to Yalcin adaptable to it? Notice first that, unlike the context-sensitive strategy, the context-shifting strategy yields a faithful representation of (1): the context-shifting strategy does not achieve its predictions by meddling with logical form. Notice further that (CA) fails on the context-shifting strategy. Suppose (11b) and (11c) are true. Then:

(12) a. $f_w(i_w \cap [\![\neg\textsc{yellow}]\!]^\sigma) \subseteq [\![\Box_\top \textsc{red}]\!]^{\sigma[\neg\textsc{yellow}]}$
b. $f_w(i_w \cap [\![\neg\textsc{red}]\!]^\sigma) \subseteq [\![\Box_\top \textsc{yellow}]\!]^{\sigma[\neg\textsc{red}]}$ Hence:

(13) a. $f_w(i_w \cap [\![\neg\textsc{yellow}]\!]^\sigma) \subseteq [\![\Box_\top \textsc{red}]\!]^{\sigma[\neg\textsc{yellow}]} \cup [\![\Box_\top \textsc{yellow}]\!]^{\sigma[\neg\textsc{red}]}$
b. $f_w(i_w \cap [\![\neg\textsc{red}]\!]^\sigma) \subseteq [\![\Box_\top \textsc{red}]\!]^{\sigma[\neg\textsc{yellow}]} \cup [\![\Box_\top \textsc{yellow}]\!]^{\sigma[\neg\textsc{red}]}$

It obviously will not follow from this that the following conditions hold. (Note that $i_w \cap [\![\neg\textsc{yellow} \vee \neg\textsc{red}]\!] = i_w$, hence that $\sigma[\neg\textsc{yellow} \vee \neg\textsc{red}] = \sigma$.)

(14) a. $f_w(i_w \cap [\![\neg\textsc{yellow} \vee \neg\textsc{red}]\!]) \subseteq [\![\Box_\top \textsc{red}]\!]^\sigma \cup [\![\Box_\top \textsc{yellow}]\!]^\sigma$
b. $[\![(\neg\textsc{yellow} \vee \neg\textsc{red}) \Rightarrow (\Box_\top \textsc{red} \vee \Box_\top \textsc{yellow})]\!]^{\sigma,w} = 1$

Something similar might be said to explain the badness of inference (1) (though I will not go through the details here): it relies on a specious application of (CA).

A semantics in this vein appears to offer the best hope for handling Yalcin's counterexample to (MT), and for handling our own counterexample to (⇒E). Gillies' is an elegant account, and allows us to define a correspondingly elegant "restriction" of (MT), on what appear to be independently motivated semantic grounds. Because it invalidates (CA) in certain cases, it is able to diagnose (1) as a specious application of (CA). I do not want to underplay the attractions of this type of account: it really does seem to me that it is the best hope for rescuing (MT) and (⇒E) from the challenges described above.

Still, there are difficulties. As before, the suspicion is that restricted operators of the following forms are true in (1): $\triangle_p q$ and $\triangle_{\neg p} q$. Then, since Gillies does not treat the restriction environment of *quantificational (e.g., modal) operators* as context-shifters, we can once again run the derivation in (8). Apply (⇒E)—more precisely, the understanding of (⇒E) appropriate to restricted operators—and we're back to square one. This response is not, however, totally satisfactory: perhaps the restriction environment of a quantificational operator *does* behave as a context-shifter (even if Gillies does not treat the environment in this way).[17]

A better (and very simple) response is to note that a semantics which invalidates (CA) cannot explain the appeal of (1c). (1) goes wrong in the move from (1c) to (1d), not in the move from (1a) and (1b) to (1c)! (I once again stress that I am not claiming that all applications of (CA) are good, only that (CA) is apparently well-applied in the move from (1a) and (1b) to (1c).)

The simplest (and most definitive) response is to note that (CA) is, ultimately, beside the point. We have presented (1c) as being "derived" from (1a) and (1b) by application of (CA). But this was an inessential feature of the presentation. (1c) is an obviously (and therefore directly) acceptable way of *summarizing* the information contained in the relevant decision table; its truth is established *directly* by the relevant context, not via inference.[18] The truth of (1c), together with the tautology $p \vee \neg p$, does not warrant inferring (1d). This is a puzzling state of affairs, no doubt. It nevertheless seems to me the only way of reading of the data.

17. For an account on which restrictors are context-shifters, see Charlow (2013b).
18. Note that Gillies' semantics renders it false, if the basic context (state) of evaluation provides a preference ordering according to which not smoking is better than smoking.

6 Diagnosis

This section will argue that the failure of (⇒E) is—in at least a range of cases—to be explained with a *normative*, rather than properly logical, account. More specifically, it will diagnose certain ways of organizing one's cognitive situation as being, in some sense, *unreasonable* (and subsequently attribute the failures of (⇒E) described in this paper to this type of unreasonability). The structure of this argument is (or may be read as) akin to a Dutch Book: if one can reason deductively from some information to an unreasonable conclusion about what to do, the appropriate conclusion is that the information—rather than the deduction—was somehow defective.

Pursuing this sort of explanation yields some curious consequences down the line, which it would be nice to avoid having to think about... if only we could develop a properly logical account of the phenomenon. The first part of this section will argue that such an account is not in the cards.

6.1 *The Simplifying Solution*

It would be nice to be able to get by without a fancy normative story, so we will try. Let us begin by noting that $(p \vee \neg p) \Rightarrow q$ plausibly entails (by Simplification of Disjunctive Antecedents) $p \Rightarrow q$ and $\neg p \Rightarrow q$. Combine this with (CA), and we have:

$$(p \vee \neg p) \Rightarrow q \dashv\vdash (p \Rightarrow q) \wedge (\neg p \Rightarrow q)$$

We cannot—if $(\vee E)$ is an invalid rule of inference, as many (including me) think—generally derive q from the premise set

$$\{(p \Rightarrow q) \wedge (\neg p \Rightarrow q), p \vee \neg p\}$$

Unsurprisingly, q cannot generally be derived from a *logically equivalent* premise set:

$$\{(p \vee \neg p) \Rightarrow q, p \vee \neg p\}$$

While I find this assessment appealing, it turns out to be fairly peripheral to the phenomena of interest in this case. First, it appears to *concede* that indicative conditionals of the form $(p \vee \neg p) \Rightarrow q$ do not generally go in for (⇒E) (while supplying a logical explanation of this failure). That is a remarkable concession, in the context of this paper.

Readers familiar with Alternative or Inquisitive Semantics (see, e.g., Alonso-Ovalle 2006; Groenendijk & Roelofsen 2009; Ciardelli & Roelofsen 2011) may think this is too quick. It is common in this literature to distinguish between simplifying and non-simplifying representations of a conditional of surface form $(p \vee \neg p) \Rightarrow q$. On the simplifying representation, the antecedent of a conditional $(p \vee \neg p) \Rightarrow q$ denotes a set of propositional alternatives: $\{p, \neg p\}$. On the non-simplifying representation, the antecedent denotes a single proposition, derived by application of a closure operator ! (such that, if S is a set of alternative propositions, $!S = \bigcup S$) to $\{p, \neg p\}$.[19]

19. I note in passing that if I were pressed to choose a semantics of the conditional and a formal theory of logical consequence, I would opt for: (i) an explanation of simplifying readings along the lines sketched here, (ii) a semantics of the conditional on which it expresses a test on an information state conditioned on the antecedent (or pairwise-conditioned information states, in the case of disjunctive antecedents) (Veltman 1996; Gillies 2004), (iii) an informational (or "good deductive inference") understanding of $\vdash$ along the lines in Yalcin (2007, 2012); Bledin (2014) (though the possible left relata of $\vdash$ will need to be restricted in accordance with §6.5 below). None of the broadly "metalogical" issues explored in this section are illuminated by these choices, however.

The only way this reply might rescue modus ponens is if the inferences at issue are *obligatorily* represented with one of the following templates:

$$\begin{array}{r|ll}
 & \{p, \neg p\} \Rightarrow \triangle q & \text{CA} \\
 & !\{p, \neg p\} & \top \\
\# & \triangle q & ??
\end{array}
\qquad\qquad
\begin{array}{r|ll}
 & \{p, \neg p\} \Rightarrow \triangle q & \text{CA} \\
 & \{p, \neg p\} & ?? \\
\# & \triangle q & \Rightarrow\text{E}
\end{array}$$

The first thing to say is that the template on the right is surely an optional representation in the cases we have considered: one can indicate that the intended reading of the bare disjunction denotes a set of alternatives $\{p, \neg p\}$, rather than $!\{p, \neg p\}$, simply by, e.g., choosing a different pattern of focal stress.

(15) A: Are you a **student or teacher**? $?!\{s, t\}$
B: Yes/No

(16) A: Are you a **student** or **teacher**? $?\{s, t\}$
B: I'm a student/I'm a teacher [#Yes/#No]

The latter pattern of focal stress seems optional (perhaps preferred) in, e.g., (1).

(17) You **will get cancer or won't**.

(18) You'll either **get cancer** or **won't**.

If the template on the right is an optional representation, as I have argued, it will be essential to this line of reply that there is no rule of deductive inference that allows one to introduce a sentence expressing $\{p, \neg p\}$ at this step of the argument. But what could justify this sort of prohibition? In introducing $\{p, \neg p\}$, one is introducing something *informationally trivial*, but with inquisitive or alternative-presenting content (Groenendijk & Roelofsen 2009). Such a move would seem deductively licensed: in introducing $\{p, \neg p\}$, one is, roughly, deciding to raise a question (is p true, or $\neg p$?). But, first, this is a question that is surely *already salient* in the contexts I have described, and so it cannot be a distortion to represent the set of alternatives $\{p, \neg p\}$ explicitly. Further, one can always introduce a set S that partitions the relevant possibilities *without collateral logical or epistemic work*, since, when S partitions the relevant possibilities, introducing S will *introduce no new information* into the discourse. And so the introduction of $\{p, \neg p\}$ would seem to be deductively well-founded, even if such a question were not already salient in the relevant contexts.[20]

More generally, such an account, even if it succeeded in explaining why the argument in (1) fails, would fail to address other questions:

- Inferring q from $(p \vee \neg p) \Rightarrow q$ is sometimes a mistake, but *often logically impeccable* (§6.6 below). Is there any general account of what distinguishes good instances of this inference from bad?

20. This isn't to say one can go around freely expressing such contents in discourse. Questions are generally *relevant* only when they address a larger question within the discourse (see, e.g., Roberts 1996), and speakers must generally make their contributions relevant, vis-à-vis the broader goal of the discourse. This would appear to be orthogonal to the logical issues under consideration here.

- Is there any clear sense in which modus ponens is regulative of deductive reasoning? Is there any clear sense in which modus ponens fails to be regulative of deductive reasoning?

The account developed in this paper's remainder will aim to address questions like these, with an eye to a more complete explanation than the Simplifying solution provides.

6.2 *Reasoning by Dominance*

As decision theorists have no doubt noticed, (1) is a textbook case of *spurious* reasoning by Dominance. Reasoning by Dominance can be initially (but, we will see, problematically) characterized with the following principle:

> **Statewise Dominance (SD)**
>
> Consider decision problem Π. (C_i is a relevant contingency, α_j an available action, $\mathcal{U}(\alpha_j \wedge C_i)$ the payoff associated with performing α_j in C_i.)
>
> | Π | C_1 | ... | C_n |
> |---|---|---|---|
> | α_1 | $\mathcal{U}(\alpha_1 \wedge C_1)$ | ... | $\mathcal{U}(\alpha_1 \wedge C_n)$ |
> | ... | ... | ... | ... |
> | α_m | $\mathcal{U}(\alpha_m \wedge C_1)$ | ... | $\mathcal{U}(\alpha_m \wedge C_n)$ |
>
> If, in Π, α_j is preferred conditional on C_i (for each i), then α_j is unconditionally preferred in Π.

Cases of "spurious" reasoning by Dominance, like (1), show that only a restricted version of SD could be true. In cases of spurious reasoning by Dominance, an agent ignores the dependence[21] of relevant contingencies on the available actions. Restricting SD to cases in which the relevant contingencies are independent of the available actions is essential to rationalize any sort of action whose performance is, by some metric costly, but which nevertheless contributes positively to one's prospects for avoiding some greatly undesirable outcome—e.g., national defense expenditures (Jeffrey 1983: 8–9), payment of protection money (Joyce 1999: 115ff), and, indeed, giving up smoking.

In decision-theoretic contexts, this sort of restriction is often understood as a restriction of the dominance principle to "*well-formed*" decision-problems: decision problems in which relevant contingencies are *independent* of the available actions.

> **Well-Formedness**
>
> Π is well-formed only if for each i, j: C_i is independent of α_j.

This notion at hand, a restricted version of SD can be stated as follows:

21. The question of how to understand dependence, in the sense relevant for formulating a precise representation of reasoning by Dominance leads headlong into the debate between Causal and Evidential formulations of Decision Theory. Probabilistic understandings of the problematic dependence are roughly associated with "Evidential" Decision Theories, causal understandings with "Causal" Decision Theories (see esp. Gibbard & Harper 1981). More on this below.

Restricted Statewise Dominance (RSD)

If, in *well-formed* Π, α_j is preferred conditional on C_i (for each i), then α_j is unconditionally preferred in Π.

For some understanding of "well-formed"—one that excludes the decision problem supplied as contextual background for (1)—RSD will express a truth.

Here, now, is a more general way of understanding the difficulty posed by (1): for any decision problem in which the relevant contingencies are jointly exhaustive, if (⇒E) is valid, there is a strong argument that (the *unrestricted* version of) SD is a *logical truth*:

$$
\begin{array}{|ll}
(C_1 \vee \ldots \vee C_n) & \top \\
C_1 \Rightarrow Pref(\alpha) & \\
\vdots & \\
C_n \Rightarrow Pref(\alpha) & \\
(C_1 \vee \ldots \vee C_n) \Rightarrow Pref(\alpha) & \text{CA} \\
Pref(\alpha) & \Rightarrow\text{E}
\end{array}
$$

As with (1), I conjecture that the use of (CA) will be well-founded in derivations of this form (whether or not (CA) is a generally correct constraint on the logic of the indicative conditional): on the intended reading, a conditional of the form $(C_1 \vee \ldots \vee C_n) \Rightarrow Pref(\alpha)$ will again summarize the information contained in each conditional of the form $C_i \Rightarrow Pref(\alpha)$. Supposing that is right, the derivation of SD may be presented in a simplified form:

$$
\begin{array}{|ll}
(C_1 \vee \ldots \vee C_n) & \top \\
(C_1 \vee \ldots \vee C_n) \Rightarrow Pref(\alpha) & \top \\
Pref(\alpha) & \Rightarrow\text{E}
\end{array}
$$

But, of course, *no version of SD*—even the true, appropriately restricted version, but certainly the false, unrestricted version—could be a *logical truth*.

6.3 *First Pass: Restricting (⇒E)*

Here is a first pass at figuring out the problem. It seems that deductive reasoning via (⇒E) about what is unconditionally preferred can *break down*—can take one from reasonable premises to unreasonable conclusions—*when the decision problem about which one is reasoning is ill-formed*.

Is it really the decision problem that's to blame? Consider the following case. The context is the Miners scenario from Kolodny & MacFarlane (2010): ten miners are trapped, in Shaft A or Shaft B, both of which are rapidly filling with water. Blocking the shaft the miners are in will save all ten; blocking the wrong shaft will kill all ten;

blocking neither will save nine, killing one.

(19) a. If the miners are in shaft A, we should block A.
  b. If the miners are in shaft B, we should block B.
  c. So, if the miners are in shaft A or in shaft B, we should block a shaft.
  d. ??So, we should block a shaft.

The objection charges that (1) and (19) are on equal footing, so far as modus ponens is concerned. With (19), however, there is no failure of act-state independence to appeal to: the miners are where they are, independent of what we choose to do. And so the explanation of (1) being developed here will not apply to (19).

In reply: in (19), the context provides enough information to guarantee that it would be best if we blocked *a shaft*—namely, the one the miners are in. And, so, one can *access a true reading of the conclusion*. In (1), by contrast, it would *not* be best if you smoked—or, at least, there is obviously no information in the context that suffices to guarantee it would be best to smoke, on any reading of this claim. The context in (1) does not guarantee that smoking is part of the best future causally accessible to you, or that the best worlds causally accessible to you are worlds in which you smoke: smoking might give you cancer, in which case you would have done better not to smoke (in which case you shouldn't have smoked). By contrast, the context in (19) does guarantee that blocking a shaft is part of the best future causally accessible to you, as well as that the best worlds causally accessible to you are worlds in which we block a shaft. This is a rather clear point of contrast between (1) and (19)—one that would seem to be explained precisely by the failure of act-state independence in (1).

To block the reasoning at issue in (1), the theory of $\vdash$ must somehow account for the fact that, when $C_1, ..., C_n$ depend on α, deductive reasoning via ($\Rightarrow$E) about what is unconditionally preferred is *unavailable*. In other words, when $C_1, ..., C_n$ depend on α:

$$(C_1 \vee ... \vee C_n)\ ,\ (C_1 \vee ... \vee C_n) \Rightarrow Pref(\alpha) \nvdash Pref(\alpha)$$

This means restricting the use of ($\Rightarrow$E) in deductive reasoning about what is unconditionally preferred to cases in which a well-formed decision problem is linguistically represented.

This is a flat-footed way of responding to the pressures raised by (1), but it has costs (an unlovely theory of $\vdash$ being the clearest). Whatever else $\vdash$ does, it should hold between premises and conclusion in virtue of their respective *logical forms*.[22] We can safely observe that inferences of the following form *often* strike us as valid:

$$\begin{array}{|ll} (p \vee q) & \\ (p \vee q) \Rightarrow r & \\ r & \Rightarrow\text{E} \end{array}$$

On the treatment under consideration, we must identify a difference at the relevant level of logical form between inferences of this form that strike us as valid and those, like

22. This is not to say that *any* argument's logical status will supervene on its logical form—in fact, I deny that this sort of supervenience holds generally. It is rather to say that when $\Gamma \vdash \phi$ and $\Delta \nvdash \psi$, these arguments must be distinct in logical form.

(1), that do not. No such difference is likely to exist. (1) fails, apparently, *because* the decision problem in question is ill-formed. That is a *substantive* fact about the content of the decision problem that (1) represents; it is not, therefore, a difference at the relevant level of logical form.

6.4 Second Pass: Restricting $\vdash$

A less direct (but, I will argue, superior) treatment begins with a similar diagnosis: deductive reasoning via ($\Rightarrow$E) about what is unconditionally preferred works when, but only when, one begins with a *reasonable representation of the information relevant to the kind of conclusion one wishes to draw*. I will assume:

- That a decision problem is a representation of the information relevant to a conclusion about what is unconditionally preferred.
- Plausibly (if controversially), that a decision problem is a reasonable representation of the information relevant to a conclusion about what is unconditionally preferred only when it is well-formed (cf. Savage 1972; Jeffrey 1983: Ch. 1).

The treatment, simply put, is to restrict $\vdash$ so that its possible relata do not include unreasonable[23] representations of the information relevant to a conclusion about what is unconditionally preferred.[24] The job of the logic of natural language is, on this view, to characterize how someone may reason deductively to certain kinds of conclusions using a reasonable representation of the information at hand; it is *not* to characterize how someone may reason deductively to such conclusions using an unreasonable representation of the information at hand. Picturesquely: the logic of natural language is the study of how reasons *transmit*, *from* premises, *to* conclusion, in virtue of logical form. In slogan form: don't let loose a good logic on bad information (unless you're seeking more bad information).

On this sort of account, there is a *logical rationale* for "restrictions on the types of decision problems to which [decision] theory can be applied"—restrictions often derided as external or ad hoc additions to a decision theory like Savage's (see, e.g., Joyce 1999: 119). As (1) well illustrates, one can reason deductively to unreasonable conclusions if one is reasoning deductively from an ill-formed decision problem. As ill-formed decision problems are not in the domain of decision problems for which a decision theory yields a recommendation, ill-formed decision problems are not possible relata of $\vdash$ (or, therefore, $\nvdash$).

23. I will not offer a general description of what it takes for a representation of the information relevant to a conclusion to be "reasonable" (though I will be clear about my meaning when I invoke the label). It is a sufficient condition for the unreasonability of one's information that one is able to use it to reason deductively to unreasonable conclusions (recall the analogy to Dutch Book arguments suggested above).

24. Formally, this will mean that $\vdash$ must come with a context subscript $\vdash_c$, with c supplying enough information to determine whether its left relata are reasonable representations of the information relevant to their right relatum. We might further formalize the account by articulating a relation of relevance between premises and conclusion, in terms of the question to which the conclusion is an answer (cf. Roberts 1996): in the case of (1), the conclusion answers a practical question (namely, which action is unconditionally preferred?), but the premises do not comprise a reasonable representation of the information relevant to deciding such a question. The discussion in the main text will proceed largely at an informal level, suppressing issues like these.

I should be upfront: even at the maximally noncommital level of description at which I've presented it here, this account has some curious features. In particular, it makes a logical and meta-logical account of the properties of the consequence relation for natural language, $\vdash$, seem as if it will depend on *normative* theoretical choices—in particular, how to understand the notion of independence invoked by the notion of Well-Formedness, a matter of special contention between Causal and Evidential decision theorists.[25] Causal and Evidential decision theorists will, very simply, disagree about questions concerning matters of logic—in particular, questions about whether reasoning by Dominance is logically permitted in cases that divide Causal and Evidential decision theorists, like the Newcomb Problem.

I regard this as an interesting (but subtle) data point for the literature on the topic-neutrality of logic, not as an objection—Causal and Evidential decision theorists *do*, in fact, apparently disagree about the logical status of reasoning by Dominance in the Newcomb Problem. I must, however, defer discussion of dependencies of this sort, and further development of the treatment I have in mind, to another place. I will end the paper by describing some attractions of the account formulated here.[26]

6.5 $\vdash$ *Obeys* ($\Rightarrow$E)

What, precisely, does the treatment sketched above have to say about our central example, (1)? In one light, relatively little: the premises are not among the relata of $\vdash$, since they are not a reasonable way of representing the relevant decision problem; the premises are related by neither $\vdash$ nor $\nvdash$ to the conclusion. It is a bad argument, but not *because* the reasoning is bad; the class of bad arguments is more expansive the class of arguments whose reasoning is bad (i.e., the arguments whose premises are related by $\nvdash$ to their conclusions).

There is, in fact, an identifiable sense in which reasoning as in (1) is *good*: identical reasoning works well—i.e., generates reasonable conclusions—in other contexts (a point to which I will return in the next section). Indeed the reasoning in this very example can be repaired by requiring that the background decision satisfy the Well-Formedness condition. Imagine, contrary to fact, that cancer did *not* depend, in any of the relevant senses, on smoking. In such a context, the reasoning in (1) is impeccable.

(20)
a. If you get cancer it's better to smoke.
b. If you don't it's better to smoke.
c. So, if you get cancer or you don't, it's better to smoke.
d. ✓ So, it's better to smoke.

According to the account proposed above, this contrast is explained by a shift in the properties of the background decision problem: the decision problem relevant for (20) is, by stipulation, well-formed.

The reasoning in (1) stands as a counterexample to ($\Rightarrow$E), in one, mostly informal, sense: like McGee (1985)'s counterexample to modus ponens, it yields an intuitive judg-

25. I owe this formulation to [XXX].

26. A prediction I will footnote concerns the inference form in (8). If we think of restricted operators as a species of conditional, (8) is an unproblematic instance of (CA). It becomes problematic only if we are able to detach the problematic unconditional claim by $\vdash$.

ment of truth for the premises of an argument with the following syntactic parsing...

$$\phi \Rightarrow \psi,\ \phi$$

...but an intuitive judgment of falsehood for the the result of applying ($\Rightarrow$E) to these premises (on this syntactic parsing).

But, if one accepts an understanding of $\vdash$ on which its role is, roughly, to represent the transmission of reasons from premises to conclusion, we will find no argument in (1) against the following version of modus ponens:

$$\phi \Rightarrow \psi\ ,\ \phi\ \vdash\ \psi$$

Though modus ponens has instances that do not represent good deductive reasoning—i.e., instances in which premises are not related to conclusion by $\vdash$—modus ponens stands as a rule that is *regulative* of good deductive reasoning. That is not because modus ponens preserves truth in all linguistically admissible contexts (as (1) shows, it doesn't), nor because someone who knows $\phi \Rightarrow \psi$ and knows ϕ is thereby in a position to know ψ (as (1) shows, they're not). It is rather because the theory of $\vdash$ has been constrained: $\vdash$ represents the inferences reasonable agents can and cannot make from, roughly, reasonably organized information. The theory does *not*—directly, anyway[27]—tell an agent who ignores cancer's dependence on smoking how to reason about whether to smoke; the agent must ameliorate their confusion before the theory issues any direct recommendations for their reasoning.

6.6 Genuine Dominance

As just noted, reasoning of the form instantiated by (1) is often—i.e., in contexts in which a well-formed decision problem is represented—*good deductive reasoning*. Suppose a businessman is contemplating whether or not to buy some property.

> He considers the outcome of the next presidential election relevant. So, to clarify the matter to himself, he asks whether he would buy if he knew[28] that the Democratic candidate were going to win, and decides that he would. Similarly, he considers whether he would buy if he knew that the Republican candidate were going to win, and again finds that he would. Seeing that he would buy *in either event*, he decides that he should buy, even though he does not know which event obtains. (Savage 1972: 21)

We may represent the inference in question as follows:

(21) a. If the Democrat wins, it is better to buy.
b. If the Republican wins, it is better to buy.
c. So, if either event obtains, it is better to buy.
d. So, it is better to buy.

27. The theory does issue verdicts of this form: if the decision problem were well-formed, inferring the conclusion from the premises would (not) be a good deductive inference. More on this below.

28. Savage mistakenly treats the (non-epistemic) 'conditional on ϕ' construction invoked in SD as interchangeable with (epistemic) conditional constructions like 'given that ϕ is known to obtain'. For why this is mistaken, see, e.g., Weirich (1980). For why conditional preferences/utilities matter for theorizing about (conditional) deontic modality, see [xxx].

My claim is that this reasoning is deductively good *because* the decision problem determined by this context is well-formed.

It is worth remembering that the "Contextualist" strategies canvassed in §5 treat (more precisely, *attempt* to treat) this reasoning as *invalid*. Context-sensitivity (§5.1) holds that the reasoning is invalid because there is equivocation around 'better'. Context-shiftiness (§5.2) holds that the reasoning is invalid because (21c) does not strictly follow from (21a) and (21b). These are incorrect predictions (even if the strategies in question did not, as I argued, succeed in actually generating these predictions). That is to say: even if there were a strategy that successfully rendered the reasoning in (1) invalid, it would presumably thereby also render the reasoning in (21) invalid. That would be an incorrect prediction.

On the account I have outlined here, (21), in contrast to (1), is valid—i.e., the premises and conclusion are related by $\vdash$—because the context represents a well-formed decision problem. It would seem that these are the correct things to say about (1) and (21): (1) and (21) are a minimal pair; the only relevant difference is in the nature of the background decision problem. Any strategy that regards (1) and (21) as being on the same logical footing—something that seems forced, if logical footing is determined solely by logical form—will have difficulty saying them.

6.7 *Probabilistic "Dominance"*

The Statewise Dominance principle is tailor-made for generating intuitive counterexamples to modus ponens. Its theoretical function is the obviation of uncertainty: it says under what conditions one can generate an unconditional conclusion (about what it is preferred, unconditionally) from conditional premises (about what is preferred, conditionally). But it can fulfill this theoretical function only if it is restricted to well-formed decision problems (on pain of sanctioning clearly spurious instances of reasoning by Dominance). This, I argued, motivated an attendant restriction in $\vdash$, so that an argument of this form is valid only when the relevant decision problem is well-formed.

$$\begin{array}{|ll} (C_1 \vee ... \vee C_n) & \top \\ (C_1 \vee ... \vee C_n) \Rightarrow Pref(\alpha) & \top \\ Pref(\alpha) & \Rightarrow\text{E} \end{array}$$

The question thus arises: are there other uncertainty-obviating principles that give rise to the same sorts of theoretical pressures as Statewise Dominance? In fact, there are. Begin by considering the following modification of Yalcin's Marble case.

	BLUE	RED
BIG	10	11
SMALL	39	40

Here are some impeccable inferences, respectively similar in form to (1) and (2):

(22) a. If the marble is big it's likely red.
b. If the marble is small it's likely red.
c. So, if the marble is big or small, it's likely red.
d. So the marble is likely red.

(23) a. If the marble is big it's likely red.
b. If the marble is small it's likely red.
c. So the marble is likely red.

I say that (22) and (23) are impeccable. Why? A natural explanation is that the following is a theorem of the probability calculus:

> **Statewise Probabilistic Dominance (SPD)**
>
> Consider a *partition* $\{C_1, ..., C_n\}$. If p is likely to degree δ (hereafter, δ-likely) conditional on C_i (for each i), then p is unconditionally δ-likely.

Like Statewise Dominance, SPD is an uncertainty-obviating principle: it says under what conditions one can generate an unconditional conclusion (about what it is likely, unconditionally) from conditional premises (about what is likely, conditionally). Notice that SPD, like RSD, is *restricted* in application to a given partition, and for good reason. When we remove the restriction to partitions, we end up with something *false*, namely:

> **Degraded Statewise Probabilistic Dominance (DSPD)**
>
> Consider *jointly exhaustive* $\{C_1, ..., C_n\}$. If p is δ-likely conditional on C_i (for each i), then p is unconditionally δ-likely.

Unless the restriction in SPD is accompanied by an attendant restriction in $\vdash$, there will be pressure to regard any argument of the following form as valid.

$$\begin{array}{|ll} (C_1 \vee ... \vee C_n) & \top \\ (C_1 \vee ... \vee C_n) \Rightarrow \delta p & \top \\ \delta p & \Rightarrow\text{E} \end{array}$$

Given what I have argued in this paper, however, we would not expect just any argument of this form to strike us as valid: only when $C_1, ..., C_n$ form a partition will δp follow, intuitively, from $(C_1 \vee ... \vee C_n) \Rightarrow \delta p$.

This is the correct prediction! Suppose we are thinking about the likeliest trajectory for a cyclone (the only options: North, South, East) and that:

- $Pr(n) = .4$
- $Pr(s) = Pr(e) = .3$
- So: $Pr(n|n \vee s) = Pr(n|n \vee e) = 4/7$

Notice that the alternatives in $\{n \vee s, n \vee e\}$ are jointly exhaustive, but not mutually exclusive: $\{n \vee s, n \vee e\}$ is not a partition. Unless $\vdash$ is restricted in tandem with SPD, the following argument will be predicted valid, contrary to fact.

(24) a. If the cyclone moves N or S, it will likely move N.
b. If the cyclone moves N or E, it will likely move N.
c. So, the cyclone moves N or S, or N or E, it will likely move N.
d. #So, the cyclone will likely move North.

Given that $Pr(n|n \vee s) = Pr(n|n \vee e) = 4/7$, we hear conditionals (24a) and (24b) as true. We also hear (24c) as true: it summarizes the probabilistic information contained in (24a) and (24b). We know that the antecedent of (24c) is true: the alternatives in $\{n \vee s, n \vee e, n \vee w\}$ are, by stipulation, jointly exhaustive. And yet the conclusion intuitively does not follow by ($\Rightarrow$E).

Here is a different example with a similar upshot, but two new features: first, it is *backward-looking* rather than forward-looking in temporal aspect; second, it doesn't rely on the relevant alternatives being syntactically disjunctive. (Examples like this one multiply easily.) Suppose we are wondering about which marble has been drawn from an urn containing 100 marbles in the following distribution:

- 30 monochromatic red marbles (all of which are small)
- 30 monochromatic blue marbles (all of which are small)
- 40 multicolored (red and blue) marbles (all of which are big)

The alternatives in $\{\textsc{red}, \textsc{blue}\}$ ('RED' here understood to mean: is red somewhere, 'BLUE' likewise) are exhaustive, but do not form a partition. As before, then:

- $Pr(\textsc{big}) = .4$
- $Pr(\textsc{big}|\textsc{red}) = Pr(\textsc{big}|\textsc{blue}) = 4/7$

Unless $\vdash$ is restricted in tandem with SPD, the following argument will be predicted valid, contrary to fact:

(25) a. If it is red, it is likely big.
b. If it is blue, it is likely big.
c. So, if it is red OR it is blue, it is likely big.
d. #So, it is likely big.

Given the data, $\vdash$ should be restricted so that it yields a verdict of valid ($\vdash$) or invalid ($\nvdash$) only for arguments whose premises characterize "well-formed" information states—information states that form a partition of the relevant space of possibilities—to conclusions.[29] As (24) and (25) well illustrate, one can reason deductively to unreasonable conclusions if one is reasoning deductively from sentences that characterize an unpartitioned information state. As ill-formed information states are not in the domain of uncertainty-obviating dominance principles like SPD, sentences that characterize ill-formed information states are not possible relata of $\vdash$ (or, therefore, $\nvdash$).

7 The End

This paper has denied modus ponens for the indicative conditional, in one sense, while attempting to describe a sense in which modus ponens is nevertheless regulative of good deductive reasoning. On the account developed in this paper, the theory of $\vdash$ gives a theory of a kind of *ideal* deductive reasoning—an account of good deductive inference from reasonable ways of structuring the relevant information—and *modus ponens is a rule of this theory*. The theory of $\vdash$ prescinds from the question of what follows from, e.g.,

29. This is a standard restriction in the decision-theoretic contexts: the relevant contingencies in a well-formed decision problem are generally required to form a partition. It is also a standard restriction in formal models of information structure in discourse (see esp. Roberts 1996). My suggestion here is simply that these sorts of restrictions have parallel motivations in the logical domain.

information characterizing an ill-formed decision problem—as it seemingly must, since, in the realm of the "non-ideal," counterexamples to modus ponens seem to abound.

Does not such a theory radically circumscribe the subject-matter of the logic of natural language? Is a theory so-circumscribed enough of a theory to be worth having? It is not ordinarily thought that $\vdash$ is "gappy" in the way it has been portrayed in this paper; in particular, the theory of $\vdash$ is ordinarily thought to tell us when a conclusion may or may not be deduced from *any* set of syntactically well-formed sentences.

It is true that, on the account described here, for many sets of syntactically well-formed sentences Γ and syntactically well-formed conclusions ϕ, it is neither the case that $\Gamma \vdash \phi$ nor that $\Gamma \nvdash \phi$. Nevertheless, the theory does not prescind from questions like: if Γ were to constitute a reasonable way of representing the relevant information, *would* it be the case that $\Gamma \vdash \phi$? Counterfactual claims such as these remain well-grounded in the theory developed here (recall §§6.5–6.6), and may be used in order to ground a (more or less standard) account of good deductive reasoning for non-ideal ways of structuring one's information.

This paper has also denied the honorific 'valid' to something as trivial as inferring the consequent of a conditional with a tautologous antecedent. More precisely, I have denied that a quantificational operator O (be it a conditional operator, modal operator, preferential operator, or probabilistic operator) restricted to a disjunction whose disjuncts are jointly logically exhaustive allows one to draw the inference in which O is vacuously restricted (i.e., to $\top$).

The evidence described in this paper does suggest strongly that the restriction environment must be regarded as *hyperintensional*: restriction to, for instance, $p \vee \neg p$ must be treated as semantically and logically distinct from trivial or vacuous restriction (i.e., to $\top$).[30] The project of devising a logic for the conditional does not die if we relinquish the principle in question. But it will need to be recast in a hyperintensional, possibly inquisitive (cf. Ciardelli & Roelofsen 2011), light.

References

Alonso-Ovalle, Luis. 2006. Disjunction in alternative semantics. http://semanticsarchive.net/Archive/TVkY2ZlM/alonso-ovalle2006.pdf. Ph.D. Dissertation, University of Massachusetts Amherst.

Ayer, A. J. 1964. Fatalism. In *The Concept of a Person and Other Essays*, 235–268. Macmillan & Co.

Bledin, Justin. 2014. Logic informed. *Mind* 123: 277–316. doi:10.1093/mind/fzu073.

Buller, David J. 1995. On the 'standard' argument for fatalism. *Philosophical Papers* 24: 111–125. doi:10.1080/05568649509506524.

Charlow, Nate. 2013b. What we know and what to do. *Synthese* 190: 2291–2323. doi:10.1007/s11229-011-9974-9.

Ciardelli, Ivano & Floris Roelofsen. 2011. Inquisitive logic. *Journal of Philosophical Logic* 40: 55–94. doi:10.1007/s10992-010-9142-6.

Dreier, James. 2009. Practical conditionals. In David Sobel & Steven Wall (eds.) *Reasons for Action*, 116–133. Cambridge: Cambridge University Press.

Gibbard, Allan & William L. Harper. 1981. Counterfactuals and two kinds of expected utility. In W. Harper, R. Stalnaker & G. Pearce (eds.) *Conditionals, Belief, Decision, Chance and Time*, 153–190. Springer Netherlands.

30. This claim is unsurprising on an alternative-semantic treatment of disjunction (recall §6.1). A similar conclusion is drawn for counterfactual conditionals in Santorio (2016).

Gillies, Anthony S. 2004. Epistemic conditionals and conditional epistemics. *Noûs* 38: 585–616. doi:10.1111/j.0029-4624.2004.00485.x.
Gillies, Anthony S. 2009. On truth-conditions for *If* (but not quite only *If*). *The Philosophical Review* 118: 325–349. doi:10.1215/00318108-2009-002.
Gillies, Anthony S. 2010. Iffiness. *Semantics and Pragmatics* 3: 1–42. doi:10.3765/sp.3.4.
Groenendijk, Jeroen & Floris Roelofsen. 2009. Inquisitive semantics and pragmatics. http://www.illc.uva.nl/inquisitive-semantics. Ms., ILLC.
Hospers, John. 1967. *An Introduction to Philosophical Analysis*. London: Routledge.
Jeffrey, Richard. 1983. *The Logic of Decision*. Chicago: University of Chicago Press, 2nd edn.
Joyce, James. 1999. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.
Kolodny, Niko & John MacFarlane. 2010. Ifs and oughts. *Journal of Philosophy* 107: 115–143. doi:10.5840/jphil2010107310.
Kratzer, Angelika. 1981. The notional category of modality. In H. Eikmeyer & H. Rieser (eds.) *Words, Worlds, and Contexts*, 38–74. Berlin: De Gruyter.
Kratzer, Angelika. 1991. Conditionals. In A. von Stechow & D. Wunderlich (eds.) *Semantics: An International Handbook of Contemporary Research*, 651–656. Berlin: De Gruyter.
McGee, Vann. 1985. A counterexample to modus ponens. *Journal of Philosophy* 82: 462–471.
Roberts, Craige. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics. In *OSU Working Papers in Linguistics, Vol 49: Papers in Semantics*. Ohio State University. http://semanticsarchive.net/Archive/WYzOTRkO/.
Santorio, Paolo. 2016. Alternatives and truthmakers in conditional semantics. http://paolosantorio.net/ac.draft10.pdf. Ms.
Savage, Leonard J. 1972. *The Foundations of Statistics*. Dover.
Stalnaker, Robert. 1975. Indicative conditionals. *Philosophia* 5: 269–286. doi:10.1007/BF02379021.
Stojnić, Una. forthcoming. One's modus ponens. *Philosophy and Phenomenological Research* .
Veltman, Frank. 1996. Defaults in update semantics. *Journal of Philosophical Logic* 25: 221–261. doi:10.1007/BF00248150.
Weirich, Paul. 1980. Conditional utility and its place in decision theory. *The Journal of Philosophy* 77: 702–715. doi:10.5840/jphil1980771110.
Willer, Malte. 2012. A remark on iffy oughts. *Journal of Philosophy* 109: 449–461. doi:10.5840/jphil2012109719.
Yalcin, Seth. 2007. Epistemic modals. *Mind* 116: 983–1026. doi:10.1093/mind/fzm983.
Yalcin, Seth. 2012. A counterexample to modus tollens. *Journal of Philosophical Logic* 41: 1001–1024. doi:10.1007/s10992-012-9228-4.