Modus Ponens and the Logic of Decision

Nate Charlow (Draft of 12/18/2020)

1 Introduction

Some time in 2013 I came to accept that the best thing for me was to quit smoking. Though I enjoyed smoking, I didn't want to get cancer from it.[1] My reasons then were mundane, grounded in equally mundane facts about my preferences. Although I preferred the outcome in which I smoked but did not get cancer to the one in which I did not smoke and did not get cancer, I greatly preferred the latter outcome to the outcome in which I smoked and got cancer. (Worst of all possible outcomes was the one in which I did not smoke and still got cancer.) I take it as obvious that I was *right* then. Smoking greatly increases your risk of cancer, and this is a risk you should obviously avoid, if you have preferences like mine. Little did I know then, there was an argument to a very different (and, it should be stressed, very implausible) conclusion, but with a surprisingly plausible claim to logical soundness.

1.1 The Puzzle

Let's begin with this schematic (but accurate) representation of the "decision problem" I faced then. I was evaluating a possible choice between smoking and not smoking, and conditional on both me getting cancer and me not getting cancer, I preferred to smoke.[2]

	CANCER	¬CANCER
SMOKE	☹	☺☺
DON'T	☹☹	☺

Given (we might say against) this way of representing my decision, the statements of conditional preference (1a) and (1b) are easily heard as true: they respectively summarize the information in the CANCER and ¬CANCER columns of the above decision table. (1c), which seems in some sense to follow from (1a) and (1b), is easily heard as true as well—again, simply by evaluating it against the way of representing my decision in the above table. But, by apparent application of modus ponens, I arrive at something unacceptable—that I cannot hear as true—that is, (1d).

(1) a. If I get cancer, it's better to smoke. (≈smoking is better than not smoking)
b. If I don't get cancer, it's better to smoke.
c. So, if I get cancer or I don't, it's better to smoke.
d. #So, it's better to smoke.

I said (1a) and (1b) are heard as true against this representation of the decision, but maybe you are doubtful. If so, apply the Ramsey Test.[3] Suppose you have preferences like mine. Now suppose that you'll get cancer. Which do you prefer: smoking or not smoking? Evidently, smoking. Bearing this in mind, you should accept (1a). Suppose that you won't get cancer. Which do you prefer: smoking or not smoking? Evidently, smoking. Bearing this in mind, you should accept (1b). And (1c)—more precisely, an intended reading of (1c)—is true, and appears to follow from (1a) and (1b).[4]

[1] Throughout I use "cancer" as shorthand for any form of cancer causally linked to smoking tobacco.

[2] I rely for now on an intuitive notion of what a decision problem is: a decision problem Π is something that determines a preference ordering over outcomes, where o is an outcome in Π iff o is the conjunction $a \wedge s$, for some action a available in Π and some state s that is relevant in Π. In this decision problem, the available actions are the agent smoking or not smoking; the relevant states are the agent getting cancer or not getting cancer. For a refinement of this intuitive notion, see §4.1.

[3] Applying the Ramsey Test to conditionals with disjunctive antecedents, like (4c), is less than straightforward. If, for example, a disjunctive antecedent denotes a question (for discussion, see 3.1), we need to think carefully about how the Ramsey Test might apply to such cases. The Ramsey Test nevertheless remains helpful for accessing the readings of (1a) and (1b) I am interested in here.

[4] A note to decision theorists: I realize the decision is ill-formed, by any metric of well-formedness you might choose. But that

1.2 *Logical Preliminaries*

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

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

(CA) is assumed by most theorists of the conditional (including both Lewis and Stalnaker), but is rejected in certain context-shifting and dynamic frameworks (a point to which we will return below). That said, we will bracket the question of its general validity here. Whether or not (CA) is generally valid, (CA) preserves truth or acceptability in this instance: given (1a) and (1b), there is a true reading of (1c). On the intended reading, (1c) *summarizes* the information contained in (1a) and (1b)—that is to say, it summarizes the information contained in each column of the relevant decision table. We can also bracket the question of whether (1c) has a false reading (as seems to me very likely). 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 of interest here.

Notice that, given (MP), (CA) allows us to derive the familiar elimination rule for disjunction, ($\vee$E), which would permit the slightly shorter argument for smoking in (2):[5]

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

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

A tentative hypothesis is that both (1) and (2) rest ultimately on applications of (MP) that are somehow deductively illicit in the envisioned context.[6] Applying (CA) seems licensed in this context: on the Intended Reading, (1c) is true/acceptable, given (1a) and (1b). Given (MP), reasoning by ($\vee$E) should be deductively licensed in this case. Evidently, though, it is not.

1.3 *Stalnaker on Fatalism*

Puzzles of this general shape like this have a long (and not very distinguished!) history in philosophy. Readers may be familiar with alleged "proofs" of Fatalism, in the form discussed by Dummett (1964); Stalnaker (1975). Commentators like Dummett and Stalnaker tend to regard this reasoning as defective in a

---

does not alter the apparent logical facts: if modus ponens is valid *and (1c) is true*, it's better to smoke, which is contrary to fact.

A note to linguists: I realize that the logical representations I have chosen for (1a)–(1c) ignore matters like tense and aspect—matters of possible significance for evaluating statements of (conditional) preference like these. But 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—tense and aspect are irrelevant to this problem, and we do well to abstract away from the complications would accompany their introduction.

And a note to Aristotelians: while it may be tempting to deny Excluded Middle for future contingents, notice that intuitive failures of modus ponens, like (12), need not utilize future contingents. Notice also that many intuitively valid inferences (e.g., the conclusion of a valid dominance argument) seem to rely on the validity of Excluded Middle for future contingents.

Finally, if you are tempted to respond to this case by appealing to the thesis that $(p \vee \neg p)$ is ambiguous between an alternative (or inquisitive) semantic value and a classical Boolean semantic value, I discuss this possibility in detail in §3.

[5] The argument in (2) is an example of a classic fallacious dominance argument. Fallacious dominance arguments similar to (2) are the focus of Cantwell (2006), and a focus Gibbard & Harper (1981). Although neither Cantwell nor Gibbard and Harper center modus ponens in their discussions, as I do here, we are all interested in the same broad target: a logical account of why intuitively good dominance reasoning is sound, and intuitively bad dominance reasoning is unsound.

[6] The suggestion here is that (2) relies tacitly on (MP), since ($\vee$E) is established by appeal to more basic principles within a proof theory for the conditional, namely, (CA) and (MP).

fairly obvious way. Here I want to explain briefly why this attitude, even if apt to the fatalistic arguments that are their focus, is inapt to a case like (1).[7]

Here is Stalnaker's rendering of the case described by Dummett.

> 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's diagnosis (which is roughly similar to Dummett's) is this:

> [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)

Stalnaker is here suggesting that conditionals (3a) and (3b) are in this case supported or established *by conditional proof* (i.e. derivation of the consequent under supposition of the antecedent).

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

Here, then, is a natural representation of the Fatalist's reasoning:

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

How would Stalnaker's diagnosis of Dummett's case apply to (1)? It's not obvious. Notice that (1a) and (1b) were provided directly by the context, not established or supported by conditional proof: (1a) and (1b) were simply assessed as *true*, in the context I provided.

Since Stalnaker accepts (CA) and (MP), and since he is committed to the falsity of (1d), he is committed to denying that both (and presumably either) of (1a) and (1b) have the (true) readings I have suggested that they have. In this vein, Cantwell (2006) argues that, given Stalnaker's (similarity-based) understanding of a conditional notion like 'the value if I do *a*', there is (independent) reason to regard (sentences like) (1a) and

[7] Dreier (2009) also discusses fatalistic ($\vee$E) arguments, concluding that they illustrate failures of (MP), on the grounds that "In some contexts [$\vee$E] is suspect, but not here, I take it" (127–8). (Thanks to [redacted] for this comparison.) I take Dreier's move here to be dialectically unavailable, as ($\vee$E) *is* typically regarded as the logically suspect principle in cases of this general type (e.g., the Miner Case) (Willer 2012; Yalcin 2012b; Bledin 2014).

(1b) as false. On a similarity-based account, the values realizable if I do *a* are those witnessed at the nearest *a*-worlds. Since smoking *makes near* worlds where you get cancer, the set of values realizable if you smoke includes some very low values indeed. This is, Cantwell suggests, sufficient to falsify 'smoking is better (than not)' in the relevant context: the values realizable if you smoke do not (generally) exceed the values if you don't. Assuming modus ponens, it follows that at least one of (1a) or (1b) must be false.

Although this is roughly the theory of the case that is suggested by a broader Stalnakerian theory, and the theory is consistent, I do not think this can be regarded as evidence *for* the Stalnakerian theory of the case. To the contrary, it is prima facie evidence *against* such a theory, as it indicates that the theory is insufficiently expressive: it cannot *semantically* generate the readings of the conditionals we are here interested in semantically representing. According to this theory, (1a) and (1b) are not jointly acceptable in the stipulated context. A theory that accounts semantically for these readings seems preferable to one that dismisses their possibility a priori.[8]

1.4 Plan

Challenges to the validity of modus ponens (Kolodny & MacFarlane 2010) and modus tollens (Yalcin 2012b) have been a focus of recent interest in the formal semantics of natural language. There is some agreement in the literature that Kolodny and MacFarlane's challenge can be avoided, if the notion of logical consequence is understood aright (Willer 2012; Yalcin 2012b; Bledin 2014). The viability of Yalcin's counterexample to modus tollens has meanwhile been challenged on the grounds that it fails to take contextual domain restriction (and context-sensitivity generally) into proper account (Stojnić 2017; Schulz 2018).

This paper argues that the strategies developed for handling extant challenges to modus ponens and modus tollens do not account for cases like (1). My aim here will be to argue that this "counterexample" (really class of "counterexamples") to modus ponens presents a genuine theoretical puzzle, one addressed neither by easy appeal to context-sensitivity nor by appeal to a novel—whether Dynamic or Informational or Inquisitive—account of the relation of logical consequence.

This paper's other aim is provide a positive account of these cases (and to tie theorizing about their logic and semantics to relevant issues in, e.g., decision theory). The basic idea is straightforward (and will ring a bell for some readers): consistent with semantic competence, there are *well-formed* and *malformed*—I do not say true or false—ways to represent a question (e.g., whether or not to smoke). Although (1c) is evaluated as true, it is so-evaluated against a malformed way of representing this question—in particular, a way of representing this question that *proof-theoretically licenses* an answer to this question, namely (1d), which we nevertheless take to be *deductively unlicensed*, in the relevant situation. Assuming that, when giving a semantics and logic for natural language, it is desirable that an inference be proof-theoretically licensed only when it is in fact deductively licensed, we have reason to treat this case as strictly outside the theory's purview (in a sense I will precisify below). This is not to say that we should not try to account semantically for the judgment that (1c) is true/acceptable in its context of evaluation—we should, and will. It is instead to note, when a set of intuitive judgments of truth/acceptability, in a context of evaluation, "violates" modus ponens, this can be taken as evidence of *latent irrationality* in that set of judgments

[8] All this said, it should be noted that Cantwell shares the sense that the spuriousness of spurious dominance arguments traces to failures of independence (although we pursue this in different theoretical frameworks). Spurious dominance arguments, which are my focus here, differ from Dummett's version of Fatalism, where it is easier to dismiss the central premises (3a) and (3b) a priori. Reasonable precautions always have *a point*, namely, *risk-reduction*; compare, e.g., Ayer (1964) and the statement from Hospers (1967) quoted at Buller (1995: 112). (In particular, in the event that you are not killed in the air raid, you would typically conclude that there is some chance that taking precautions *saved your life* by reducing your risk and so had a clear point after all.) Since the truth of (3a) and (3b) 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 considered in Kolodny & MacFarlane (2010). As they note, Subjectivism's main difficulty is empirical: it is a cost to render conditionals like (1a) and (1b) false, when competent speakers find them acceptable in the stipulated context. See Kolodny & MacFarlane (2010: §1.2) for another argument against Subjectivism.

(but need not be taken to imply that any one of those judgments is strictly false, in the relevant context of evaluation). Since such judgments are consistent with semantic competence, the idea that natural language semantics (construed as the empirical study of states of semantic competence with respect to a natural language) can deliver a "logic of natural language"—a theory of which inferences are deductively licensed, given a set of sentences evaluated as true/accepted—should be reconsidered. That, I will suggest, is better construed as the job of a normative (e.g., decision) theory.

2 Modus Tollens and Context-Sensitivity

This section will review Yalcin (2012b)'s alleged counterexample to modus tollens. It will then argue that none of the obvious strategies for dealing with Yalcin's counterexample supply any explanation of why we err in going from (1c) to (1d). The argument here is technical, but the section may be skimmed or skipped, since the broader point is simple: although context-sensitivity plausibly helps to understand Yalcin's counterexamples, it offers no real traction on our puzzle.

In Yalcin's counterexample, 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 (4a) and (4b) are judged acceptable in this scenario. But, by apparent application of modus tollens, we arrive at something unacceptable: (4c).

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

This section describes two responses 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 (4) are incorrect (because under-described). On another, this is because (MT) is appropriately understood as rule governing preservation of truth with respect to a single context, and in (4) there is an illicit context-shift. I am not primarily interested in whether these replies blunt the force of Yalcin's counterexample (though the second reply is, in my view, quite powerful). My aim in this section is to establish that they do not extend to the counterexample to (MP) presented above.

2.1 Domain Restriction

It is commonly thought that 'if'-clauses semantically function to introduce *domain restrictions* for downstream quantificational expressions (in particular: modal, preferential, probabilistic, or epistemic operators).[9] Letting $\triangle$ be any such operator, the idea is that the relevant operators are binary, i.e., have as arguments a restriction and nuclear scope:

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

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

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

The suggestion (drawn from Stojnić 2017; Schulz 2018) is that, once we disambiguate (4) along these lines, we see it is not an instance of (MT).

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

[9] The notion that 'if'-clauses are restrictors for downstream quantifiers has been developed extensively by Angelika Kratzer (see a.o. Kratzer 1981, 1991, 2012). The implementation here is not quite Kratzer's, since she treats the semantics of the 'if'-clause as *exhausted* by its restrictive function. I consider Kratzer's actual analysis of the relevant conditionals below.

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

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

Similarly, perhaps, the move from (1c) to (1d), is invalid, if represented as follows.[10]

$$\begin{array}{ll} (p \vee \neg p) \Rightarrow \triangle_{p \vee \neg p} q & \\ \triangle q & \# \# \end{array}$$

My concern with this analysis of (1) is that, if modus ponens is valid for indicative conditionals in natural language, we would expect $\triangle q$ to follow logically from $\triangle_{p \vee \neg p} q$ (which follows logically from $(p \vee \neg p) \Rightarrow \triangle_{p \vee \neg p} q$ by MP). Commitment to modus ponens for the natural language indicative conditional would seem to bring in tow commitment to the principle that $\triangle q$ follows validly from $\triangle_{p \vee \neg p} q$.[11]

Note first that, in the context that I have described, there is the sense that the restricted operators—(5) and (6)—are simply *true*.

(5) It is better to smoke, assuming/given CANCER. $\triangle_{p} q$

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

That's unsurprising, as (5) and (6) have *precisely* the logical forms that Kratzer (1981, 1991) proposes for conditionals like (1a) and (1b). On Kratzer's (widely accepted) account, the compositional semantic function of an 'if'-clause is *exhausted* by its restriction of some downstream operator.

This in mind, consider this schematic derivation:

$$\begin{array}{ll} \triangle_{p} q & \\ \triangle_{\neg p} q & \\ \triangle_{p \vee \neg p} q & \text{OCA} \end{array}$$

Given Kratzer's analysis, we may read derivations of this schematic form as representing the logical form of many instances of (CA) in natural language. The OCA derivation appears to preserve truth/acceptability in the context for (1): in this context, the acceptability of the following restricted operator seems to follow from the acceptability of (5) and (6)—as we would expect, given Kratzer's analysis, since (7) just is the logical form that Kratzer assigns (by default) to (1c).

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

The difficulty here is that, if restricted operators are conditional in interpretation—as they surely *are*—and (MP) 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 (7), given that CANCER ∨ ¬CANCER is a tautology.

Insofar as restrictable operators in natural language in conditional in interpretation, we expect (and indeed observe) that they generally go in for (something that at least looks a great deal like!) modus ponens. Here is an example (it is trivial to replicate).

(8) a. The likelihood that the marble is red, given that it is big, is 75%.
b. The marble is big.
c. ✓ So the likelihood that the marble is red is 75%.

This is as we would expect, if (i) (MP) is valid for natural language indicative conditionals, (ii) Kratzer is correct in thinking that (8a) is equivalent to 'if the marble is big, there's a 75% chance it's red.'

This is reason to formulate (MP), like (CA), as a proof-theoretic principle for all conditional operators (with the conditional operator ⇒ understood as a special case), along the following lines:

[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 §2.2.

[11] Charlow (2013b); Schulz (2018) register a different view, but they simply overlook the thesis that logical consequence is informational or dynamic in character. More on this below.

$$\begin{array}{ll} \triangle_{\phi}\psi & \\ \phi & \\ \triangle\psi & \text{OMP} \end{array}$$

One might of course worry whether there is a notion of logical consequence to "cover" OMP inferences like (8), given that the truth of $\triangle_p q$ and p at a world or situation of evaluation w does not generally guarantee the truth of $\triangle q$ at w (on this point, see e.g. Charlow 2013b,c; Schulz 2018). Worry not: such an understanding has been independently suggested as a way of rescuing modus ponens from the challenge developed in Kolodny & MacFarlane (2010) (see e.g. Willer 2012; Yalcin 2012b; Bledin 2014). On this understanding, logical consequence is *informational* or *epistemic* in nature. Valid arguments—what Bledin (2014) dubs "good deductive inferences"—are roughly, on this understanding, *knowledge-* or *information-preserving*.

Good deductive inference does not generally track *preservation of truth with respect to a world of evaluation*, as a case like (8) shows. The fact that (i) a credal measure Pr that evaluates q as .75 likely given p at w, together with the fact that (ii) p is true at w does not imply that (iii) Pr evaluates q as .75 likely at w. Pr is not omniscient! However, (iii) does follow from (i), on the assumption that (ii') conditionalizing Pr on the information that p is idle, just as we would expect, if validity tracks the preservation of *belief* or *information*, rather than truth. More generally, although instances of (OMP) in natural language do not generally preserve truth with respect to a world of evaluation, they are ordinarily judged impeccable. Bledin's understanding of logical consequence would help to explain why: updating or conditioning on the information expressed by the premises makes available the information expressed by the conclusion.

Our present difficulty is that, on this understanding, we would still apparently predict that a claim like (7) will allow us to infer the preferability of smoking by (OMP). Evidently, though, this is a *bad* deductive inference to draw: the information made available by (7) does not deductively license the conclusion. This is our original puzzle, only now rendered in a metalanguage of restrictable conditional operators.

2.2 *Context-Sensitivity Via Context-Shifting*

Another strategy, due to Gillies (2009, 2010), assigns the conditionals familiar logical forms, but gives the conditional a distinctive *information-sensitive* and *context-shifting* interpretation.[12] For Gillies, the conditional has a dual compositional function: relative to a world of evaluation w, it quantifies universally over worlds selected by applying a selection function f_w to the antecedent-witnessing possibilities in a domain i_w, while also *shifting the context of interpretation* for context-sensitive material in the consequent (to a context that incorporates the information contained in the antecedent). Let $\sigma = \langle i, f \rangle$. Then:

Definition 1. $[\![\phi \Rightarrow \psi]\!]^{\sigma,w} = 1$ *iff* $f_w(i_w \cap [\![\phi]\!]^{\sigma}) \subseteq [\![\psi]\!]^{\sigma[\phi]}$. $\qquad \sigma[\phi] := \langle \lambda w.i_w \cap [\![\phi]\!]^{\sigma}, f \rangle$

If conditional antecedents are *literally* context-shifters, it is obvious that we need to exercise care in formulating rules like (MT) and (MP). According to Gillies' semantics, in Yalcin's counterexample to (MT), a sentence of the form $\triangle q$ is evaluated relative to an informationally enriched context—an information state enriched with the information expressed by the antecedent of (4a). Meanwhile $\triangle q$ is evaluated relative to an unenriched information state—the global information state appropriate to the distribution of marbles Yalcin describes. There is *no antecedent reason* to understand (MT) so that, when hypothetically/locally conditioning the context on ϕ *affects the truth-conditional content of* ψ, we are pressured to regard $\neg\phi$ as derivable from $\phi \rightarrow \psi$ and $\neg\psi$ via (MT) (for more on this point, see Stojnić 2017). This would be a principled (and to my eye rather compelling) explanation of why Yalcin's (4) is invalid (but would not be appropriately regarded as an instance of MT).

Let's return to case (1). Is this analysis adaptable to it? Notice (CA) fails on the context-shifting strategy. Consider an example that Gillies uses: it is presupposed that there is exactly one red or yellow marble concealed in an opaque box (Gillies 2010: 13). Then each of the following are true:

(9) 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.

[12] I will not discuss dynamic accounts separately, since they, like Gillies' account, tend to avoid problematic predictions surrounding certain ($\vee$E) arguments by invalidating (CA) (see, e.g., Willer 2012).

From (9b) and (9c) it doesn't seem to follow that (indeed, Gillies' semantics predicts it is false that):

(10) If it's RED or it's YELLOW, then, for some color k, it must be k.

Given (MP), (10) would imply that, for some $k \in \{\text{YELLOW}, \text{RED}\}$, the marble must be k. Given (9a), it is *neither* the case that it must be RED nor that it must be YELLOW; given (9a) and (MP), then, (10) cannot be true. Something similar might be said to explain the badness of inference (1): as when we go from (9b) and (9c) to (10), the move from (1a) and (1b) to (1c) rests on a specious application of (CA).

Although Gillies' semantics offers a nice treatment of Yalcin's counterexample (and coincidentally invalidates (CA)) it offers us basically no traction on our puzzle: invalidating (CA) does not defeat the intuitive appeal of (1c).[13] Where (1) goes wrong is in the move from (1c) to (1d), not in the move from (1a) and (1b) to (1c). (I again stress that I am not presupposing that (CA) is valid, only that the move from (1a) and (1b) to (1c) preserves truth or acceptability in its context of use.) Although we presented (1c) as being "derived" from (1a) and (1b) by application of (CA), this was inessential to the presentation. (1c) is a 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.

The puzzle is simple: the truth of (1c) (on the relevant reading, in the stipulated context of evaluation), does not warrant inferring (1d) (under *any* way of reading (1d)). This is a puzzling state of affairs. But I am not sure that there is any other way of reading the data.

3 Taking the Case Seriously

I will ultimately argue that such "counterexamples" to (MP) are to be explained with a *normative*, rather than logical, account (on which the "unsoundness" of (1) is due to the fact that the way of representing the decision made salient in this context is an unreasonable way of representing a decision).

It would be if nothing else be easier if we could get by without such a fancy normative story, while also taking this case seriously (by which I mean we take the judgment that (1c) is acceptable in the relevant context as data, rather than something to be explained away). So that is where we will put our attention first. The best prospects here lie with a family of related views known as Alternative or Inquisitive Semantics (see, e.g., Alonso-Ovalle 2006; Groenendijk & Roelofsen 2009; Ciardelli & Roelofsen 2011). This section provides a general overview of such views, eventually zeroing in the theory of Bledin (2020). Although Bledin's account does better with respect to (1) than any other account we have so far considered, I will suggest that his account *under-generates* good deductive inferences (in the broader setting of what I will call a logic of decision). The theory I go on to state will attempt to find a middle ground on which certain intuitively sound episodes of dominance reasoning are deductively licensed, although not of course the episode of dominance reasoning dramatized in (1).

3.1 Inquisitive Logic

Let us begin with a basic observation: (the relevant reading of) $(p \vee \neg p) \Rightarrow q$ plausibly entails both $p \Rightarrow q$ and $\neg p \Rightarrow q$. Combine this with (CA), and we have:

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

If $(\vee E)$ is not (as many believe) a valid rule of inference, we cannot derive q from the premise set Γ_1. Unsurprisingly, q is not generally derivable from a *logically equivalent* premise set Γ_2.

$$\Gamma_1 = \{(p \Rightarrow q) \wedge (\neg p \Rightarrow q), p \vee \neg p\}$$
$$\Gamma_2 = \{(p \vee \neg p) \Rightarrow q, p \vee \neg p\}$$

There is no doubt something to this. Strikingly, however, it *appears* to concede that indicative conditionals of the form $(p \vee \neg p) \Rightarrow q$ do not generally go in for (MP) (while supplying an explanation of this fact).[14]

[13] Indeed, we might (and I do) worry that Gillies' account doesn't try to explain why we can get a true reading of (10), given (9b) and (9c). Given that the marble is yellow *or* that it is red, one is a position to conclude that there is some color (yellow or red) such that the marble must be that color. Moreover, there is a sense in which the consequent of this conditional, on its own, is perfectly acceptable in this context: a speaker can claim that there there is some color (yellow or red) such that the marble must be that color—or more simply that: *either* it must be yellow *or* it must be red—*seemingly* without thereby contradicting (9a).

[14] Bledin (2020) would resist this characterization of the theory, for reasons we will see below.

This prompts an obvious question: under exactly what conditions *is* inferring q from $(p \vee \neg p) \Rightarrow q$ logically permitted? Alternative/Inquisitive Semanticists may see an answer ready to hand. Following this tradition, let us 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 an inquisitive (question-like) content, namely the set of propositional alternatives $\{p, \neg p\}$. On the non-simplifying representation, the antecedent denotes a proposition, derived by application of a flattening operator ! (such that, if S is a set of alternative propositions, $!S = \bigcup S$) to $\{p, \neg p\}$. This leaves two options for representing the move from (1c) to (1d).

$\{p, \neg p\} \Rightarrow \triangle q$		$\{p, \neg p\} \Rightarrow \triangle q$	
$!\{p, \neg p\}$	$\top$	$\{p, \neg p\}$	##
$\triangle q$	##	$\triangle q$	MP

Supposing $\ulcorner \{p, \neg p\} \Rightarrow \triangle q \urcorner$ is an available representation of (1c), it is (as indicated by the ##) essential to this line of reply that there is no rule of deductive inference that allows one to introduce a sentence expressing $\{p, \neg p\}$ in the derviation. Here is a natural (if naïve) question: what justifies this prohibition? In introducing $\{p, \neg p\}$, one is introducing something *informationally trivial*, but with inquisitive or alternative-presenting content (Groenendijk & Roelofsen 2009). It is natural to regard such a move as deductively licensed: in introducing $\{p, \neg p\}$, one is, roughly, deciding to consider or raise to salience 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, it seems that it is always possible to introduce a set S that partitions the relevant possibilities *without collateral 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 permitted, even if such a question were not already salient in the relevant contexts.[15]

Bledin (2020), building on Ciardelli (2018), uses an Inquisitve Semantic framework to state an account of cases like (1) that answers this naïve challenge. Bledin's core suggestion is this: if we represent an agent as reasoning from an inquisitive premise (or supposition) $?p$, we thereby represent them as having an (*indeterminate*) opinion on the question $?p$.

> *How much is one Bitcoin worth in US dollars?* stands in for one of the various ways in which the question expressed by this sentence can be resolved. When this constituent interrogative appears as a premise in an inference, we can come to learn things about any information state that settles it... For example, if we infer the polar interrogative *Is one Bitcoin worth more than a thousand dollars?*, we thereby establish that the more specific type of information that settles how much one Bitcoin is worth in US dollars yields the less specific type of information that settles whether one Bitcoin is worth more than a thousand dollars.
>
> ...[W]e can think of interrogatives as placeholders for arbitrary decision states of a given type. In [(1)], the premise ['I will or will not get cancer'] stands in for an arbitrary decision state that settles the question of whether you are going to [get cancer]. (149)

On the Ciardelli-Bledin account, the argument from $?p$ and $?p \Rightarrow \triangle q$ to $\triangle q$ is to be regarded as valid, in this rough sense: a decision state that satisfies the condition on decision states expressed by $?p$ and $?p \Rightarrow \triangle q$ *is* a decision state that accepts $\triangle q$. An agent who merely entertains the question *is it the case that p?* is not, in general, to be understood as satisfying the condition on decision states expressed by $?p$: satisfying this condition, in the logically relevant sense, implies either accepting p or accepting $\neg p$. On this account, for an agent to accept (satisfy the condition on decision states) $?p \Rightarrow \triangle q$ is for them to be such that both:

- If they accept p, they must accept $\triangle q$.
- If they accept $\neg p$, they must accept $\triangle q$.

It is obvious that, on this account, accepting $?p \Rightarrow \triangle q$ does not require accepting $\triangle q$, since an agent can accept $?p \Rightarrow \triangle q$ without thereby accepting p or accepting $\neg p$.[16]

[15] This isn't to say one can go around freely expressing such contents in discourse. Questions are typically *not relevant* unless they address a larger question in an extant discourse (see, e.g., Roberts 1996), and speakers must generally make their contributions relevant, vis-à-vis the broader goal of the extant discourse. I take this to be orthogonal to the logical issues under consideration here.

[16] Ciardelli and Bledin do not say much by way of explicating the idea of a theory of logically valid reasoning making use of

In addition to explaining why the argument in (1) fails, Bledin aims to say under what conditions inferring $\triangle q$ from $(p \vee \neg p) \Rightarrow \triangle q$ *is* logically permitted. According to Bledin, what distinguishes good instances of this form, like (11), from bad, like (1), is a *distinctive semantic property of the consequent*.

(11) a. If it's raining, the match is likely be cancelled.
b. If it's not raining, the match is likely be cancelled.
c. So, if it's raining or it isn't, the match is likely be cancelled.
d. So, the match is likely be cancelled.

Specifically, Bledin says an inference of this form is good just when the consequent $\triangle q$ has a property he dubs Coarse Distributivity. (Note: here $s \subseteq W$, while $\preceq$ is a weak partial order on subsets of s.[17])

> **Coarse Distributivity.** ϕ is Coarsely Distributive iff for any decision state $\langle s, \preceq \rangle$ and partition $\{s_1, ..., s_n\}$ of s, if $[\![\phi]\!]^{s_i, \preceq} = 1$ (for all $1 \leq i \leq n$), then $[\![\phi]\!]^{s, \preceq} = 1$. (147)

Informally, ϕ is Coarsely Distributive just when acceptance of ϕ relative to every possibility in a partition implies outright (unconditional) acceptance of ϕ. For Coarsely Distributive ϕ, Bledin's account predicts it is impossible for an agent to accept $?p \Rightarrow \phi$ without also accepting ϕ.

Bledin's thought here is that sentences expressing *comparative preference* are a paradigm of *Non-Distributivity*: conditional on the proposition that I get cancer (from smoking), as well as conditional on the proposition that I do not get cancer (from smoking), it is better for me to smoke; nevertheless, unconditionally, it is obviously not the case that it is better for me to smoke. In stark contrast, sentences expressing comparative probability ('ϕ is likelier than ψ') are a paradigm of Coarse Distributivity. Note first the following fact about probability/comparative likelihood.

> **Likelihood Coarsely Distributes.** Consider any state $\langle s, \preceq \rangle$ and partition $\{s_1, ..., s_n\}$ of s. If, for each i, $q \cap s_i$ is likely in $\langle s, \preceq \rangle$ ($:= \neg q \cap s_i \prec q \cap s_i$), then q is likely in $\langle s, \preceq \rangle$ ($:= \neg q \prec q$).

This follows from a more general fact about probability measures (roughly, if q is likely conditional on each cell of a partition of s, then q is likely conditional on s):

> **Statewise Probabilistic Dominance (SPD).** Consider any s and $\{s_1, ..., s_n\}$ that partitions s, and let Pr be a probability measure. If $\Pr(q \mid s_i) \geq x$ (for all $1 \leq i \leq n$), then $\Pr(q \mid s) \geq x$.

For Bledin, the contrast between (1) and (11) rests on a substantive difference between statements of comparative likelihood and statements of comparative preference: the first are coarsely distributive, the second aren't.[18] (11) is *not* an instance of modus ponens (since for it to be so, we would need to appeal to {it's raining, it's not raining} as a premise, which, for reasons already seen, Bledin's notion of logical consequence will prohibit in this case). To see this more clearly, it is helpful to examine a case where a

---

premises whose content is *indeterminate*. I tend to understand their account of logical consequence supervaluationally: ψ is a logical consequence of ϕ just when, for any *determinate assignment* g of semantic content to sentences, $g(\psi)$ is a logical consequence of $g(\phi)$. (Ciardelli compares his account to the Intuitionistic treatment of disjunction, which is also an appropriate comparison.) Even with this understanding, the account has puzzling features: for the reasons canvassed above, one would naïvely think it should be possible to represent an agent as reasoning validly from a disjunction $p \vee \neg p$, interpreted inquisitively as $\{p, \neg p\}$, without thereby representing that agent as taking a stance on whether p. The account I will ultimately sketch does not share this puzzling feature.

[17] As the ensuing discussion illustrates, an ordering like $\preceq$ may be used to represent a binary preference relation, as when we are interpreting a construction of comparative preference ('better'), but also to represent a binary probabilistic relation, as when we are interpreting a construction of comparative probability ('likely' or 'likelier than not').

[18] Another way of putting this point is that Bledin is uninterested in giving a *proof theory* according to which (11) is a deductively permitted inference. This is a bit surprising: to say $\triangle$ is Coarsely Distributive is basically to say that $\triangle$ (understood as a binary, conditional operator) validates (a restricted form of) modus ponens! This is related to the issues raised in §2.1. We would like an account of why certain inferences with binary operators (that happen to look an awful lot like modus ponens) are deductively licensed and others are not. Bledin's account offers a (non-proof-theoretic) justification of the inference in (11), since it is obvious that likelihood *is*, in fact, Coarsely Distributive. Bledin proposes to explain cases like (1) by appeal to the assumption (which he takes to be obvious, in light of such cases) that preferability is *not* Coarsely Distributive. But the logic of preferability, in the *relevant* sense, *is* plausibly, just like the logic of likelihood, Coarsely Distributive (more on this below).

formally indistinguishable inference fails. A marble has been drawn from an urn containing 100 marbles in the following distribution:

- 30 completely red marbles (all of these small)
- 30 completely blue marbles (all of these small)
- 40 half-red and half-blue marbles (all of these big)

In this case, the alternatives in {it's red, it's blue} ('it's red' here understood to mean that it's somewhere red, 'it's blue' that it's somewhere blue) are exhaustive: all the marbles are somewhere red or blue.

(12) 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.

Although (11) and (12) share a logical form, (11) is impeccable and (12) is outrageous. The only difference between (11) and (12) is of a substantive character: the set of alternatives in (11) is a *partition* of the set of relevant possibilities, while the set of alternatives in case (12) is not (since the elements of {it's red, it's blue}, though jointly exhaustive, are also jointly compatible). Although neither (11) nor (12) amounts to an instance of modus ponens, (11) is nevertheless valid, its validity owing to a substantive (necessary) truth about comparative likelihood (rather than to modus ponens).

3.2 Reasoning by Dominance

Contra Bledin, I will now argue that (1) fails because it is (a textbook case of) Spurious Reasoning by Dominance—not because constructions of comparative preference ('smoking is preferred to not smoking') are in general not Coarsely Distributive. The contrast between Spurious and Non-Spurious Reasoning by Dominance is accounted for as follows: rules of deductive inference, like modus ponens, are to be understood as rules of *rational deductive inference*. A theory of rational deductive inference is a theory of which inferences are deductively licensed, *given* a reasonable way of representing on some question.

The Dominance principle can be preliminarily characterized thus: if, for every relevant contingency s, the value of a given s exceeds the value of any other available action given s, then the unconditional value of a exceeds the unconditional value of any other available action. More formally:

> **Statewise Dominance (SD).** Consider decision problem Π. (s_i is a relevant contingency, a_j an available action, $\mathrm{Val}_{\preceq}(a_j \mid s_i)$ the degree to which performing a_j, given s_i, satisfies the preferences encoded in $\preceq$, which I will abbreviate as its 'value'.)
>
> | Π | s_1 | ... | s_n |
> |---|---|---|---|
> | a_1 | $\mathrm{Val}_{\preceq}(a_1 \mid s_1)$ | ... | $\mathrm{Val}_{\preceq}(a_1 \mid s_n)$ |
> | ... | ... | ... | ... |
> | a_m | $\mathrm{Val}_{\preceq}(a_m \mid s_1)$ | ... | $\mathrm{Val}_{\preceq}(a_m \mid s_n)$ |
>
> If $\mathrm{Val}_{\preceq}(a_k \mid s_i) < \mathrm{Val}_{\preceq}(a_j \mid s_i)$ (for all $k \neq j$ and $1 \leq i \leq n$), then $\mathrm{Val}_{\preceq}(a_k) < \mathrm{Val}_{\preceq}(a_j)$ (for all $k \neq j$).

Cases of "spurious" reasoning by Dominance, like (1), show that SD cannot hold in full generality: if we allow Statewise Dominance to apply to the decision problem associated with (1), we will fail to take proper account of the (causal or evidential) dependence of relevant contingencies on the available actions.[19] Restricting SD to cases in which the relevant contingencies are *independent* of the available actions is essential, if we want to rationalize any action that is in some respect undesirable, but which nevertheless *contributes to* one's prospects for avoiding a worse situation—defense expenditures (Jeffrey 1983: 8–9), payment of protection money (Joyce 1999: 115ff), and, indeed, refraining from smoking.

[19]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; Joyce 1999). More on this below.

In decision-theoretic contexts, this sort of restriction is sometimes understood as a restriction of the dominance principle to "*well-formed*" decision-problems: decision problems in which relevant contingencies are, simply, independent of the available actions.[20]

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

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

> **Restricted Statewise Dominance (RSD).** For *well-formed* Π: if $\mathrm{Val}_{\preceq}(a_k \mid s_i) < \mathrm{Val}_{\preceq}(a_j \mid s_i)$ (for all $k \neq j$ and $1 \leq i \leq n$), then $\mathrm{Val}_{\preceq}(a_k) < \mathrm{Val}_{\preceq}(a_j)$ (for all $k \neq j$).

On the account I propose here, (1) fails roughly because the decision problem I made salient in describing the case is malformed, not because Coarse Distributivity fails for the comparative preferability operator.

Is it really the nature of the relevant decision situation that accounts for (1)? I'll say yes,[21] but let us consider an argument otherwise. Consider this well-known 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.

(13) a. If the miners are in shaft A, blocking a shaft is best.
b. If the miners are in shaft B, blocking a shaft is best.
c. So, if the miners are in shaft A or in shaft B, blocking a shaft is best.
d. ??So, blocking a shaft is best.

In this case, there is no failure of act-state independence: the miners are where they are, independent of what we choose to do. (13) is not explained by appeal to RSD; what reason is there to think that (1) is?

Here, though, is a key difference. In (13), unlike (1), the context *guarantees a true reading of the conclusion*: there is a clear sense in which either blocking A is (unconditionally) preferable or blocking B is (unconditionally) preferable, which of these depending on the miners' actual location. I'll henceforth refer to this sense of preferability as *primary preferability*.[22]

Primary preferability appears to be coarsely distributive, if not in absolute generality, then certainly in contexts like (13). This raises a question that Bledin's account does not answer: if primary preferability coarsely distributes in contexts like (13), why not in contexts like (1)? Notice that in case (1), it would *not* be best if you smoked—or, at least, nothing about the context suffices to guarantee that *smoking is preferred to not smoking*, on *any* way of reading this claim. The context in (1) does not guarantee that the best possibilities (within your practical or causal reach) are possibilities in which you smoke: *smoking might give you cancer*, in which event it would have been better not to smoke. By contrast, the context in (13) does guarantee—via, it seems, a trivial modus ponens inference from (13c)—that the best possibilities (within your practical or causal reach) are possibilities where you block a shaft. This point of contrast between (1) and (13) is evidently explained by the failure of act-state independence in (1) but not (13).

4 Logic of Decision

This section will develop the idea about deductive logical consequence I have been intimating above. On the account I'll pitch in this section, normative questions encroach on questions about how to model

[20] The stipulation that a decision theory's choice function is defined only over well-formed decision problems is most closely associated with Savage (1972). Alternatives to Savage's theory, like Jeffrey (1983), also restrict dominance reasoning to a certain class of decision problems: those in which the relevant states are states (that the agent regards as) evidentially independent of the actions (the agent regards as) available (to her). By "well-formed decision problem", I mean simply a decision problem in which dominance reasoning is well-applied. For an overview of the dialectic, see esp. Joyce (1999: Ch. 5).

[21] This is also the diagnosis that Gibbard & Harper (1981); Cantwell (2006) offer of bad ($\vee$E) dominance arguments like (2).

[22] There is also evidently a sense of 'best'—which we might roughly gloss as 'selected by the correct decision theory', and to which I will refer as *deliberative preferability*—according to which blocking a shaft is *clearly not* best. Reading 'best' this way probably accounts for our squeamishness with respect to (13d) (appearing, as it does, at the end of an implicit episode of practical deliberation). To screen off this sense in reading (13), let me reiterate an implicit supposition: *we have no idea where the miners are*. Evidently, (13a) still has a true reading, even though, even on the supposition that the miners are in A, no plausible decision theory will, on the supposition that we have no idea where the miners are, recommend blocking a shaft. I conclude that when we judge (13a)–(13c) acceptable in the miners context, we are reading 'best' as primary preferability.

good deductive reasoning. I will try to explain why it is natural to regard good deductive reasoning (and the understanding of logical consequence that follows from it) as being *relative to*, or *dependent upon*, a substantive normative (e.g., decision) theory. And I will argue that this offers an appealing way of theorizing apparent "counterexamples" to (MP), in the mold of (1).

4.1 A Toy Theory

I will illustrate the idea by providing a toy theory (for a toy language $\mathcal{L}$ built from a base propositional language, binary likelihood operator $\blacktriangle$, binary preference operator $\star$, and two-place conditional operator $\Rightarrow$[23]). The theory begins with this refinement of the notion of a decision state.

Definition 2. *A **decision state** $\Pi = \langle \mathcal{S}, \mathcal{A}, \Pr, \mathrm{Val}_{\preceq} \rangle$ where:*

- $\mathcal{S} = \{s_1, ..., s_n\}$ *is a set of states represented in* Π.
- $\mathcal{A} = \{a_1, ..., a_m\}$ *is a set of actions represented in* Π.
- $\Pr(\cdot \mid \cdot)$ *is a conditional credence function, where* $\Pr(\cdot) := \Pr(\cdot \mid \top)$.
- $\mathrm{Val}_{\preceq}(\cdot \mid \cdot)$ *is a conditional value function, where* $\mathrm{Val}_{\preceq}(a \mid s)$ *is the degree to which a is preferred conditional on s and* $\mathrm{Val}_{\preceq}(\cdot) := \mathrm{Val}_{\preceq}(\cdot \mid \top)$.

Definition 3. $\Pi[\phi]$*, the **update** of* $\Pi = \langle \mathcal{S}, \mathcal{A}, \Pr, \mathrm{Val}_{\preceq} \rangle$ *on* $\phi \in \mathcal{L}$*, is the result of conditioning* $\mathcal{S}$*,* Pr*, and* Val *on x:* $\Pi_{\phi} = \langle \mathcal{S}[\phi], \mathcal{A}, \Pr(\cdot \mid \phi), \mathrm{Val}_{\preceq}(\cdot \mid \phi) \rangle$.

$$\mathcal{S}[p] = \{s \in \mathcal{S} : s \subseteq p\}$$
$$\mathcal{S}[\neg\phi] = \mathcal{S} - \mathcal{S}[\phi]$$
$$\mathcal{S}[\phi \wedge \psi] = \mathcal{S}[\phi] \cap \mathcal{S}[\psi]$$
$$\mathcal{S}[\phi \vee \psi] = \mathcal{S}[\phi] \cup \mathcal{S}[\psi]$$

Like Bledin (as well as Cariani et al. 2013; Charlow 2016, 2018), we define satisfaction for $\mathcal{L}$ relative to a decision state $\Pi = \langle \mathcal{S}, \mathcal{A}, \Pr, \mathrm{Val}_{\preceq} \rangle$. Satisfaction in general is *idle update*, while entailment is simply *preservation of satisfaction* with respect to an arbitrary decision state:

$$\Pi \vDash \phi \text{ iff } \Pi[\phi] = \Pi$$
$$\phi_1, ..., \phi_n \vDash \psi \text{ iff } \forall \Pi \vDash \phi_1 ... \vDash \phi_n : \Pi \vDash \psi$$

I won't define updates for $\blacktriangle$ or $\star$ or $\Rightarrow$,[24] though I will provide their satisfaction conditions. Let $[\![\phi]\!]^{\Pi}$ designate the set of states represented in Π that support ϕ: $[\![\phi]\!]^{\Pi} = \{s \in \mathcal{S} : s \subseteq \phi\}$.[25]

$$\Pi \vDash \blacktriangle(\psi \mid \phi) \text{ iff } \Pr([\![\psi]\!]^{\Pi} \mid [\![\phi]\!]^{\Pi}) > \Pr([\![\neg\psi]\!]^{\Pi} \mid [\![\phi]\!]^{\Pi})$$
$$\Pi \vDash \star(\psi \mid \phi) \text{ iff } \mathrm{Val}_{\preceq}([\![\psi]\!]^{\Pi} \mid [\![\phi]\!]^{\Pi}) > \mathrm{Val}_{\preceq}([\![\neg\psi]\!]^{\Pi} \mid [\![\phi]\!]^{\Pi})$$
$$\Pi \vDash \phi \Rightarrow \psi \text{ iff } \forall s \in [\![\phi]\!]^{\Pi} : \Pi[s] \vDash \psi$$

Supposing $\star$ is primary preferability, $\mathrm{Val}_{\preceq}([\![\psi]\!]^{\Pi} \mid s)$ represents, to a first pass, the degree of value an agent whose decision situation is representable with $\Pi[s]$ can intentionally cause or realize by making it the case that ψ (ψ-ing, for short). Primary preferability is here construed akin to (relevantly) informed advisability: if you are "lost in the woods without a map or compass," the primarily preferable thing—the thing a relevantly informed advisor would advise you to do—would be to walk in the direction of your car (whereas the deliberatively preferable thing to do is to "pursue one of the standard strategies for getting out

[23] I will assume for simplicity that these operators take only sentences from the base propositional language (or Boolean combinations thereof) as restrictors.

[24] Were I to do so, I would simply treat them as tests, in the sense of Veltman (1996): updates that return the original state in the event that the sentence is satisfied, and which return an absurd state otherwise (compare Charlow 2015).

[25] I am here assuming that $[\![\phi]\!]^{\Pi}$ and $\Pi[\phi]$ are defined iff, for all $s \in \mathcal{S}$, $s \cap \phi = s$ or $s \cap \phi = \emptyset$ (that is to say, ϕ presupposes, in a loose sense, that the relevant information is partitioned ϕ-wise). This is a visibility presupposition, in the sense of Yalcin (2011). $\Pr([\![\psi]\!]^{\Pi} \mid [\![\phi]\!]^{\Pi})$ is naturally defined as the sum of the probabilities (given ϕ) for the possibilities in $\mathcal{S}$ that entail ψ: $\Pr([\![\psi]\!]^{\Pi} \mid [\![\phi]\!]^{\Pi}) := \Pr(\bigcup [\![\psi]\!]^{\Pi} \mid \bigcup [\![\phi]\!]^{\Pi})$.

of trackless woods: walk carefully in a straight line by sighting along trees, or go consistently downhill") (Gibbard 1990: 18–19).

To further explicate the notion, consider this representation of the decision situation for (13):

	they're in A	they're in B
block A	10 live	10 die
block B	10 die	10 live
don't block A or B	9 live	9 live

A key fact of this decision situation is that, whatever the miners' location, the degree of value you're positioned to bring about if you block a shaft exceeds the degree of value you're positioned to bring about if you don't block a shaft. Blocking a shaft is what you would choose to do, if only you knew where the miners were—something that is *guaranteed by* the fact that blocking a shaft is what you would choose to do if you had settled which state of your decision situation was actual.

This case supports the conditional $(A \vee \neg A) \Rightarrow \star BL$, read: if the miners are in A or aren't in A, blocking a shaft is (primarily) preferable. On the assumption that dominance reasoning to conclusions about primary preferability is logical—on the assumption that $\star$ is coarsely distributive—we predict (it would appear correctly) that $\star BL$ is a logical consequence of $(A \vee \neg A) \Rightarrow \star BL$.

$$\begin{aligned}\Pi \vDash (A \vee \neg A) \Rightarrow \star BL \text{ implies } &\forall s \in [\![A \vee \neg A]\!]^{\Pi} : \Pi[s] \vDash \star BL \\ \text{implies } &\Pi[A] \vDash \star BL \text{ and } \Pi[\neg A] \vDash \star BL \\ \text{implies } &\Pi \vDash \star BL\end{aligned}$$

But, of course, on the assumption that $\star$ is coarsely distributive, we also predict (it would appear incorrectly) that (1d) is a logical consequence of (1c). I'll now consider two ways to avoid this prediction (while also managing to construe dominance reasoning as logical), each of which answers to a different set of theoretical desiderata for what I will call a "logic of decision."

4.2 Rational Deductive Inference

I'll first consider this very direct fix: we stipulate that the theory we have provided is a theory covering all, but *only*, the reasonable decision states (i.e., those witnessing act-state independence).

Definition 4. $\Pi = \langle \mathcal{S}, \mathcal{A}, \Pr, \mathrm{Val}_{\preceq} \rangle$ *is* ***reasonable*** *only if, for all* $s \in \mathcal{S}$ *and* $a \in \mathcal{A}$, $\Pr(s \mid a) = \Pr(s)$.

This is intended as an extra clause for Definition 2: Π is a decision state if and only if Π is reasonable (and satisfies the other conditions in that definition). Cases like (1) are then theorized as follows. The case makes salient a (type of) decision state that is unreasonable. This theory is, strictly speaking, silent on the logical commitments (i.e., the set of inferences that are deductively licensed), given such a decision state.

Whether dominance reasoning (to conclusions about primary preferability) is deductively licensed (at a given context of evaluation) thus *depends* on (decision-theoretically significant) features of the context. Notice that—for reasons already seen—for any (reasonable) decision state Π such that $\Pi \vDash$ (1c), $\Pi \vDash$ (1d). Imagine (obviously contrary to fact) that one's risk of cancer was *independent* (or was *represented as* independent) of smoking. In such a practical context, the reasoning in (1) would be impeccable.

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

On the present account, the contrast is explained by a shift in the context of evaluation: the decision state Π against which (14) is evaluated is, by assumption, reasonable; so $\Pi \vDash$ (14a), and hence $\Pi \vDash$ (14b). So, although the satisfaction of (1c) by Π *would* logically necessitate the satisfaction of (1d) by Π, the decision state against which (1) is evaluated for satisfaction is unreasonable. On the present theory, *no* inferences—certainly no inferences to conclusions bearing on whether or not to smoke—are strictly licensed, from the vantage point of such a decision state.[26]

[26] With one further natural stipulation, this theory also yields an account of dominance reasoning in the more familiar sense (which

The aim of this section was to formulate a theory of logical commitment (by way of a theory of $\vDash$) in which modus ponens, understood as a principle for *all two-place conditional operators* $\triangle$ (recall §2.1), would preserve logical commitment.

$$\forall \phi, \psi : \phi, \triangle_{\phi}\psi \vDash \triangle\psi$$

I have here suggested that a theory that "validates modus ponens", in this particular sense, is possible, provided we require that decision states appealed to in our model theory for deductive inference satisfy substantive criteria of what I have called reasonability.[27]

How should we regard the assumption that the decision states appealed to in our model theory for deductive inference are reasonable? The assumption is, I'll suggest, best regarded as a *rational idealization* (similar in motivation, as I will explain, to the assumption that Pr is, by definition, a *probability* measure).

Philosophers who model the notion (and logic) of degreed belief commonly assume, as I have done here, that a state of degreed belief consists in an agent's relation to a probability measure (or set thereof) that is updated by via (pointwise) conditionalization.[28] But the assumption that probability measures are (i) probabilistically coherent and (ii) are updated via conditionalization is standardly justified by appeal to a *normative argument* (although the style of argument varies with the explanatory interests of the theorist): probabilistic incoherence or failure to conditionalize "subjects" or "disposes" an agent to epistemically undesirable inferences (in the context of Accuracy-Dominance arguments for probabilistic coherence or conditionalization) or practically undesirable inferences (in the context of Dutch Book arguments for probabilistic coherence or conditionalization).[29]

The requirement that decision states witness act-state independence can be seen as motivated by broadly normative considerations: failures of reasonability "commit" or "subject" an agent to bad modus ponens inferences, and, for certain modeling purposes, it makes sense to idealize away from this (as we often idealize away incoherent credences when modeling probabilistic talk and thought). A theory that assumes Reasonability (and Means-End Coherence) yields a model theory appropriate to representing the logical commitments of a certain type of *minimally rational agent*—roughly speaking, an agent (or syntactically specified class thereof) whose decision situation is representable as reasonable (and who takes means appropriate to their ends). If you are in need of a logic of practical talk and thought—a logic of decision—on which canonical forms of deductive practical reasoning (e.g., dominance reasoning) can be represented *as logical*, this is the type of theory you need.

I will suppose is represented as reasoning from (exhaustive) premises about primary preferability to conclusions about deliberative preferability, the latter of which I'll represent using the operator $\bigstar$).

Means-End Coherence (MEC). For all Π, ϕ, a: if a is available in Π and $\Pi[\phi] \vDash a$: $\Pi \vDash \star\phi$ implies $\Pi \vDash \bigstar a$.

MEC says that if ϕ is (primarily) preferred and there is an available action a such that ϕ implies doing a, then a is (deliberatively) preferred. In the Miners case, blocking a shaft is primarily preferred, roughly because it's the only way to save all ten miners. But the available actions are: block A, block B, or block neither; no available action is such that blocking a shaft implies doing that action, and so MEC doesn't require that you deliberatively prefer to block a shaft. To contrast, in (14), MEC does require that you deliberatively prefer smoking, since smoking is (i) primarily preferred and also (ii) identical to an action that is available to you.

[27] The present theory does not quite make good on this: recalling case (12), the following probabilistic inference is deductively licensed when ϕ and ψ are represented as incompatible, but not otherwise.

$$\phi \vee \psi, (\phi \vee \psi) \Rightarrow \blacktriangle\chi \vdash \blacktriangle\chi$$

One, again very direct, possibility for enforcing this is simply to expand the criteria of reasonability, so that a decision state $\Pi = \langle \mathcal{S}, \mathcal{A}, \Pr, \mathrm{Val}_{\preceq} \rangle$ is reasonable (for $(\phi \vee \psi)$) only if $\mathcal{S}[\phi] \cap \mathcal{S}[\psi] = \emptyset$. This raises the question of what, exactly, is unreasonable about representing $\phi \vee \psi$ when you regard ϕ and ψ as compatible. (We might raise a similar question about the decision state against which (1) is evaluated.) More on the nature of this unreasonability (which I will propose to theorize as a kind of incoherence) below.

It is significant, for the sake of comparing the theories, that Bledin (2020), following Hamblin (1958); Groenendijk & Stokhof (1984); Biezma & Rawlins (2012), assumes that a semantically well-formed question (here understood to include an inquisitively interpreted disjunction) is such that its answers *must be* understood as mutually exclusive. Although this assumption *entails* that a decision state Π can satisfy $\phi \vee \psi$ only if $\mathcal{S}[\phi] \cap \mathcal{S}[\psi] = \emptyset$—and thus has basically the same effect as supposing that abstract representations of states of decision are reasonable for semantically well-formed disjunctions, as a matter of definition—I nevertheless do not believe it is warranted, in light of cases like (12).

[28] See (a.o.) Rothschild (2012); Yalcin (2011, 2012a); Moss (2015, 2018).

[29] For an Accuracy-Dominance argument for probabilism, see Joyce (1998). For an Accuracy-Dominance argument for conditionalization, see Briggs & Pettigrew (2020). For an overview of Dutch Book arguments for probabilism, see Hájek (2009). For a Dutch Book argument for conditionalization, see Lewis (1999).

4.3 *"Pure" Deductive Inference*

These remarks raise something of a puzzle: in what sense is an unreasonable agent "committed" or "subject" to an inference, if the logic and model theory are construed as presupposing reasonability, as I've suggested above? I will draw an analogy to Dutch Book arguments to help illustrate my answer. Dutch Book arguments standardly make use of the following thesis about fair prices:

> **Ramsey's Thesis (RT).** "Suppose your credence in X is p. Consider a $\$S$ bet on X [that pays $\$S$ if X and \$0 otherwise]... You are rationally required to pay $\$x$ for this bet, if $x < pS$." (Pettigrew 2019: 4) (cf. also Christensen 2004; Hedden 2013)

Thus, for example, a probabilistically incoherent agent who assigns X credence .6 and Y credence .6 (when Y is logically equivalent to $\neg X$) is required to pay \$.55 for a bet that pays \$1 if X and to pay \$.55 for a bet that pays \$1 if Y. Taken together, these bets logically guarantee a loss of \$.10.

	Y	X
Bet 1	−\$.55	+\$.45
Bet 2	+\$.45	−\$.55

Given RT, it is natural to think of credence as a special kind of conditional attitude: having a .6 credence in X means being (inter alia) *logically committed to regarding Bet 1 as acceptable if it is offered for \$.55*. Having a .6 credence in X *and* a .6 credence in Y means being logically committed to regarding *both* Bets 1 and 2 as acceptable if they are each offered for \$.55. And so having a .6 credence in X *and* a .6 credence in Y means being logically committed to regarding as preferable something that one is also logically committed to not regarding as preferable, namely, the conjunction of Bets 1 and 2.

The "impure" theory of the previous section *does not* account for the type of logical commitment used in such a Dutch Book. That theory provides a way to work out the logical commitments of reasonable states of decision (and syntactically specified classes thereof). But there is *no* reasonable decision state in which one has a .6 credence in X *and* a .6 credence in Y.[30]

The appropriate "relaxation" of the theory seems straightforward: just drop the constraint requiring decision states to be reasonable (e.g., by allowing decision states to be constructed with credal measures Pr such that $\Pr(X) + \Pr(\neg X) > 1$). On the relaxed version of this theory (and in contrast with the theory of the prior section), it would be possible for (the credal component of) a decision state to satisfy both:

(15) a. There's a 60% chance that X.
b. There's a 60% chance that Y.

Reading RT as a thesis about the logical commitments of states of partial belief, any decision state that satisfies both of these sentences will also satisfy:

(16) a. If I'm offered Bet 1 for \$.55, it's preferable to take it.
b. If I'm offered Bet 2 for \$.55, it's preferable to take it.

Thus, supposing you are offered both Bets 1 and 2, you are (on the assumption that you are logically committed to reasoning via modus ponens) logically committed to concluding:

(17) a. It is preferable to take Bet 1 for \$.55.
b. It is preferable to take Bet 2 for \$.55.

So you are committed to regarding it as preferable to take both bets.[31] Of course this contradicts another commitment of yours: since the payout of both bets is negative, you in fact prefer taking *neither* bet to

[30] There is also, to be sure, a fairly clear sense in which this sort of agent shouldn't be regarded as logically committed to this conclusion. An agent in this decision situation *makes a logical error* if she is offered both bets and decides to accept both (given that both bets logically guarantee a loss of \$.10).

[31] To see this more clearly, consider a standard assumption about the logic of strict preference, known as conjunctive expansion, according to which p is preferred to q iff $(p \wedge \neg q)$ is preferred to $(\neg p \wedge q)$. Note first that the agent in our Dutch Book is logically committed to prefer taking Bet 1 to rejecting Bet 2. Given conjunctive expansion taking both Bets 1 and 2 is preferred to rejecting both Bet 1 and Bet 2 iff taking Bet 1 is preferred to rejecting Bet 2.

taking both. And so, given this understanding of logical commitment, any decision state that satisfies (15) is logically committed to a contradiction-in-preference—one we could gloss by saying that, relative to that state, it is and is not preferable to take both bets if you are offered both.[32]

This is the notion of logical commitment—I'll refer to it as the "pure" notion—that philosophers employ when they argue that agents whose credences violate the laws of probability are logically committed to accepting conclusions that are unacceptable from their own point of view. This notion has a specialized theoretical application (for example, when we want to make the point that agents who are not representable with reasonable decision states are incoherent, in the sense of being logically committed to reasoning from premises they would accept to conclusions they would reject).[33]

4.4 Plural Commitment

"Pure" and "impure" understandings of logical commitment each have their role to play.[34] Here I will identify another role for the latter: it has the right characteristics to account for theoretical disagreement about the deductive status of dominance arguments, in contested cases like the Newcomb Problem. I'll then draw on this discussion to say a final word about how I think cases like (1) should be theorized.

In the Newcomb Problem, there are two boxes before you, A and B. You keep what cash is inside any box you open. At the time of your decision, an exceptionally reliable predictor has already acted as follows:

- It put \$1,000 in A.
- It predicted whether you would open A in addition to B.
- If it predicted you would take one box (you would not open A), it put \$1,000,000 in B.
- If it predicted you would take two boxes (you would open both A and B), it put nothing in B.

The decision situation can be represented with the following table:

	Predicted one-boxing	Predicted two-boxing
One-box	\$1,000,000	\$0
Two-box	\$1,001,000	\$1,000

There is a notorious disagreement between Causal and Evidential Decision Theorists about the permissibility of dominance reasoning in this case: EDT rejects it (since EDT permits dominance reasoning only in decision problems where the actions do not provide any evidence about which state is actual, which does not hold in Newcomb), CDT embraces it (since CDT permits dominance reasoning in decision problems where the actions are causally independent of the relevant states, as they are in the Newcomb Problem).

There should be no disagreement about the acceptability of the following, *relative to* or *against* this

[32] The notion of being "logically committed" to a conclusion is perhaps something of a misnomer for the somewhat more permissive notion I mean to invoke here. On the rough notion I have in mind, being logically committed to ϕ amounts to *being in a position to conclude* ϕ (compare Pettigrew 2019). Thus, an agent in a Dutch Book is in a position to conclude something they are also in a position to reject, namely, that the package consisting of both bets is superior to the package consisting of neither. I take the incoherence of this sort of position as given. Here it is worth mentioning that Dutch Book arguments are often (mis-)read as supplying a *prudential* reason to conform one's credences to the laws of probability: roughly, if you don't so-conform, you're liable to become a money pump. This reading is impluasible, as Christensen (2004) emphasizes. It is true that Dutch Book arguments do not provide an *epistemic* (i.e., truth-related) reason to conform one's credences to the laws of probability, but they do nevertheless establish that agents whose credences violate the laws of probabilities are committed to certain forms of logical inconsistency or incoherence.

[33] Decision theorists have puzzled over how to motivate the restriction to well-formed decision problems that characterizes a decision theory like Savage (1972), sometimes deriding them as ad hoc or external (see, e.g., Joyce 1999: Ch. 4). The discussion here is suggestive of what would amount to a Dutch Book argument for this restriction.

[34] A related ambiguity is identified in work on "wide" and "narrow scope" normative requirements (see e.g. Broome 1999; Kolodny 2005). There is a sense in which believing $\neg\neg p$ supports believing p: if you believe $\neg\neg p$, you ought to believe p. But suppose you believe $\neg\neg p$ for *no good reason*. Does it follow (as it seems it should, by modus ponens) that you ought to believe p? Intuitively it does not: indeed, you ought not believe $\neg\neg p$, and since not believing $\neg\neg p$ logically commits you to not believing p, you *ought not* believe p. My best sense is that there are simply two notions of logical commitment at issue in this debate (which parallel the two notions of logical commitment developed in this section and the last). On the latter notion, you are logically committed to believing p, supposing you irrationally believe $\neg\neg p$; on the former notion, you are not.

decision situation. (Here read 'it's better to take two' simply as 'I get more money by taking two'.)

(18) a. If the predictor predicted I'd take one box, it's better to take two.
b. If the predictor predicted I'd I take two boxes, it's better to take two.
c. If the predictor predicted I'd take one box or it predicted I'd take two, it's better to take two.
d. ?So, it is better to take two boxes.

From the premises here, CDT-ers conclude (while EDT-ers reject) that it is (unconditionally) preferable to take both boxes.[35] The central difficulty of EDT's position on the Newcomb Problem, of course, is that the CDT-er's dominance argument is *facially sound* (as assessed from the natural representation of the decision situation represented in the above decision table).[36]

Here I will sketch a way for EDT to resist this assessment (by invoking the impure notion of logical commitment described above).[37] By EDT's lights, this representation of the decision situation violates act-state (evidential) independence. I have said that decision states witnessing violations of act-state (evidential) independence deductively commit an agent in such a state to conclusions that are unacceptable (by EDT's lights), to wit, that it is preferable to take two boxes (and almost certainly walk away with $1,000) than to take one box (and almost certainly walk away with $1,000,000). The proponent of EDT should concede that (18c) is satisfied *when evaluated against such a decision state* (and that such a decision state is logically committed, in the pure sense, to a preference for taking two boxes). A way for the EDT-er to resist this argument's claim to being deductively licensed is to say that:

- The representation of the decision situation that is made salient in the Newcomb Problem is unreasonable (since the states evidentially depend on the actions).
- There is an impure notion of satisfaction, according to which it is not the case that (18c) satisfied in its context of evaluation, even though the relevant decision state meets the condition on decision states semantically encoded in (18c).

EDT can explain the pull of CDT's dominance argument in the Newcomb Problem as follows: although (18c) is strictly speaking unacceptable, it is smoothly evaluated as acceptable relative to the decision state supplied above. "For" EDT-ers, dominance reasoning in the Newcomb Problem is logically akin to dominance reasoning to a preference for smoking (or ignoring the air sirens).

This is *not*, as I'll now explain, to say that competent speakers assess (18c) as *false* in the context associated with the Newcomb Problem. The context provides a decision state that is logically committed (in the pure sense) to (18c), and so there is a sense in which (18c), like (1c), is simply evaluated as true in its context. The proponent of EDT can agree with *all of this* and still consistently maintain that they are not logically committed to a preference for two-boxing (provided they invoke an "impure" notion of logical commitment, on which unreasonable states of decision are strictly speaking without logical commitments).

Here is a slightly more formal spin. Let c be a context that provides a decision state (or class thereof). Let $\vDash$ be the pure (total) satisfaction relation, and let $\vDash_T$ be a satisfaction relation that is defined only for decision states witnessing act-state independence by the lights of decision theory T. Then:

Definition 5. *ϕ is a minimal commitment in c (notation: $\vDash_c \phi$) iff for any c-relevant Π, $\Pi \vDash \phi$.*[38]

Definition 6. *ϕ is a T-commitment in c (notation: $\vDash_{T,c} \phi$) iff for any c-relevant Π, $\Pi \vDash_T \phi$.*

(18c) is a minimal commitment in its context—competent speakers judge it acceptable in that context. (18c) is also a CDT-commitment in this context (given a CDT-compatible understanding of decision states); given

[35] EDT-ers are fond of a Dutch Book-ish argument against Dominance reasoning in Newcomb: if you're so smart why ain't you rich? The thrust of the EDT-er's point is that a logic of decision should favor a course of action that will, with a high degree of probability, make you a millionaire over an action that will, with a high degree of probability, net you only $1,000.

[36] The intuitive soundness of this sort of dominance argument is surely a large part of the reason that authors like Lewis (1981); Egan (2007) (a.o.) are inclined to treat the Newcomb Problem as a *counterexample* to EDT.

[37] Jeffrey (1983)'s formulation of EDT is noted for abandoning the requirement of Savage (1972) that decision states be well-formed. This discussion can be read as suggesting that the EDT-er has reasons (akin to the reason provided by the Dutch Book argument for Probabilism) for invoking a notion of well-formed-ness in their own logic of decision.

[38] The quantifier 'for any c-relevant Π' should be read as requiring a non-empty domain: if no Π is salient or relevant in c, then it is not the case that $\star\phi$ is minimally accepted in c.

CDT, we *are* logically committed to (18c) and therefore to (18d). But neither (18c) nor its negation is an EDT-commitment in this context. That is not because (18c) has one satisfaction-condition in the mouth of the Evidential Decision Theorist and another in the mouth of the Causal Decision Theorist—it doesn't. It is instead because EDT does not regard the decision state provided in the Newcomb Problem as one from which an agent has a "license" to reason deductively, i.e., by dominance.

Like (18c), (1c) is minimally accepted in its context of use. Unlike (18c), (1c) is unratifiable in its context of use, given the following notion of ratifiability.

Definition 7. *ϕ is $\mathcal{T}$-ratifiable in c iff, for some $T \in \mathcal{T}$: $\vDash_{T,c} \phi$.*

Although (1c) is evaluated as acceptable relative to the salient decision state, *no* decision theory permits dominance reasoning in such a decision state (given the dependence, in any relevant sense of that word, of cancer on smoking). So (1c), unlike (18c), is *$\mathcal{T}$-unratifiable* in its context (for any plausible class of decision theories $\mathcal{T}$). This should somewhat dispel the feeling of paradox around (1): although $\vDash_{T,c}$ (1c) implies $\vDash_{T,c}$ (1d) (and T-commitment, more generally, is closed under modus ponens), for *no* plausible T and c resembling the context I described for (1) is it the case that $\vDash_{T,c}$ (1c).

5 Conclusion

This paper grappled with the following puzzle: the context I described for (1) appeared to make available the information in (1c), without deductively licensing the inference to (1d). While this paper considered several ways we could deal with this puzzle, most appeared, in one way or another, to under-generate good deductive inferences. Instead I suggested a theory of deductive inference (and a corresponding notion of logical consequence) on which an inference from Γ to ϕ is deductively licensed in c roughly when Γ represents a reasonable state of logical commitment in c, and $\Gamma \vDash \phi$. In essence I have proposed that the notion of good deductive inference is epistemic, rather than doxastic or informational, in character. Whether a deductive inference is *good* (in context) is a normatively laden question. In the quest to model and systematize our judgments of good and bad deductive inference, the semantics of natural language can only take us so far.

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.

Biezma, María & Kyle Rawlins. 2012. Responding to alternative and polar questions. *Linguistics and Philosophy* 35: 361–406. doi:10.1007/s10988-012-9123-z.

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

Bledin, Justin. 2020. Fatalism and the logic of unconditionals. *Noûs* 54: 126–161. doi:10.1111/nous.12257.

Briggs, R. A. & Richard Pettigrew. 2020. An Accuracy-Dominance argument for conditionalization. *Noûs* 54(1): 162–181. doi:10.1111/nous.12258.

Broome, John. 1999. Normative requirements. *Ratio* 12: 398–419. doi:10.1111/1467-9329.00101.

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

Cantwell, John. 2006. The logic of dominance reasoning. *Journal of Philosophical Logic* 35: 41–63. http://www.jstor.org/stable/30226858.

Cariani, Fabrizio, Magdalena Kaufmann & Stefan Kaufmann. 2013. Deliberative modality under epistemic uncertainty. *Linguistics and Philosophy* 36: 225–259. doi:10.1007/s10988-013-9134-4.

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

Charlow, Nate. 2013c. Conditional preferences and practical conditionals. *Linguistics and Philosophy* 36: 463–511. doi:10.1007/s10988-013-9143-3.

Charlow, Nate. 2015. Prospects for an expressivist theory of meaning. *Philosophers' Imprint* 15: 1–43.

Charlow, Nate. 2016. Decision theory: Yes! Truth conditions: No! In N. Charlow & M. Chrisman (eds.) *Deontic Modality*, 47–81. Oxford University Press.

Charlow, Nate. 2018. Decision-theoretic relativity in deontic modality. *Linguistics and Philosophy* 41: 251–287. doi:10.1007/s10988-017-9211-1.

Christensen, David. 2004. *Putting Logic in its Place: Formal Constraints on Rational Belief*. Oxford: Oxford University Press.

Ciardelli, Ivano. 2018. Questions as information types. *Synthese* 195: 321–365. doi:10.1007/s11229-016-1221-y.

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.

Dummett, Michael. 1964. Bringing about the past. *Philosophical Review* 73: 338–359. doi:10.2307/2183661.

Egan, Andy. 2007. Some counterexamples to causal decision theory. *The Philosophical review* 116: 93–114. http://www.jstor.org/stable/20446939.
Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge: Harvard 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.
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.
Groenendijk, Jeroen & Martin Stokhof. 1984. Studies on the semantics of questions and the pragmatics of answers. http://dare.uva.nl/record/123669. Ph.D. Dissertation, University of Amsterdam.
Hájek, Alan. 2009. Dutch book arguments. In P. Anand, P. Pattanaik & C. Puppe (eds.) *Oxford Handbook of Rational and Social Choice*, 173–195. Oxford: Oxford University Press.
Hamblin, Charles. 1958. Questions. *Australasian Journal of Philosophy* 36: 159–168. doi:10.1080/00048405885200211.
Hedden, Brian. 2013. Incoherence without exploitability. *Noûs* 47: 482–495.
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. 1998. A nonpragmatic vindication of probabilism. *Philosophy of Science* 65: 575–603. doi:10.1086/392661.
Joyce, James. 1999. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.
Kolodny, Niko. 2005. Why be rational. *Mind* 114: 509–563. doi:10.1093/mind/fzi509.
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.
Kratzer, Angelika. 2012. *Modals and Conditionals*. Oxford: Oxford University Press.
Lewis, David. 1981. Causal decision theory. *Australasian journal of philosophy* 59: 5–30. doi:10.1080/00048408112340011.
Lewis, David. 1999. Why conditionalize? In *Papers in Metaphysics and Epistemology*, 403–407. Cambridge University Press. doi:10.1017/CBO9780511625343.024.
Moss, Sarah. 2015. On the semantics and pragmatics of epistemic modals. *Semantics and Pragmatics* 8: 1–81.
Moss, Sarah. 2018. *Probabilistic Knowledge*. Oxford: Oxford University Press.
Pettigrew, Richard. 2019. On the expected utility objection to the dutch book argument for probabilism. *Noûs* doi:10.1111/nous.12286.
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/.
Rothschild, Daniel. 2012. Expressing credences. *Proceedings of the Aristotelian Society* CXII, Part 1: 99–114. doi:10.1111/j.1467-9264.2012.00327.x.
Savage, Leonard J. 1972. *The Foundations of Statistics*. Dover.
Schulz, Moritz. 2018. Modus ponens under the restrictor view. *Journal of Philosophical Logic* 47: 1001–1028. doi:10.1007/s10992-018-9459-0.
Stalnaker, Robert. 1975. Indicative conditionals. *Philosophia* 5: 269–286. doi:10.1007/BF02379021.
Stojnić, Una. 2017. One's modus ponens: Modality, coherence and logic. *Philosophy and Phenomenological Research* 95: 167–214. doi:10.1111/phpr.12307.
Veltman, Frank. 1996. Defaults in update semantics. *Journal of Philosophical Logic* 25: 221–261. doi:10.1007/BF00248150.
Willer, Malte. 2012. A remark on iffy oughts. *Journal of Philosophy* 109: 449–461. doi:10.5840/jphil2012109719.
Yalcin, Seth. 2011. Nonfactualism about epistemic modality. In A. Egan & B. Weatherson (eds.) *Epistemic Modality*, 295–332. Oxford: Oxford University Press.
Yalcin, Seth. 2012a. Bayesian expressivism. *Proceedings of the Aristotelian Society* CXII, Part 2: 123–160. doi:10.1111/j.1467-9264.2012.00329.x.
Yalcin, Seth. 2012b. A counterexample to modus tollens. *Journal of Philosophical Logic* 41: 1001–1024. doi:10.1007/s10992-012-9228-4.