Considering the Exceptions

Ryan Simonelli

October 8, 2020

0 Introduction

According to existing accounts of indicative conditionals, any argument of the following form is valid:[1]

$$\frac{\varphi \to \psi \quad (\varphi \wedge \psi) \to \chi}{\varphi \to \chi}$$

Here, I present one main counterexample, three auxiliary counterexamples, and a general procedure for generating indefinitely many more counterexamples, to show that there exist invalid arguments of this form. I argue that this data poses serious problems to variably strict accounts of conditionals (Lewis 1973;

[1] This includes accounts of truth-conditional varieties (Lewis 1973; Stalnaker 1968), including even McGee's (1985) semantics that invalidates modus ponens, suppositional varieties (Adams 1966, 1975; Edgington 1995, 2001), and dynamic varieties (von Fintel 2001; Gillies 2004, 2007, 2009; Willer 2017). Some of these accounts, such as Lewis (1973), von Fintel (2001), and Gilles (2007), are proposed for counterfactuals, but they can be straightforwardly carried over as accounts of indicatives. In what follows, I will talk as if they are simply proposed as accounts of indicatives. Following Rothschild (2020), I do not take Kratzer's (1981, 1986) restrictor view to be a view of conditionals competing with these other views, but, rather, a view of the meaning of "if" that is compatible with multiple views of the meanings of full "If . . . then . . ." sentences, but, in any case, it should be clear that it validates arguments of this form in much the same way the Lewis/Stalnaker account does. I should note from the outset that I do not explicitly consider suppositional views here, but it should be clear from the discussion of the variably strict view that there is a structurally analogous problem here. Just to be clear that such views fall within that target range, note that Adams (1975) describes this inference pattern as "universally probabilistically sound in that the uncertainty of the conclusion can never exceed the sum of the uncertainty of the premises," (22).

Stalnaker 1968), as such accounts are structurally unable to accommodate it. Dynamic strict accounts (von Fintel 2001; Gillies, 2007; Willer 2017), however, are a different story. While existing dynamic strict accounts do not accommodate the data, they are in principle able to, and I propose a modified dynamic strict account, drawing from von Fintel (2001), that does. The key modification is this: whereas existing dynamic strict accounts take into account only the effects of conditional *antecedents* in changing the semantic context, the data shows that we must also take into account the effects of conditional *consequents* in changing the semantic context. After proposing a modified dynamic strict account, I argue that pragmatic alternatives (Moss 2012, Lewis 2017) fall short. I conclude by considering a philospohical upshot of the semantic conclusions reached here: we cannot really make sense of the idea of closing an idealized rational agent's beliefs under defeasible implication, as the formal models proposed by some authors, such as Horty (2007a, 2007b, 2012), suggest we can.

1 The Case(s)

Suppose you're at a very large party, with hundreds of people, and an open bar where one can get anything at all that one might want to drink. A friend of yours has told you that there is some woman named Maddy at this party. You've never met Maddy, and you don't know anything about her other than that she's at the party. Now, consider the following sentence:

1. If Maddy's drinking a beer, then she's drinking an alcoholic beverage.

(1) seems true. If Maddy's drinking a beer, then she's drinking an alcoholic beverage. Now, consider the following sentence:

2. If Maddy's drinking a beer and she's drinking an alcoholic beverage, then she's not drinking an O'Doul's.

(2) also seems true. If Maddy's drinking a beer and she's drinking an alcoholic beverage, then, since, if she's drinking an O'Doul's, she's drinking a non-alcoholic beverage, she's not drinking an O'Doul's.[2] Finally, consider the following sentence:

3. If Maddy's drinking a beer, then she's not drinking an O'Doul's.

(3) doesn't seem true. It seems that the truth of (3) would rule out the possibility that Maddy's drinking a beer and the beer she's drinking is an O'Doul's. Since you don't know anything about Maddy, and she may very well be a teetotaler, such a possibility surely can't be ruled out. Accordingly, (3) is not true. Indeed, we may well say that it's false.

Taking these intuitions at face value, it follows from the truth of (1), the truth of (2), and the falsity of (3), relative to the context you've just considered, that the following argument is invalid:

The Maddy Argument

1. If Maddy's drinking a beer, then she's drinking an alcoholic beverage.
2. If Maddy's drinking a beer and she's drinking an alcoholic beverage, then she's not drinking an O'Doul's.

So, 3. If Maddy's drinking a beer, then she's not drinking an O'Doul's.

This argument might be formally represented as follows:

$$\frac{b \rightarrow a \quad (b \wedge a) \rightarrow \neg o}{b \rightarrow \neg o}$$

[2] If you didn't already know, O'Doul's is a popular non-alcoholic beer.

Since it is invalid, it follows that there are invalid arguments of the following form:

$$\frac{\varphi \to \psi \quad (\varphi \wedge \psi) \to \chi}{\varphi \to \chi}$$

The principle of inference displayed by this argument schema might be called *Cumulative Transitivity*.[3] Existing semantic proposals for indicative conditionals tell us that Cumulative Transitivity is a valid principle of inference. So, given the meanings of "If . . then . . ." sentences, one should be able to reason from the truth of (1) and the truth of (2) to the truth of (3). However, as the Maddy Argument demonstrates, one cannot do that.

This is the basic bit of data with which I'll be working here. For most people, the intuitions of the truth of (1), the truth of (2), and the falsity of (3), when these sentences are presented in the order in which I've just presented them, relative to the context I've just specified, are quite strong.[4] Still, for certain people—for instance, those for whom non-alcoholic beer comes quickly to mind—this particular example might not work. Perhaps you were one of those people. No matter. Nothing hangs on this particular example. The Maddy Argument is just one example of an argument in which the schema of Cumulative Transitivity fails. Here are three more arguments capable of demonstrating the same general failure:

The Norm Argument:

1n. ✓ If Norm gave Maddy a rose, then he gave her a red flower.

[3] To keep the terminology here consistent with the terminology deployed in discussions of substructural logics (for instance, Makinson (2005) and Brandom (2018)), in which "Cumulative Transitivity" is taken to pick out a *structural* rather than *operational* principle, we might want to call this principle "*Conditionalized* Cumulative Transitivity," but I'll just stick with the shorter name here.

[4] Or, at least, people who are competent with the relevant vocabulary, who know what an O'Doul's is.

2n. ✓ If Norm gave Maddy a rose and he gave her a red flower, then he didn't give her a white rose.

So, 3n. # If Norm gave Maddy a rose, then he didn't give her a white rose.

The Frank Argument:

1f. ✓ If Frank is a fish, then he can't walk on land.

2f. ✓ If Frank is a fish and he can't walk on land, then he's not a mudskipper.

So, 3f. # If Frank is a fish, then he's not a mudskipper.

The Bella Argument:

1b. ✓ If Bella is a bird, then she flies.

2b. ✓ If Bella is a bird and she flies, then she's not a penguin.

So, 3b. # If Bella is a bird, then she's not a penguin.

I'll leave it as an exercise for the reader to specify, for each of these arguments, a context against which speakers will reliably deem (1) to be true, (2) to be true, and (3) to be false. It's not hard, and the fact that it's not means that there is a class of arguments exemplifying the schema of Cumulative Transitivity that speakers intuitively take to be invalid.

Not only is it not hard to generate these arguments, it also not hard to specify a general procedure for generating them. Take some general kind *K* (such as *beer*, *rose*, *fish*, or *bird*), instances of which generally have some feature *F* (such as *being alcoholic*, *being red*, *being unable to walk on land*, or *being able to fly*). Now, find a sub-kind *K*′ (such as *O'Doul's*, *white rose*, *mudskipper*, or *penguin*) that is exceptional with respect to *F*, such that, while that *K*s generally have feature *F*, *K*′s, while still being *K*s, have a materially contrary feature F^* (such as *being non-alcoholic*, *being white*, *being able to walk on land*, or *being unable to fly*), which precludes them from having feature *F*. If you do that, you'll generally have one of these triads, since there will generally be (1) a licit inference from the proposition ascribing K to proposition ascribing F, (2) a licit inference from the conjunction of proposition ascribing K and the proposition ascribing F to the negation of the proposition ascribing K', but (3) an illicit inference from the proposition ascribing K to the negation of the proposition ascribing K'.

These failures of Cumulative Transitivity pose a serious problem for the standard truth-conditional semantics for indicative conditionals, owed to Lewis (1973) and Stalnaker (1968).

2 The Standard Truth-Conditional Account

On a truth-conditional semantic theory, to know the meaning of a sentence is to know what the world would have to be like in order for that sentence to be true. Truth-conditions are standardly thought of in a possible-worlds framework. On such a framework, we think of the truth-conditions of a sentence in terms of the set of the ways for the world to be, among all the ways it could possibly be, such that, if the world is any of those ways, that sentence is true.[5] Since the truth-conditions of many sentences depend on features of the context in which they are uttered, we assign these truth-conditions—sets of possible worlds—relative to contexts.[6] For an atomic sentence p, we do this directly, as follows:

$$\mathbf{S}_{\mathcal{A}} \text{ (truth-conditional)} : [\![p]\!]^c = \{w \mid p \text{ is true in } w\text{, relative to } c\}$$

Once we've assigned semantic values to atomic sentences in this way, we can

[5] We can leave the notion of a possible world intuitive and informal for our purposes here, thinking of a possible world as just a completely determinate way for the world as a whole to be. All that is technically required is that, for each atomic sentence p and world w, p is either true in w or false in w, and not both true and false in w.

[6] If one wants to build context-dependency into semantic values, one might, following Kaplan (1989), model semantic values as context-invariant functions from contexts to propositions. This extra complexity can be ignored here.

recursively assign semantic values to conjunctions and negations in terms of the operations of set subtraction and intersection as follows:

$\mathbf{S}_{\neg}$ (truth-conditional) : $[\![\neg\varphi]\!]^c = W - [\![\varphi]\!]^c$
$\mathbf{S}_{\wedge}$ (truth-conditional) : $[\![(\varphi \wedge \psi)]\!]^c = [\![\varphi]\!]^c \cap [\![\psi]\!]^c$

Negation and conjunction are widely thought to be the most unproblematic logical operators of natural language, and practically zero proponents of truth-conditional semantic theories challenge these assignments of semantic values. We will follow suit here, concerning ourselves solely with the semantics for the conditional offered by the truth-conditional theorist.

The textbook "variably strict" semantics for conditionals comes from Lewis (1973) and Stalnaker (1968, 1975).[7] The basic idea is that a sentence of the form $\varphi \rightarrow \psi$ is true in a world w and context c just in case, given a relation of "closeness" determined by c, all of the "closest" worlds in which φ is true are worlds in which ψ is true. For indicative conditionals, the closest worlds to w in a context c might be thought of as the worlds that are most like the way that one expects w to be in c.[8] Officially, we suppose that a context c supplies an ordering relation between worlds, $\leq$, such that, for each world w, $w' \leq_w w''$ just in case w' is as much or more like the way one expects w to be in c than w''. We can then define, for each world w, a function, $\min_{\leq_w}$, that takes a sentence φ, and returns

[7] For a textbook presentation see, for instance, von Fintel and Heim (2011) 63-66. It's worth noting that Lewis himself only endorsed this semantics for counterfactual conditionals, endorsing a truth-functional analysis of indicatives.

[8] On the way I am thinking about things here, "closeness," at least for indicatives, must be at least a partly epistemic, rather than purely metaphysical, matter. I take it that, in the context of such a semantics it's best to think of the "actual world," which figures into the semantics and is as close to itself as any world, as a *representation* of the actual world, rather than the actual world itself. Really, then, one is assigning conditions for the judgment of the truth of sentences—acceptability conditions, rather than truth-conditions per se. I think this is the right way to think about truth-conditional semantics in general, but I'll bracket this fundamental issue here to the extent that I can.

the set of minimally distant φ-worlds, relative to w, as follows:[9]

> $\min_{\leq_w}(\varphi) = \{w' \mid w' \in [\![\varphi]\!]^c$ and, for all w'', if $w'' \in [\![\varphi]\!]^c$, then $w' \leq_w w''\}$

Having defined such a function, the semantic value of a conditional sentence $\varphi \to \psi$ is defined as follows:

> $\mathbf{S}_{\to}$ (truth-conditional) : $[\![\varphi \to \psi]\!]^c = \{w \mid \min_{\leq_w}(\varphi) \subseteq [\![\psi]\!]^c\}$

So the semantic value of a conditional of the form $\varphi \to \psi$, relative to a context c, is the set of worlds such that, for each world w in this set, the closest φ-worlds, relative to w, are also ψ-worlds.

One of the main merits of the variably strict semantics, compared to its truth-functional and fully strict alternatives, is that it does not validate several arguments that are intuitively invalid for natural language conditionals. For a crucial such case, consider that, on truth-functional and fully strict accounts of conditionals, any argument of the following form is valid:

$$\frac{\varphi \to \psi}{(\varphi \wedge \chi) \to \psi}$$

This is a bad result, since there are intuitively invalid arguments of this form, such as the following:

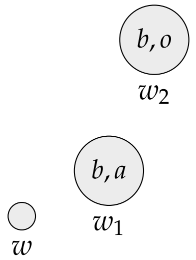
1. ✓ If Maddy's drinking a beer, then she's drinking an alcoholic beverage.

So, 4. # If Maddy's drinking a beer and she's drinking an O'Doul's, then she's drinking an alcoholic beverage.

In contrast to truth-functional and fully strict accounts, a variably strict account is able to accommodate the invalidity of this argument. To see this, let b be

[9] I make neither the uniqueness nor the limit assumption in this presentation. I'll rely somewhat not making the uniqueness assumption in the presentation that follows, but I don't wish to take a stand on the limit assumption one way or the other.

“Maddy is drinking a beer,” a be “Maddy is drinking an alcoholic beverage,” o be “Maddy’s drinking an O’Doul’s,” and consider the following diagram:



Here, the closest world to w in which Maddy’s drinking a beer is one in which she’s drinking an alcoholic beverage, but the closest world to w in which she’s both drinking a beer and drinking an O’Doul’s is not, so (1) is true, but (4) is false.

In a similar way, a variably strict semantics invalidates what we might call “Simple Transitivity”:

$$\frac{\varphi \to \psi \quad \psi \to \chi}{\varphi \to \chi}$$

which also has intuitive counterexamples.[10] However, a variably strict semantics will inevitably validate the principle that we’ve called *Cumulative* Transitivity:

$$\frac{\varphi \to \psi \quad (\varphi \wedge \psi) \to \chi}{\varphi \to \chi}$$

To see this, note first that, on a truth-conditional conception of meaning, an argument of the form

$$\frac{\varphi \quad \psi}{\chi}$$

[10] This principle is usually just called “Transitivity” or “Hypothetical Syllogism.” For a discussion, see von Fintel and Heim (2011), 64-66.

is valid just in case, for every world w and context c, if φ is true in w at c, and ψ is true in w at c, then χ is true in w at c. Now consider an arbitrary world w and context c. If $\varphi \to \psi$ is true in w at c, then the closest φ-worlds, relative to w, are also ψ-worlds. If $(\varphi \wedge \psi) \to \chi$ is true in w at c, then the closest worlds, relative to w, that are both φ-worlds and ψ-worlds are also χ-worlds. Since the closest φ-worlds are ψ-worlds, and the closest worlds that are both φ-worlds and ψ-worlds are χ-worlds, it follows that the closest φ-worlds are χ-worlds. So, $\varphi \to \chi$ is true in w at c. Thus, any argument of the above form is valid. Since the Maddy Argument is an invalid argument of this form, as demonstrated by the fact that we've specified a context relative to which (1) is true, (2) is true, and (3) is false, the standard truth-conditional account fails.

3 The Dynamic Strict Account

In the last few decades, several authors have proposed a dynamic alternative to thinking about meaning solely in terms of truth-conditions, and a new theory of conditionals has emerged from this paradigm.[11] The basic idea of a dynamic semantics is this: rather thinking of the meaning of a sentence solely in terms of the conditions under which it is true, we can think of the meaning of a sentence, at least in part, in terms of its potential, when uttered in a given context, to change (or "update") that context. In a slogan, the meaning of a sentence is, at least in part, its context change potential. Of course, the stronger, and far catchier, slogan gets rid of the parenthetical "at least in part," but I do not intend to argue for this stronger slogan here, and I do not think I need to in order to encounter

[11] For seminal pieces of dynamic semantics, see Heim (1982), Groenendijk and Stokhof (1991), and Veltmann (1996). Unsurprisingly, there has been some push-back against the dynamic turn from proponents of the truth-conditional paradigm. See, for instance, Dever (2013) and Lewis (2014).

significant resistance from the orthodoxy. Even the weaker slogan marks a radical divergence from the truth-conditional paradigm, bringing aspects of what is normally relegated to the pragmatics—the effect of the utterance of a sentence on a discursive context—into the semantics proper. This, I am going to suggest, is precisely what is needed to accommodate our data.

The basic view I'll propose to accommodate our data is based on three intuitive ideas. First, conditionals are evaluated relative to a set of possibilities that are "in view." Second, a conditional is true just in case, for every possibility in the set possibilities in view in which the antecedent holds, the consequent holds. Third, the consideration of some conditionals can function to expand the set of possibilities that one has in view when one evaluates the truth of that conditional. So, conditionals are strict conditionals over a set of accessible possibilities that they themselves have the potential to change. The account of conditionals based on these three ideas is accordingly called the "dynamic strict" account of conditionals, as conditionals are treated as strict conditionals over a dynamically evolving set of possibilities. The basic distinction the version of the dynamic strict account to be proposed here and existing versions of the dynamic strict account (von Fintel 2001, Gillies 2007, Willer 2017) is that existing dynamic strict accounts consider only the effects of conditional *antecedents* in expanding the set of possibilities considered for the evaluation of the conditional, whereas the account proposed here will also consider the effects of conditional *consequents* in expanding the set of possibilities. Before I officially propose the modified account that I will endorse here, let me start by presenting the dynamic strict account as it is endorsed by proponents today and showing how it does not accommodate our data.

There are different formal frameworks in which the dynamic strict account can be presented. For ease of exposition, I'll present it here in the framework proposed by von Fintel (2001), with slight modification, but little hangs on this decision for our purposes here.[12] As with the variably strict account, we once again suppose that a context c supplies an ordering relation between worlds, $\leq$, such that, for each world w, $w' \leq_w w''$ just in case w' is more like the way one expects w to be in c than w''. We additionally take a context c to include an accessibility function f_c, which takes a world w and selects a corresponding set of accessible worlds, those that we've informally described as those that are "in view" of a participant in that context, which von Fintel calls the "modal horizon." Von Fintel's proposal is that conditionals have a semantics that contains both a dynamic component and a truth-conditional component. First, they have the potential to update the accessibility function by making it select additional worlds, expanding the modal horizon. Specifically, von Fintel proposes that conditionals expand the modal horizon by making the accessibility function select the closest worlds in which the antecedent holds. Then, they have strict truth-conditions relative to this updated accessibility function, being true just in case all the antecedent-worlds in the modal horizon are consequent-worlds. Where φ and ψ are neither modals nor conditionals, von Fintel's dynamic strict account can be put as follows:

S$_{\rightarrow}$ (Dynamic Strict):

a **Context Change Potential:** $f_{c[\varphi \rightarrow \psi]} = \lambda w. f_c(w) \cup \{w' \in [\![\varphi]\!]^c$:

[12] The framework here differs in detail from the "spheres"-based framework proposed by Gillies (2007), though they make basically the same predictions. Willer (2013) motivates a more complicated dynamic variant of the spheres framework which is developed in Willer (2017). The basic proposal that I make here is implementable in any of these frameworks, and the question which framework one should opt for hangs on issues outside the scope of this paper. For that reason, I do not take the time to develop this system here so that it is able to accommodate embedded conditionals and conditionals containing modal operators, since I take it that an adequate development will require abandoning the truth-conditional dynamic strict model considered here for a more complex dynamic model, which would distract from the main point of this paper.

$\forall w'' \in [\![\varphi]\!]^c : w' \leq_w w''\}$

b **Truth-Conditions:** $[\![\varphi \rightarrow \psi]\!]^{c,w} = 1$ iff $\forall w' \in f_{c[\varphi\rightarrow\psi]}(w) \cap [\![\varphi]\!]^c$: $[\![\psi]\!]^{c,w'} = 1$

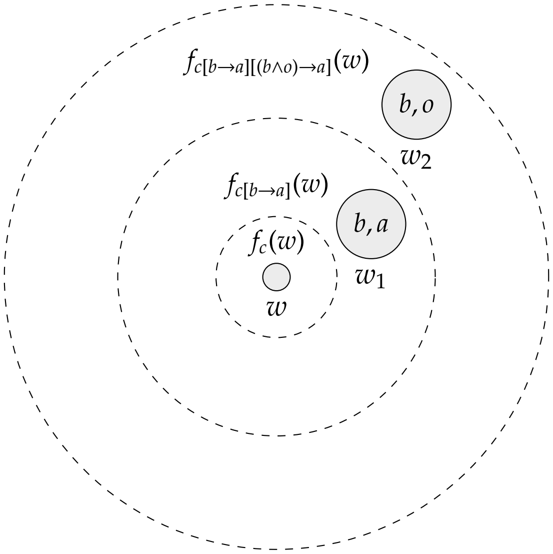
On this account, the context change potentials of conditionals enables us to accommodate the same invalidities that the variably strict semantics accommodates, but to do so while retaining a strict analysis of their truth-conditions.

Consider again the invalidity of Strengthening the Antecedent, as demonstrated by the following argument:

1. ✓ If Maddy's drinking a beer, then she's drinking an alcoholic beverage.

So, 4. # If Maddy's drinking a beer and she's drinking an O'Doul's, then she's drinking an alcoholic beverage.

To see how this argument comes out invalid on the dynamic strict account, consider the following diagram:


$f_{c[b\rightarrow a][(b\wedge o)\rightarrow a]}(w)$
b, o
w_2
$f_{c[b\rightarrow a]}(w)$
b, a
$f_c(w)$
w_1
w

Here, updating the original context c with $b \rightarrow a$ expands the modal horizon to include the closest b-worlds, and, since all of the b-worlds in this updated modal horizon are a-worlds, $b \rightarrow a$ comes out as true relative to $c[b \rightarrow a]$. However, updating $c[b \rightarrow a]$ with $(b \wedge o) \rightarrow a$ expands the modal horizon to include the closest $(b \wedge o)$-worlds, and these worlds aren't a-worlds, so $(b \wedge o) \rightarrow a$ comes out false relative to $c[b \rightarrow a][(b \wedge o) \rightarrow a]$.

Proponents of the dynamic strict account have argued that it has significant virtues over its variably strict alternative.[13] However, at least as it stands, it does not help in dealing with the data that concerns us here. On the dynamic conception of meaning against which the dynamic strict account is proposed, an argument of the form

$$\frac{\varphi \quad \psi}{\chi}$$

is valid just in case, for every world w and context c, if φ is true in w at $c[\varphi]$, and ψ is true in w at $c[\varphi][\psi]$, then χ is true in w at $c[\varphi][\psi][\chi]$.[14] So, an argument is valid just in case, when the premises are successively considered, the context evolving accordingly, and they are all judged to be true, and the conclusion is judged to be true. It follows from these definitions that any argument of the following form is valid:

$$\frac{\varphi \rightarrow \psi \quad (\varphi \wedge \psi) \rightarrow \chi}{\varphi \rightarrow \chi}$$

Take an arbitrary world w and context c, containing an accessibility function f_c and ordering relation $\leq$, and suppose that $\varphi \rightarrow \psi$ is true in w and at $c[\varphi \rightarrow \psi]$

[13] The main virtue of the dynamic strict account, according to proponents such as von Fintel, Gillies, and Willer, is that it is able to accommodate "Reverse Sobel Sequences." Proponents of the dynamic strict account, such as Moss (2012) and Lewis (2017), respond that these sequences are to be dealt with pragmatically. We'll consider how the analogues of these responses fair in response to the data here in §5.

[14] Note that this is not quite von Fintel's definition. On von Fintel's definition, an argument of this form is valid just in case, for every world w and context c, if φ is true at c, and ψ is true at $c[\varphi]$, then χ is true at $c[\varphi][\psi]$, (2001, 142). While this suffices for the data with which von Fintel concerns himself, it will prove crucial in dealing with the data that concerns us here that the updating effects of a sentence are always processed before evaluation of the truth of that sentence.

and $(\varphi \wedge \psi) \rightarrow \chi$ is true in w at $c[\varphi \rightarrow \psi][(\varphi \wedge \psi) \rightarrow \chi]$. Since $\varphi \rightarrow \psi$ is true in w and at $c[\varphi \rightarrow \psi]$, all of the φ-worlds in $f_{c[\varphi \rightarrow \psi]}(w)$, which includes all the worlds that $f_c(w)$ includes and, in addition, includes all the closet φ-worlds, are ψ-worlds. Now, updating $c[\varphi \rightarrow \psi]$ with $(\varphi \wedge \psi) \rightarrow \chi$ would function to bring into view the closest $(\varphi \wedge \psi)$-worlds, if they aren't already included in $f_{c[\varphi \rightarrow \psi]}(w)$. But, since the closest φ-worlds are already included in $f_{c[\varphi \rightarrow \psi]}(w)$, and all of these worlds are ψ-worlds, the closest $\varphi \wedge \psi$-worlds *are* already included in $f_{c[\varphi \rightarrow \psi]}(w)$. Accordingly, $f_{c[\varphi \rightarrow \psi][(\varphi \wedge \psi) \rightarrow \chi]}(w) = f_{c[\varphi \rightarrow \psi]}(w)$. Now, since $(\varphi \wedge \psi) \rightarrow \chi$ is true in w at $c[\varphi \rightarrow \psi][(\varphi \wedge \psi) \rightarrow \chi]$, all of the $\varphi \wedge \psi$-worlds in $f_{c[\varphi \rightarrow \psi][(\varphi \wedge \psi) \rightarrow \chi]}(w)$ are χ-worlds. Finally, we consider $\varphi \rightarrow \chi$ relative to $c[\varphi \rightarrow \psi][(\varphi \wedge \psi) \rightarrow \chi][\varphi \rightarrow \chi]$. $f_{c[\varphi \rightarrow \psi][(\varphi \wedge \psi) \rightarrow \chi][\varphi \rightarrow \chi]}(w)$ includes all the worlds that $f_{c[\varphi \rightarrow \psi][(\varphi \wedge \psi) \rightarrow \chi]}(w)$ includes, and, in addition, includes the closest φ-worlds, but, since these worlds are already included in $f_{c[\varphi \rightarrow \psi]}(w)$, $f_{c[\varphi \rightarrow \psi][(\varphi \wedge \psi) \rightarrow \chi][\varphi \rightarrow \chi]}(w) = f_{c[\varphi \rightarrow \psi][(\varphi \wedge \psi) \rightarrow \chi]}(w) = f_{c[\varphi \rightarrow \psi]}(w)$. Now we just check whether all the φ-worlds in the modal horizon are χ-worlds. Since all the φ-worlds are ψ-worlds, and all the worlds that are both φ-worlds and ψ-worlds are χ-worlds, all the φ-worlds are χ-worlds. So, $\varphi \rightarrow \chi$ is true in w at $c[\varphi \rightarrow \psi][(\varphi \wedge \psi) \rightarrow \chi][\varphi \rightarrow \chi]$. Thus, any argument of the above form is valid. Since the Maddy Argument is an invalid argument of this form, the dynamic strict account, at least as it stands, fails.

The way in which the dynamic strict semantics goes wrong in validating the Maddy Argument is quite clear: neither (2) nor (3) actually function to change the modal horizon. Given that (1) is true, there is no mechanism for (2) or (3) to expand the modal horizon. (1) updates the context so that the modal horizon includes all the closest beer-worlds. Since, (1) is true, all of the beer-worlds are alcoholic beverage-worlds, so the modal horizon already includes all the closest beer and alcoholic beverage-worlds. Accordingly, an update with (2) idles, and, since the antecedent of (3) is the same as that of (1), so does an update with (3). Clearly, what needs to happen is that an update (2) or (3) needs to bring into view some O'Doul's worlds, so that (3) will be judged to be false, relative to the updated context against which it ends up being considered. But that's not what happens on the existing version of the dynamic strict account. Unlike the truth-conditional account, however, which is structurally unable to accommodate the data, the dynamic account can easily be modified to accommodate the data.

4 The Modified Dynamic Strict Account

The data here calls for a modification of the context change potentials of conditionals. Now, I have already said what I take the required modification to be: conditional *consequents* must function to bring possibilities into view, in addition to conditional antecedents. However, given what I've said so far, this conclusion might seem unwarranted. One might think that what is doing the work is not the consequent of (2) and (3), but the conjunctive antecedent of (2). There is, as it turns out, a principled reason to maintain that the conjunctive antecedent of (2) may actually function to bring into view possibilities in which Maddy's drinking an O'Doul's. Suppose the original context includes no possibilities in which Maddy's drinking a non-alcoholic beer. In such a case, the conjunctive antecedent in (2) would violate (the conjunctive analogue of) Hurford's (1974) constraint, the second conjunct being redundant, given the first.[15] If we modify the existing dynamic strict account so that a conditional not only semantically presupposes the possibility of its antecedent but also semantically presupposes the semantic presuppositions of its antecedent, we can plausibly get a context

[15] Hurford's constraint is originally articulated as a constraint on disjunctions, but it is straightforwardly extended to conjunctions.

change potential for (2) that invalidates the Maddy Argument.[16] In considering (2), the potential violation of Hurford's constraint forces the hearer to add to the modal horizon the closest possibilities in which Maddy's drinking a non-alcoholic beer, for it is the elimination of such possibilities that give the second conjunct a function after the first. Since O'Doul's is among the most popular kinds of non-alcoholic beers, the closest possibilities in which Maddy's drinking a non-alcoholic beer includes some in which she's drinking an O'Doul's, and the presence of these possibilities defeats (3).

I don't intend to reject this proposal entirely—the conjunctive antecedent of (2) may well bring into view possibilities in which Maddy's drinking a non-alcoholic beer for the reasons just stated. At the very least, this seems to be a possibility worth exploring. However, this can't be all that's going on in the case of the Maddy Argument. Suppose, instead of presenting (1), (2), and (3), relative to the context originally specified, I skipped (2), presenting (1) and then jumping straight to (3). In this case, you would have still judged (1) to be true and (3) to be false. Since (1) and (3) have the same antecedent, it's got to be the consequent of (3) that is functioning to bring O'Doul's-worlds into view. Somehow, considering a conditional with the consequent that Maddy's not drinking an O'Doul's brings into view worlds in which she is. The minimal modification required to deliver the right results, accordingly, is simply one that

[16] For some additional evidence for this proposal, consider the following sentence:

5. If Maddy's drinking a beer and she's either drinking an O'Doul's or an alcoholic beer, then she's drinking an alchohlic beverage.

After judging (1) to be true, speakers will generally not judge (5) to be true, but that's not what the unmodified dynamic strict account predicts. The closest worlds in which the antecedent is true will be worlds in which the first conjunct is true and the second disjunct of the second conjunct is true, but these world's aren't O'Doul's-worlds. On this proposal, the addition to the modal horizon of O'Doul's worlds is explained by the fact that a disjunction semantically presupposes the possibility of both of the disjuncts, so O'Doul's-worlds get added to the modal horizon against which (5) is judged to be false.

makes this the case:

S$_{\rightarrow}$ (Dynamic Strict, Modified):

a **Context Change Potential:** $f_{c[\varphi\rightarrow\psi]} = \lambda w.f_c(w) \cup \{w' \in [\![\varphi]\!]^c : \forall w'' \in [\![\varphi]\!]^c : w' \leq_w w''\} \cup \{w' \in W - [\![\psi]\!]^c : \forall w'' \in W - [\![\psi]\!]^c : w' \leq_w w''\}$

b **Truth-Conditions:** $[\![\varphi \rightarrow \psi]\!]^{c,w} = 1$ iff $\forall w' \in f_{c[\varphi\rightarrow\psi]}(w) \cap [\![\varphi]\!]^c : [\![\psi]\!]^{c,w'} = 1$

On this modified proposal, updating a context with a conditional brings into view the closest possibilities in which the antecedent holds, but also the closest possibilities in which the consequent doesn't hold, and a conditional is true, relative to the updated context, if every possibility in view in which the antecedent holds is a possibility in which the consequent holds.

The minimal modification of the dynamic strict account is motivated by the simple idea that to assert a conditional sentence is, in the fully felicitous case, to assert that a *connection* obtains between the antecedent and the consequent: the truth of the antecedent *ensures* the truth consequent.[17] In a possibility-based framework, one can see there to be such a connection between the antecedent and the consequent only if there are some possibilities in view in which the consequent *doesn't* hold, so that these possibilities that can be seen to be *ruled out* by the holding of the antecedent. Only by seeing that the possibilities in which the consequent doesn't hold are excluded from the set of possibilities in which the antecedent does hold can one see that the antecedent and the consequent are connected in the right way for the conditional to be judged to be true. If there are no possibilities in view in which the consequent doesn't hold, then the antecedent cannot be distinguished from any other sentence as one that ensures the truth

[17] When I speak of "ensurance" here, I mean to be speaking of ensurance relative to a set of salient possibilities, not absolute ensurance relative to the total set of possibilities. Hardly any indicative conditional, I would claim, expresses the latter kind of ensurance.

of the consequent.[18] So, the judgment of the truth of a non-trivial conditional requires, if there aren't any possibilities in view in which the consequent doesn't hold, that such possibilities be brought into view so that they can be seen to be excluded from the set of possibilities in which the antecedent holds. Now, there may well be motivation for a greater modification of the dynamic strict view, for instance, one according to which a conditional consequent not only brings into view possibilities in which it doesn't hold but also possibilities in which it does hold, but I will not consider such a further modification here.[19] Since the minimal modification is motivated, and that is enough to resolve our issue, that's the proposal with which I'll settle here.

Let me now explicitly state how this account resolves our issue. Once again, on the dynamic strict account proposed here, an argument of the form

$$\frac{\varphi \quad \psi}{\chi}$$

is valid just in case, for every world w and context c, if φ is true in w at $c[\varphi]$, and ψ is true in w at $c[\varphi][\psi]$, then χ is true in w at $c[\varphi][\psi][\chi]$. With the modification to the dynamic strict account, it now does *not* follow that any argument of the following form is valid:

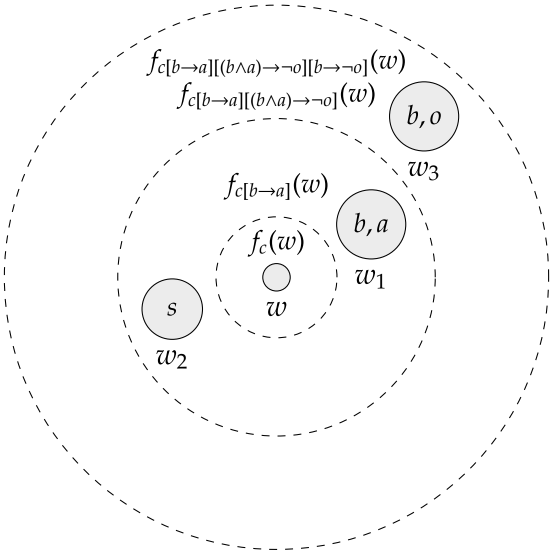
$$\frac{\varphi \rightarrow \psi \quad (\varphi \wedge \psi) \rightarrow \chi}{\varphi \rightarrow \chi}$$

To see this, consider the Maddy Argument. Let b be "Maddy's drinking a beer, a be "Maddy's drinking an alcoholic beverage," o be "Maddy's drinking

[18] I bracket the consideration of mathematical conditionals here, such as "If 48 is divisible by 6, then it is divisible by 3," which is felicitous and has a necessary consequent, since such conditionals pose problems generally for any possibility-based semantics for conditionals rather than posing any specific problem for the semantics proposed here.

[19] One might, for instance, motivate such a view by drawing inspiration from Starr (2014) and arguing that conditionals presuppose that both their antecedent and consequent are live questions, requiring, for both the antecedent and consequent, possibilities in which it does and does not obtain.

an O'Doul's, and s be "Maddy's drinking a soda," and consider the following diagram:


$f_{c[b\to a][(b\wedge a)\to\neg o][b\to\neg o]}(w)$
$f_{c[b\to a][(b\wedge a)\to\neg o]}(w)$
b, o
w_3
$f_{c[b\to a]}(w)$
b, a
w_1
$f_c(w)$
w
s
w_2

Here, updating the original context c with $b \to a$ brings into view the closest b-worlds and not-a-worlds, and $b \to a$ is true, relative to $c[b \to a]$, since all the b-worlds in $f_{c[b\to a]}(w)$ are a-worlds.[20] Now, updating $c[b \to a]$ with $(b \wedge a) \to \neg o$ adds the closet o-world, and $(b \wedge a) \to \neg o$ is true, relative to $c[b \to a][(b \wedge a) \to \neg o]$, since all the $(b \wedge a)$-worlds in $f_{c[b\to a][(b\wedge a)\to\neg o]}(w)$ are $\neg o$-worlds. Nevertheless, since the updated context now includes an o-world, which is a b-world, it's not the case that all of the b-worlds in $f_{c[b\to a][(b\wedge a)\to\neg o][b\to\neg o]}(w)$ are $\neg o$-worlds, so, $b \to \neg o$ is false, relative to $c[b \to a][(b \wedge a) \to \neg o][b \to \neg o]$. Thus, the Maddy Argument is invalid.

[20] It is worth being clear that, on this proposal, if the original modal horizon doesn't include any possibilities in which Maddy's not drinking an alcoholic beverage, $b \to a$ functions to bring the closest such possibilities into view. I assume, reasonably I think, that the closest such possibilities are ones in which she's, say, drinking soda, not one's in which she's drinking an O'Doul's.

5 Pragmatic Alternatives

When I originally presented the Maddy Argument, I presented a context and then presented the three sentences that constitute it: (1), (2), and (3), in that order. For those with standard intuitions, (1) is deemed true, relative to the context against which it gets evaluated, (2) is deemed true, relative to the context against which it gets evaluated, and (3) is deemed false, relative to the context against which it gets evaluated. When we considered how the argument is validated by the variably strict semantics, we worked on the assumption that (1), (2), and (3) were all evaluated relative to the same context. We have now presented a semantics according to argument is invalid in virtue of the fact that the context does not stay fixed, but, rather, shifts through the process of evaluating the sentences presented. When the first premise is evaluated, there are no O'Doul's-worlds in view, but the presentation of (2) updates the context such that, when the conclusion is evaluated, there are O'Doul's-worlds in view, and that explains why it is evaluated as false. It seems clear that taking into account how the context changes through the evaluation of these sentences in this sort of way is necessary to understand what is going on in these cases. One may still ask, however, whether this contextual evolution needs to be understood as a matter of *semantics* or whether it can be understood as a pragmatic matter, with the standard variably strict semantics being maintained. I will consider two attempts to understand this data pragmatically, the first owed to Sarah Moss and the second owed to Karen Lewis, both proposed for related , and argue that they both fall short for the cases that concern us here.

Let me start with the sort of account proposed by Moss. It surely seems that (3) is false, relative to the context against which it gets considered. Of course, things are not always as they seem, and, if one endorses a pragmatic account of the sort endorsed by Moss, one will say that things are not as they seem in this case: though (3) seems false, it is really true.[21] If one goes in for this route, one must then offer an error theory, explaining why things are not as they seem, and that is what Moss does.[22] Adapting Moss's proposal for the cases that concern us here, the thought would be that, though (3) seems false, when it is presented, that is because speakers systematically mistake pragmatic wrongness—the epistemic irresponsibility of uttering (3)—for semantic wrongness—the falsity of (3) itself. Appearances notwithstanding, (3) is true, relative to the context against which it gets considered. It's true because the closest worlds in which Maddy's drinking a beer are worlds in which she's not drinking an O'Doul's, a fact which follows from the truth of (1). However, it seems wrong to utter (3), relative to the context

[21] Saying that either (1) or (2) is false does not seem like a viable option. One could try to say that (1) is false, relative to the original context specified, since Maddy might be drinking an O'Doul's, or one could try to say that (2) is false, since Maddy might be double-fisting, drinking an alcoholic beverage with one hand and an O'Doul's with the other. However, the obvious issue with saying either of these things is that, for any conditional that is not absolutely necessary, there are always going to be some possibilities in which the antecedent is true and the consequent is false. Requiring that there be no such possibilities in order for (1) or (2) to be true would be to give up the variably strict semantics for a fully strict semantics with the obviously untenable consequence, at least in the context of natural language semantics, that the only true conditionals are the necessary ones. (I qualify "untenable" with "in the context of natural language semantics" because, if one shifts from the context of natural language semantics to the context of metaphysics, all bets are off. Hajek (2014), for instance, accepts this conclusion for counterfactuals, claiming that the only true counterfactuals are the strictly necessary ones or the ones with impossible antecedents, and one might try to take the same route here. Such a route, however, requires using "true" in such a way that the vast majority of sentences that speakers take to be true may be in fact false because speakers are systematically "mistaken about the fundamental nature of the universe." If we're using "true" in the way that Hajek does, for instance, then, if we take what Sellars (1963) calls the "manifest image" to be an illusion, we'll take nearly all of the sentences that English speakers utter are false. But then the word "true" would be useless in the context of natural language semantics, since it would not at all track semantic acceptability by competent speakers of our object language. If truth is a relevant property of sentences for the purposes of doing natural language semantics, it must track acceptability. If one insists that *truth* doesn't do this, then let us say that relevant property for natural language semantics is *schmuth*. The point then becomes that it is untenable to assign schmuth-conditions to conditionals in such a way that the only schmue conditionals are the necessary ones, and, surely, no one can deny that.)

[22] The specific cases with which Moss concerns herself are "Reverse Sobel Sequences" involving counterfactuals, the main motivation for previous dynamic strict proposals.

against which it gets considered, because the presentation of (2) functions bring into view worlds in which Maddy's drinking an O'Doul's, and, once these worlds have been made salient, it's epistemically irresponsible to utter this sentence, given that these worlds can't be ruled out. Crucially, even though it is epistemically irresponsible to assent to "If Maddy's drinking a beer, then she's not drinking an O'Doul's" once O'Doul's-worlds have been made salient, it is nevertheless still the case that closest beer-worlds are not O'Doul's-worlds, and so this sentence is still true. The crucial idea of this sort of pragmatic account is to make a sharp distinction between the pragmatic properties of epistemic responsibility and irresponsibility, which are taken to vary depending on which possibilities have been made salient, and the semantic properties of truth and falsity, which are not taken to vary in this way. The basic claim of the error theory, meant to explain speakers' judgment that (3) is false as erroneous, is that speakers systematically mistake the pragmatic property of epistemic irresponsibility for the semantic property of falsity.

Accounts along the lines of that proposed by Moss get much of their plausibility from an apparent analogy with other cases in which it seems wrong to utter certain sentences, not because they're false, but because it's epistemically irresponsible to make these utterances once certain possibilities have been raised to salience. Consider, for instance, the famous case from Dretske (1970). One goes to the zoo and sees a black and white striped animal in the zebra pen. Intuitively, it seems fine for one to say, in this case, "That's a zebra." However, suppose, prior to one's saying this, one's friend raises the possibility that the animal in the zebra pen might be a cleverly disguised mule. If one's friend does this, then, assuming one can't rule out the possibility that it's a cleverly disguised mule, it's going to seem wrong to then say "That's a zebra." Now, unless one's friend can change the world simply by suggesting possibilities, it's not going to be that one's friend has turned what has been said from something true to something false. So, if our semantics is a truth-conditional one, the explanation of why the first utterance seems fine and why the second utterance seems wrong is not going to be a semantic one—it's not going to be that utterance seems wrong in the second case because, in that case, the sentence uttered is false.[23] Rather, it's going to be a pragmatic one—the utterance of "That's a zebra" in the second case seems wrong because, though the sentence is true, it's irresponsible to utter it once the possibility that the animal in the pen is a cleverly disguised mule has been made salient. The proposal is that just the same sort of thing is going on in the cases that concern us here.

While the analogy of the cases that concern us here with the zebra case might seem to help Moss's case, this analogy works only if we are able to maintain that, like the utterance of "That's a zebra" in the zebra case, the possibilities that are contextually salient doesn't affect the truth value of the sentence uttered and only affects the responsibility of uttering of it. Though one might be able to maintain this for the cases of counterfactual conditionals with which Moss concerns herself, it is much harder to maintain for the cases of indicative conditionals that concern us here. Unlike the truth-conditions of non-modal declaratives like "That's a zebra," which do not seem to depend on which possibilities are salient, it seems that the truth-conditions of indicative conditionals

[23] If our semantics is not truth-conditional, then the explanation may well be a semantic one. For instance, if one goes fully dynamic here, thinking of the meanings of entirely in terms of their context change potential, then one will likely substitute the notion of a sentence's being *true*, relative to a context, with the notion of a sentence's being *supported* by a context, where this means that the sentence is informationally redundant with respect to a context, such that updating that context with the sentence does not change that context. If one then proposes a semantics for epistemic "might"s according to which such expressions function to add possibilities to the context, then one can say that the original context supports "That's a zebra," but the original context, once updated with "That might be a cleverly disguised mule" does not.

really do depend on which possibilities are salient. In this regard, indicative conditionals are more like epistemic modals like “That must be a zebra” than non-modal declaratives like “That’s a zebra.” While the truth of “That’s a zebra” doesn’t vary depending on whether your friend has raised the possibility that it’s a cleverly disguised mule, it seems that the truth of “That must be a zebra” does vary in this way. Your friend’s raising the possibility that the animal in the pen is a cleverly disguised mule seems to result in a context relative to which the sentence “That must be a zebra” is not just irresponsible to utter, but false. Likewise, for the cases that concern us here: the truth of “If Maddy’s drinking a beer, then she’s drinking an alcoholic beverage” seems to vary depending on whether the possibility that Maddy’s drinking an O’Doul’s has been made salient. An account along the lines proposed by Moss, which ignores which possibilities are salient in the evaluation of the truth of an indicative conditional, while likely right for non-modal declaratives, and potentially even right for counterfactuals, just seems wrong for epistemic modals and indicative conditionals. Of course, one can always dig in one’s heels here, but, at this point, it’s hard to see what could motivate one to do so.

A different sort of pragmatic account aims to resolve the above issue with Moss’s account by making contextual salience, still taken to be determined by pragmatic mechanisms, feed directly into the semantics so that the truth-conditions of the conditionals under consideration do vary depending on whether certain possibilities are salient. Such an account has been proposed by Karen Lewis (2017). On Lewis’s account, as new possibilities become contextually salient through pragmatic mechanisms, the contextually supplied relation of “closeness” that figures into the variably strict semantics changes, with the newly salient possibilities becoming among the closest worlds. So, when the original context gets updated with (3), which mentions the possibility that Maddy's drinking an O'Doul's, and the truth of (3) gets considered, relative to this updated context, the ordering relation on words that belongs to this context shifts so that worlds in which Maddy's drinking an O'Doul's become among the closest worlds, and so (3) comes out false according to the variably strict semantics. An account along these lines maintains that, while the Maddy Argument is valid, since, relative to any context, if (1) and (2) are true, relative to that context, then (3) is also true relative to that context, as the variably strict semantics dictates, in any actual consideration of the truth of (1), (2), and (3), the context shifts through the course of this evaluation, and so (1) and (2) may be true, relative to the context against which they are considered, but (3) may be false, relative to the context against which it gets considered. This basic idea of this approach is to grant, with the dynamic semanticist, that the semantic properties of sentences change as the discourse context evolves, but to nevertheless maintain, in opposition to the dynamic semanticist, that the dynamics of discourse evolution belongs squarely in the pragmatics, rather than the semantics.

There are two issues I'd like to raise with an account along these lines. The first and most immediate issue with Lewis's approach is that, while it is proposed in defense of the variably strict semantics, it actually undercuts the main motivation for the variably strict semantics. If one goes in for Lewis's approach for the cases that concern us here, it's hard to see why one wouldn't apply it more generally and just maintain a strict semantics in which indicative conditionals are strict conditionals over the set of contextually salient possibilities, which change through pragmatic mechanisms. Recall, one of the main motivations for the variably strict semantics is that it enables us to maintain that arguments like the following are invalid:

1. ✓ If Maddy's drinking a beer, then she's drinking an alcoholic beverage.

So, 4. # If Maddy's drinking a beer and she's drinking an O'Doul's, then she's drinking an alcoholic beverage.

If one endorses Lewis's account for cases like the Maddy Argument, then it's hard to see why one wouldn't take the same line here, maintaining that this argument is valid but appears invalid in virtue of the fact that the context shifts from the evaluation of (1) to the evaluation of (4). Specifically, the thought would be that, originally, no non-alcoholic beer-worlds are salient, but the presentation of (4) updates the discourse context by making salient possibilities in which Maddy's drinking an O'Doul's, and, when (4) is evaluated, relative to this updated context, it is false. It's hard to see why one who endorses an account along the lines of that proposed by Lewis wouldn't endorse a strict semantics and just say this. At the very least, one who endorses an account along the lines of that proposed by Lewis owes an explanation of why this approach should be taken for the Maddy Argument and the other cases that concern us here but not for the argument above. That's the first issue. Let me now turn to the second issue, which gets at the root of the issue with pragmatic approaches more generally.

I take it that, when it comes to the above argument from (1) to (4), the reason to endorse a variably strict semantics or a dynamic strict semantics over a fully strict semantics is that doing so enables us to maintain that this intuitively invalid argument above really is invalid. As a general principle, if our semantic theory says that an intuitively invalid argument is valid, this is, all else being equal, a bad result. Now, if this semantic theory is good in enough other places, we may be willing to cope with a bad result in this one place, but, in general, the value of the notion of validity defined by a semantic framework directly corresponds to the extent to which it tracks speakers' judgments of intuitive validity, as manifested by their judgments of the truth or falsity of the elements of a series of sentences, expressed relative to an initial context. The problem is that, on Lewis's account, what is actually valid and what is intuitively valid may entirely swing free of one another. Lewis acknowledges that the context may change as the sentences constituting an argument are presented, with new possibilities becoming salient as these sentences are presented, and the truth-values of sentences vary depending on which possibilities have been made salient, and so these truth-values can change as the context evolves. However, the notion of "validity" that is defined in the truth-conditional framework to which Lewis adheres does not take into account contextual evolution, but, rather, is defined in terms of contexts that are supposed to stay fixed. Since, as Lewis acknowledges, contexts rarely do stay fixed, the truth-conditional notion of validity to which Lewis adheres can systematically fail, by Lewis's own lights, to track intuitive validity. This is what we observe in the cases that concern us, but there is no reason to think that failures of this sort are not completely ubiquitous across natural language. If that's so, it's hard to see what good the truth-conditional notion of validity is for the purposes of natural language semantics at all. It seems that our purposes would be much better served by a notion of validity that incorporates the context change potentials of sentences, and so actually tracks judgments of intuitive validity and invalidity which often depend on contextual evolution. That, of course, is just what our dynamic notion of validity does. So, contra Lewis, there is a decisive reason to incorporate the evolution of context into the semantics, rather than treating it as a pragmatic matter.

A third sort of pragmatic response might be worth considering. I have just been working on the assumption that one who endorses the variably strict truth-conditional semantics would have to maintain that the Maddy Argument is valid, and so maintain that either (1) or (2) is false, relative to the original context, or that (3) is true. However, one line of response that has been brought to my attention is to maintain that, in fact, (1) and (2) do not entail (3) because "beer" is ambiguous between (1) and (3).[24] In (1), "beer" means specifically *alcoholic* beer, and so it is trivially true that if Maddy's drinking a beer, she's drinking an alcoholic beverage. However, in (3), once the possibility of non-alcoholic beer has been introduced through pragmatic factors, the speaker's interpretation of "beer" shifts so that the extension of "beer" includes both *alcoholic or non-alcoholic* beer. As such, inferring (3) from (1) and (2) would be committing a fallacy of equivocation. Now, if one goes this route, then one would surely want to take it generally, and so one would likewise maintain that "fish" in the Frank argument is ambiguous in (1f) and (3f). But this seems to me to be a wildly implausible thing to maintain. On this account, the default semantic value of terms such as "fish" is not a function that maps each world to the set of *fish* in that world, but a function that maps each world to the set of *normal fish* in that world. But if that were the case, the following dialogue should be felicitous:

> Norm: Do you have a pet fish?
> Maddy: No, I have a pet mudskipper.

If "fish," as Norm used it, meant specifically fish that didn't walk on land, then Maddy would be right to respond negatively to his question if she has a pet mudskipper. But clearly, she should respond positively. She *does* have a pet fish—a mudskipper. Mudskippers are fish. Someone who goes the third route is thus forced to denying such trivialities as the claim that, given the default interpretation of "fish" by English speakers, the sentence "Mudskippers are

[24] This was suggested to me by XXXX.

fish" is true, and such a position surely cannot be reasonably maintained. So, this third sort of pragmatic response is out.

These three sorts of pragmatic responses seem to exhaust the space of possible responses by the proponent of a truth-conditional variably strict semantics for indicative conditionals, and they all have serious problems. Of course, more could be said in defense of any of these pragmatic responses, but I conclude that the best thing to say, in response to the data presented here, is that Cumulative Transitivity really is invalid. So, the truth-conditional variably strict semantics for indicative conditionals semantics must be abandoned for a semantics that invalidates Cumulative Transitivity. I have presented one such semantics.

6 The Problem of Closure

Let me close by considering a potential philosophical upshot of the semantic conclusions reached here. There is a widespread belief that, even though we do not in fact believe all of the implications of what we believe, we can at least imagine an ideal rational agent who does so, and, indeed, believing all of the implications of what one believes would be one of the things that an ideal rational agent would do. That is, being ideally rational would involve having beliefs that conform to the following constraint:

> **Closure of Belief Under Implication:** An agent α's beliefs are closed under implication just in case, for any set of propositions Γ and proposition φ, if α believes Γ, and Γ implies φ, then α believes φ.

Once again, of course, as finite rational agents, our own beliefs are not closed under implication, since that would require believing, for instance, all of the tautologies of classical logic, and our limited cognitive capacities preclude us from being able to do that. However, in thinking about how we ought to believe, it can be helpful to imagine a hypothetical ideal reasoner, with unlimited cognitive capacities, who is capable of believing all of the implications of their beliefs and indeed does so. As Horty (2012) clarifies the thought, "Ideal reasoners do not exist, of course, but the myth is nevertheless useful as a competence model for actual reasoners," (15).

Now, when most philosophers consider the way in which an idealized rational agent's beliefs may be closed under implication, they have in mind specifically the set of implications that are called "entailments" in a standard truth-conditional semantics, which obtain between a set of sentences Γ and a sentence φ just in case it is strictly impossible for all the sentences in Γ to be true and φ to be false. This set of implications includes logical implications like the one from "Maddy's wearing a red hat" to "Maddy's wearing a hat," but it also includes certain non-logical semantic implications like the one from "Maddy's wearing a crimson hat" to "Maddy's wearing a red hat," which hold not in virtue of the logical forms of these sentences but in virtue of their material contents, in this case, the contents expressed by the words "crimson" and "red." In considering the potential closure of a rational agent's belief under implication, philosophers generally exclude from consideration defeasible semantic implications like the one from "Maddy's drinking a beer" to "Maddy's drinking an alcoholic beverage." For instance, formal models of belief in semantics and epistemology, drawing on the work of Hintakka (1962), consider beliefs as closed under truth-conditional entailments, but have nothing to say about how believing a proposition also involves believing its defeasible semantic implications. This seems to me to be a rather artificial restriction, and, over the past several decades, proponents of non-monotonic logics have developed formal models according to which we can extend the notion of closure of belief to defeasible implications as well. While I applaud these efforts to formally understand defeasible implication relations, the data here raises a serious problem for the majority of approaches to non-monotonic logic, which build in Cumulative Transitivity in their attempt to articulate the structure of defeasible implication relations.

Let me illustrate the problem caused by the data presented here by showing how it arises in one specific theory that is supposed to enable us to make sense of the idea of an idealized rational agent whose beliefs are not just closed under strict entailment but also under the sort of defeasible implication of concern to us here: Horty's (2007a, 2007b, 2012) default logic. On Horty's model, an agent will be able to take their set of beliefs $\mathcal{W}$ and arrive at an extended set of beliefs, $\mathcal{E}$, which, intuitively, will be the set $\mathcal{W}$ closed under defeasible implication, just in case they have a default theory.[25] Officially, a default theory is a triple $\langle \mathcal{W}, \mathcal{D}, < \rangle$, where $\mathcal{W}$ is a set of propositions codifying the information that is given to an agent, $\mathcal{D}$ is a set of default rules enabling that agent to defeasibly reason from that information, and $<$ is a strict partial ordering on $\mathcal{D}$ codifying priority relations between defaults, the role of which is to be explicated below. In order to determine how our agent is to reason from $\mathcal{W}$, we need to determine the specific set of defaults $\mathcal{S} \subseteq \mathcal{D}$ whose conclusions our agent is to accept. If as our agent's default theory is consistent, there will be only one such set, and this will be the set of defaults $\mathcal{S} \subseteq \mathcal{D}$ that is *stable*, such that all of and only the

[25] The terminology, though in line with how the term "closure" is standardly used in philosophical discussions (see Kvang (2006)) can be a bit misleading here, since this is not, strictly speaking, a closure in the topological sense, as, though we have the properties of extensivity and indempotency, we don't have monotonicity. That is, where $\mathcal{E} = Cl(\mathcal{W})$ we have $\mathcal{W} \subseteq Cl(\mathcal{W})$ and $Cl(Cl(\mathcal{W})) = Cl(\mathcal{W})$, but not it's not the case that, if $\mathcal{W} \subseteq \mathcal{W}'$, then $Cl(\mathcal{W}) \subseteq Cl(\mathcal{W}')$, as illustrated by the example below. Given the way that the term "closure" is used in philosophical contexts, the crucial properties are extensivity and indempotency.

defaults in $\mathcal{S}$ are *binding*, relative to $\mathcal{S}$.[26] Defining the notion of a binding default is then the key task of Horty's theory.

According to Horty, the set of defaults that are *binding*, relative to a set of defaults $\mathcal{S}$, is the set of defaults that are *triggered* by $\mathcal{S}$ and neither *conflicted* nor *defeated* by $\mathcal{S}$. Let me define these three notions in turn. A default $\delta \in \mathcal{D}$ is triggered by a set of defaults $\mathcal{S}$ just in case $\mathcal{W}$ along with the set of conclusions of $\mathcal{S}$ entails the premise of δ. That is:

$$\mathit{triggered}(\mathcal{S}) = \{\delta \in \mathcal{D} : \mathcal{W} \cup \mathit{conclusion}(\mathcal{S}) \vdash \mathit{premise}(\delta)\}$$

A default $\delta \in \mathcal{D}$ is conflicted by a set of defaults $\mathcal{S}$ just in case $\mathcal{W}$ along with the set of conclusions of $\mathcal{S}$ entails the negation of the conclusion of δ. That is:

$$\mathit{conflicted}(\mathcal{S}) = \{\delta \in \mathcal{D} : \mathcal{W} \cup \mathit{conclusion}(\mathcal{S}) \vdash \neg\mathit{conclusion}(\delta)\}$$

A default δ is defeated by a set of defaults $\mathcal{S}$ just in case $\mathcal{S}$ triggers a default δ' that has higher priority than δ and a conclusion that entails the negation of the conclusion of δ. That is:

$$\begin{aligned}&\mathit{defeated}(\mathcal{S}) = \{\delta \in \mathcal{D} : \text{ such that there is a } \delta' \in \mathit{triggered}(\mathcal{S}) \text{ such that}\\ &\qquad (1)\ \delta' > \delta \text{ and}\\ &\qquad (2)\ \mathcal{W} \cap \{\mathit{conclusion}(\delta')\} \vdash \neg\mathit{conclusion}(\delta)\}\end{aligned}$$

So, once again, the set of defaults that are *binding*, relative to a set of defaults $\mathcal{S}$, is the set of defaults that are triggered by $\mathcal{S}$ and neither conflicted nor defeated by $\mathcal{S}$. That is:

$$\mathit{binding}(\mathcal{S}) = \{\delta \in \mathcal{D} : \text{ such that}$$

[26] Now, it is possible, in cases where there are conflicting default rules, that there is more than one stable set of defaults, but, since such cases will not arise here, we will suppose that there will always be a unique such set.

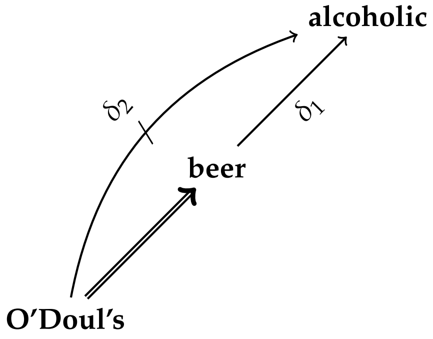
(1) $\delta \in \mathit{triggered}(\mathcal{S})$
(2) $\delta \notin \mathit{conflicted}(\mathcal{S})$
(3) $\delta \notin \mathit{binding}(\mathcal{S})\}$

A set of defaults $\mathcal{S}$ is *stable* just in case $\mathcal{S} = \mathit{binding}(\mathcal{S})$. Where $\mathcal{S}$ is a stable set of defaults, the extension $\mathcal{E}$ of the set of beliefs $\mathcal{W}$ is straightforwardly determined as follows, where $Th(X)$ indicates the closure of a set of propositions X under entailment:

$$\mathcal{E} = Th(\mathcal{W} \cup \mathit{Conclusion}(\mathcal{S}))$$

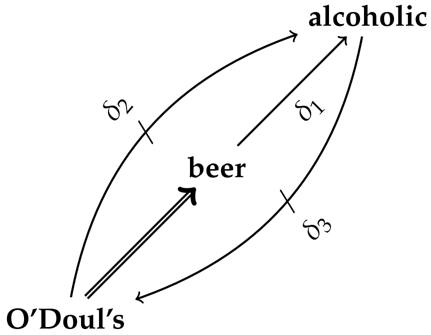
This amounts to a closure of $\mathcal{W}$ under, not only entailment, but defeasible implication as well.

To see how Horty's default logic is supposed to apply in our case, consider the following set of relations, where a double arrow indicates an entailment, a regular arrow indicates a default rule, and a struck-through arrow indicates a default rule to the negation of the proposition that the arrow goes to:



We suppose here that $\delta_2 > \delta_1$ on account of δ_2's being more specific than δ_1. Now, suppose first that $\mathcal{W}$ = {**O'Doul's**}. δ_1 and δ_2 are both triggered in this case. However, since $\delta_2 > \delta_1$, $\mathit{conclusion}(\delta_2) = \neg$(**alcoholic**), $\mathit{conclusion}(\delta_1)$ = **alcoholic**, and {**O'Doul's**} $\cup$ {$\neg$(**alcoholic**)} $\vdash \neg$(**alcoholic**), δ_1 is defeated. δ_2, on the other hand, is not defeated. Thus, the only stable set of defaults is $\{\delta_2\}$. Taking $Th(\mathcal{W} \cup conclusion(\delta_2))$, we get {**O'Doul', beer,** ¬(**alcoholic**)}, along with all the logical consequences of these sentences. Now suppose that $\mathcal{W}$ = {beer}. In this case, only δ_1 is triggered. Since δ_1 is neither conflicted or defeated, the only stable set is $\{\delta_1\}$. Taking $Th(\mathcal{W} \cup conclusion(\delta_1))$, we get {**beer, alcoholic**}, along with all of the logical consequences of this set. This, it seems, is all well and good. However, there is a problem hiding here.

What our examples reveal, is that, in any case of this sort, there is an additional default rule at play. The full diagram looks like the following:



Letting $\mathcal{W}$ = {**beer**} again and considering the possible sets of defaults, we see that, since δ_1 is triggered, relative to $\emptyset$, δ_3 is triggered, relative to $\{\delta_1\}$, and neither δ_1 nor δ_3 is conflicted or defeated relative to $\{\delta_1, \delta_3\}$, the stable set here is $\{\delta_1, \delta_3\}$. Taking $Th(\mathcal{W} \cup\ conclusion(\{\delta_1, \delta_3\}))$ to arrive at the extension of $\mathcal{W}$, we get {**beer, alcoholic**, ¬(**O'Doul's**)}, along with all of the logical consequences of this set. Thus, an idealized rational agent who believes that someone's drinking a beer should, according to this model, extend their beliefs to include the belief that this person's not drinking an O'Doul's. But this is not right. One should not reason from the belief that someone's drinking a beer to the belief that they're not drinking an O'Doul's. Someone's drinking a beer is not, by itself, a reason to think that they're not drinking an O'Doul's.[27] This is so even though someone's drinking a beer is reason to think they're drinking an alcoholic beverage, and someone's drinking a beer along with their drinking an alcoholic beverage is reason to think that they're not drinking an O'Doul's. In general, though something's being a kind *K* is generally a reason to think that it has a characteristic feature *F*, and something's being kind *K* along with it's having characteristic feature *F* is a reason to think it's not an exceptional subkind K', which doesn't have characteristic feature *F*, something's being a *K* is not, by itself, a reason to think that it's not a K'. Reason relations are not universally cumulatively transitive. Accordingly, the idea of closing an idealized rational agent's beliefs under defeasible implication does not make proper sense.

Though this is a problem for Horty's version of non-monotonic logic, it is not just a problem for Horty. Almost all non-monotonic logics yield a consequence relation that is, though non-monotonic, nevertheless cumulatively transitive.[28] Indeed, it is often advertised as a virtue of a non-monotonic logic that, though it lets go of Monotonicity, it holds onto Cumulative Transitivity and various other structural rules that we've come to expect of a logic. The considerations advanced here, however, suggest that, when we have exceptions to Monotonicity involving general rules which have exceptions, we also have exceptions to

[27] According to Horty's *official* theory of reasons, it seems that this is what we should say. On Horty's official definition of reasons, φ is a reason for ψ, in the context of a default theory Δ, just in case there is some triggered default of the form $\varphi \to \psi$ in the stable set $\mathcal{S}$ based on Δ. Since there's no default going directly from from **beer** to ¬(**O'Doul's**), the former is not a reason for the latter, on this official definition. However, though **beer** is not officially a reason for ¬(**O'Doul's**) according to Horty's official definition, it's hard to see how, according to Horty's actual default theory, **beer** could *not* be a reason for ¬(**O'Doul's**), for the theory tells us that we can conclude ¬(**O'Doul's**) from **beer**.

[28] This includes not just Reiter's (1980) version of default logic, but also the preferential models approach of One notable exception is the version of non-monotonic logic proposed by Brandom (2018), developed by Hlobil (2017, 2018), Kaplan (2018), building on Brandom's (1994, 2001, 2008) semantic inferentialism and logical expressivism.

Cumulative Transitivity. These exceptions to Cumulative Transitivity have, for most of the history of non-monotonic logic, been ignored. But, as non-monotonic logic teaches us, we must consider the exceptions.

References

[1] Adams, Ernest. 1965. "The Logic of Conditionals." *Inquiry* 8: 166-197.

[2] Adams, Ernest. 1975. *The Logic of Conditionals*. Dordrecht: Reidel.

[3] Brandom, Robert. 1994. *Making It Explicit*. Cambridge, MA: Harvard University Press.

[4] Brandom, Robert. 2001. *Articulating Reasons*. Cambridge, MA: Harvard University Press.

[5] Brandom, Robert. 2008. *Between Saying and Doing*. Oxford: Oxford University Press.

[6] Brandom, Robert. 2018. "From Logical Expressivism to Expressivist Logic: Sketch of a Program and Some Implementations." *Philosophical Issues* 28, no. 1: 70-88.

[7] Dever, Josh. 2013. "The Revenge of the Semantics-Pragmatics Distinction." *Philosophical Perspectives* 27: 104-144.

[8] Edgington, Dorathy. 1995. "On Conditionals." *Mind*.

[9] Edgington, Dorathy. 2001. "Indicative Conditionals." *The Stanford Encyclopedia of Philosophy*.

[10] von Fintel, Kai. 2001. "Counterfactuals in Dynamic contexts." In *Ken Hale: A Life in Language,* ed. M. J. Kenstowicz, 123-152. Cambridge, MA: MIT University Press.

[11] von Fintel, Kai and Irene Heim. 2011. *Intensional Semantics.* MIT Lecture Notes.

[12] Gillies, Anthony. 2007. "Counterfactual scorekeeping." *Linguistics and Philosophy* 30: 329–360.

[13] Gillies, Anthony. 2009. "On the truth conditions of if (but not quite only if)." *Philosophical Review* 118: 325–349.

[14] Gillies, Anthony. 2004. "Epistemic conditionals and conditional epistemics." *Noûs* 38: 585–616.

[15] Grice, Paul.1975. 'Logic and conversation'. In D. Davidson and G. Harman (eds.), The Logic of Grammar. Dickenson. Encino, CA. 64–75.

[16] Hajek, Alan. 2014. "Most Counterfactuals Are False." Manuscript. https://philpapers.org/rec/HJEMCA.

[17] Hintikka, Jaakko. 1962. *Knowledge and Belief*. Ithaca: Cornell University Press.

[18] Hlobil, Ulf. 2016. "A Nonmonotonic Sequent Calculus for Inferentialist Expressivists." In *The Logica Yearbook 2015*, ed. P. Arazim and T. Lákiva, 87-105. London: College Publications.

[19] Hlobil, Ulf. 2017. "When Structural Principles Hold Merely Locally. In *The Logica Yearbook 2016*, ed. P. Arazim and T. Lákiva, 53-67. London: College Publications.

[20] Horty, John. 2007a. “Reasons as Defaults.” *Philosopher's Imprint* 7, no 3: 1-28.

[21] Horty, John. 2007b. “Defaults with Priorities.” *Journal of Philosophical Logic* 36: 367-413.

[22] Horty, John. 2012. *Reasons as Defaults*. Oxford: Oxford University Press.

[23] Hurford, James. 1974. “Inclusive or Exclusive Disjunction.” *Foundations of Language* 11, no. 3: 409-411.

[24] Kaplan, Daniel. 2018. “A Multi-Succident Sequent Calculus for Logical Expressivists.” In *The Logica Yearbook 2017*, ed. P. Arazim and T. Lákiva, 139-154. London: College Publications..

[25] Kaplan, David. 1989. “Demonstratives.” In *Themes From Kaplan*, ed. J. Almog, J. Perry, and H. Wettstein, 481-563. Oxford: Oxford University Press.

[26] Kratzer, Angelika. 1986. “Conditionals.” *Chicago Linguistics Society* 22, no. 2:1-15.

[27] Kraus, S., Lehmann, D., and Magidor, M. (1990). “Nonmonotonic Reasoning, Preferential Models and Cumulative Logics.” *Artificial Intelligence* 44, no. 1-2: 167-207.

[28] Kvang, Jonathan. 2006. “Closure Principles.” *Philosophy Compass* 1/3: 256-267.

[29] Lewis, David. 1973. *Counterfactuals.* Cambridge, MA: Harvard University Press.

[30] Lewis, Karen. 2014. "Do we need dynamic semantics?" In *Metasemantics: New Essays on the Foundations of Meaning*, ed. A. Burgess and B. Sherman, 231-258. Oxford: Oxford University Press.

[31] Lewis, Karen. 2018. "Counterfactual Discourse in Context." *Nous* 52, no. 3: 481-507

[32] Makinson, David. 2005. *Bridges from Classical to Non-Monotonic Logic*. London: King's College Publications.

[33] McGee, Vann. 1985. "A Counterexample to Modus Ponens." *The Journal of Transcendental Philosophy* 82, no. 9: 462-471.

[34] Moss, Sarah. 2012. "On the Pragmatics of Counterfactuals." *Noûs* 46, no. 3: 561–86.

[35] Rothschild, Daniel. 2020 "A Note on Conditionals and Restrictors." In *Conditionals, Probability, and Paradox: themes from the philosophy of Dorothy Edgington*, ed. L. Waters and J. Hawthorne. Oxford: Oxford University Press.

[36] Sellars, Wilfrid. 1953. "Inference and Meaning."

[37] Sellars, Wilfrid. 1954. "Some Reflections on Language Games."

[38] Sellars, Wilfrid. 1974. "Meaning as Functional Classifcation."

[39] Stalnaker, Robert. 1968. "A Theory of Conditionals. In *Studies in Logical Theory,* ed. N. Rescher, 98-112. Oxford: Blackwell.

[40] Stalnaker, Robert. 1975. "Indicative Conditionals." *Philosophia* 5, no. 3: 269-286.

[41] Stalnaker, Robert. 1984. *Inquiry*. Cambridge, MA: MIT Press.

[42] Starr, William. 2014. “What ‘If’?” *Philosophers’ Imprint* 14, no. 10: 1-27.

[43] Veltman, Frank. 1996. “Defaults in Update Semantics.” *Journal of Philosophical Logic*, 25, 221-261.

[44] Willer, Malte. 2013. “Indicative Scorekeeping.” In *Amsterdam Colloquium*, 249-256.

[45] Willer, Malte. 2017. “Lessons from Sobel Sequences.” *Semantics and Pragmatics* 10, no. 4: 1-57.

[46] Yalcin, Seth. 2014. “Semantics and Metasemantics in the Context of Generative Grammar.” In *Metasemantics: New Essays on the Foundations of Meaning*, edited by Alexis Burgess and Brett Sherman, 17-54. Oxford: Oxford University Press.