How to (Not) Derive Negative Introspection

Many think there are special connections between our attitudes towards conditionals and our attitudes towards non-conditional facts. Take a claim like

(1) If it rains, there will be no picnic.

Some think that, following Ramsey (1925), believing (1) amounts to believing that there will be no picnic, conditional on learning that it will rain. Similarly, Stalnaker (1970) and others have suggested that one's probability in claims like (1) should line up with one's conditional probabilities. Call these *linking* principles.

I will show that, on standard approaches to conditionals, strong *epistemological* consequences result from combining these linking principles with other generally accepted principles about conditionals. They entail that our information obeys the principle of *Negative Introspection*, which says that if our information leaves open ϕ, then it is part of our information that our information leaves open ϕ. This principle is subject to some controversy: it is generally rejected by *externalist* approaches to epistemology, on which access to our information can be highly limited.

This seems to leave us in the uncomfortable position of choosing between our conditional epistemology and epistemology at large. To resolve this tension, I take a different route. Most agree that conditionals make assertions about some relevant body of information in the context; call this property *information-sensitivity*. Traditional approaches spell this out in a particular way: indicatives are sensitive only to the global information in the context and not local linguistic information. This means that, so long as they are assessed in the same global context, the contribution of (1) is the same unembedded as it is in embedded contexts, like:

(2) John thinks that if it rains there will be no picnic.

More recently, theorists working in both static and dynamic traditions have challenged this, claiming that the contribution of the conditional *can* depend on its local informational environment. I show that this thesis, which following [redacted] I'll call *Conditional Locality*, offers us a way out of the dilemma, reconciling our linking principles with externalism.

1 Information-sensitivity of conditionals

My statement of the linking principles will take for granted that conditionals are information-sensitive. So let us first rehearse at least one argument for thinking that they are indeed information-sensitive.[1]

Or-to-if reasoning seems very reliable. If Alice knows:

(3) Matt is either in London or Los Angeles.

then Alice can infer:

(4) If Matt is not in London, he's in Los Angeles.

In general, whenever someone knows some disjunction $\phi \vee \psi$ (and they leave open which), then they also (are in a position to) know the indicative *if* $\neg\phi$*, then* ψ.

But now suppose that Billy knows:

(5) Matt is either in London or Oxford.

and so concludes:

(6) If Matt is not in London, he's in Oxford.

The conclusions that Alice and Billy have reached seem incompatible. In general, conditionals with shared antecedents and contradictory conclusions themselves sound contradictory.[2] No one can happily say:

[1] See, e.g., Stalnaker (1984) ch.6 for an argument along these lines.

[2] This is the principle of Conditional Non-Contradiction. Stalnaker (1968) and Lewis (1973) validate restricted versions of this principle.

(7) #If Matt is not in London, he's in Los Angeles and if Matt is not in London, he's in Oxford.

But Alice's and Billy's starting points are compatible: Matt might be in London. Supposing he is, Alice and Billy have used reliable reasoning to work from known premises to incompatible conclusions. How can this be?

Enter information-sensitivity. Their conclusions are compatible after all, we might say, because (4) and (6) were uttered in different contexts by people with different information. Certainly no one can say both (4) and (6) in one and the same context. But the truth of (4) *as said by Alice* is compatible with the truth of (6) *as said by Billy*. Information-sensitivity preserves the reliability of Or-to-if reasoning, while maintaining that the conjunction of (4) and (6) is incoherent: different contexts, different information, different propositions.

From here, I take it as read that conditionals are information-sensitive. I will assume specifically a *contextualist* account of information-sensitivity: the relevant information in the context of utterance determines the proposition expressed by a conditional. I will also assume that the relevant information is what the speakers in the context (distributively) *fully believe*, given their evidence; to avoid confusion with the notion of weak belief from Hawthorne et al. (2016), Dorst (2019b) and Rothschild (forthcoming), I'll also say this is what the speakers are *sure of*. However, relativist and expressivists will be equally susceptible to the arguments that follow; and we will later see how to generalise my arguments, if the relevant information is stated in terms of other attitudes like knowledge or weak belief.

2 The Key Premises

Rather than working directly with either the Ramsey Test or Stalnaker's Thesis, my argument will be more indirect. I will focus on a weaker principle, Stability, that is plausibly entailed by both linking principles. My argument for Negative Introspection will rely on this principle, together with a principle about the presuppositions of indicatives.

Stability, the first main premise of my argument, says that when the speakers are sure that $[\![\psi]\!]$ and are not sure whether $[\![\phi]\!]$, then they are sure of the proposition that ⌜If ϕ then ψ⌝ expresses in their context. To state the principle precisely, let's add to our language an operator S_c which tracks the relevant body of information in a context c: ⌜$S_c\phi$⌝ is true at $\langle c, w\rangle$ just in case the relevant agents in c distributively should be sure of ϕ at w. Writing > for the indicative, this gives us:

Stability. $\neg S_c \neg\phi, S_c\psi \models S_c(\phi > \psi)$

A version of this natural principle was first formulated in Dorst (2019a).[3]

This principle is a downstream consequence of the linking principles from the introduction. For instance, given standard assumptions about belief revision, it follows from a localised version of the Ramsey Test:[4]

Local Ramsey Test. $S_c(\phi > \psi) \equiv S_c^{\phi}\psi$.

This thesis says that speakers are sure of the indicative expressed by $\phi > \psi$ in their context just in case they are sure of ψ *conditional* on ϕ.

Plausibly, it is also a consequence of a localised version of Stalnaker's Thesis:

Local Stalnaker's Thesis. $Pr_{c,w}([\![\phi > \psi]\!]^c) = Pr_{c,w}([\![\psi]\!]^c | [\![\phi]\!]^c)$, when $Pr_{c,w}([\![\phi]\!]^c) > 0$.

This thesis says speakers have probability 1 in what ⌜$\phi > \psi$⌝ expresses in their context iff they have probability 1 in the consequent conditional on the antecedent.

[3] Note that this principle is crucially different from the following principle, which applies not just to the speaker, but any agent more generally:

Global Stability. $\neg S \neg\phi, S\psi \models S(\phi > \psi)$

I focus on localised versions of Stability, as well as of the Ramsey Test and Stalnaker's Thesis, as these versions avoid triviality results for the principles. See §6.3 for further discussion of triviality for Stability. van Fraassen (1976), Bacon (2015) and Khoo (2019), among others, localise Stalnaker's Thesis to avoid the triviality results of Lewis (1976). A similar strategy defuses the triviality results of Gärdenfors (1986) for the Ramsey Test.

[4] The relevant assumptions here are:

Success. $\models S_c^{\phi}\phi$

Preservation. $\neg S_c \neg\phi, S_c\psi \models S_c^{\phi}\psi$

If we assume that one is sure of a proposition just in case one assigns it probability 1, then Stability follows from Local Stalnaker's Thesis.

But we can also motivate Stability directly by appeal to examples. In each of these discourses, the (b)-sentences are incoherent:

(8) a. I'm sure that there won't be a picnic and I'm not sure whether it will rain.
b. #I'm also not sure whether if it rains, there will be a picnic.

(9) a. John is sure that Bob is in his office but he isn't sure whether Vann is in his office.
b. #He also isn't sure whether if Vann is in his office, Bob is in his office.

Stability explains why the (b)-sentences sound off: if the (a)-sentence is true, the (b)-sentence follows.

The second main premise of my argument is *Antecedent Presupposition*, the principle that a conditional $\ulcorner \phi > \psi \urcorner$ presupposes that ϕ is consistent with the relevant information in the context:[5]

Antecedent Presupposition. $[\![\phi > \psi]\!]^c$ presupposes $[\![\neg S_c \neg \phi]\!]^c$

Why think it holds? Because it is generally infelicitous to assert both a conditional and the denial of its antecedent:

(10) # If it rains, there won't be a picnic but (I'm sure) it isn't going to rain.

(11) #If John was at the party, I didn't see him but (I'm sure) he wasn't at the party.

Absent special pragmatic pressure, one does not say things of the form $\ulcorner \phi > \psi$ and (I'm sure that) $\neg\phi \urcorner$. This is well explained by Antecedent Presupposition.

Standardly, presuppositions are thought of as definedness conditions: ϕ presupposes ψ just in case when ϕ is either true or false, ψ is true.[6] On standard

[5] This has been defended by, among others, von Fintel (1998), Gillies (2009) and Leahy (2011).

[6] There is another, less standard *multi-dimensional* approach to presupposition, given by Herzberger (1973) and Karttunen and Peters (1979). Here sentences are assigned two values, a truth-value and a separate value to reflect whether its presuppositions are satisfied. On this approach to

assumptions, this means presuppositions are entailments: when ϕ presupposes ψ, ϕ entails ψ. Together with Antecedent Presupposition, this yields the principle I will call *If-to-might*:

If-to-might. $\models (\phi > \psi) \supset \neg S_c \neg \phi$

If-to-might is the principle I will use in the argument below; but remember it is Antecedent Presupposition that motivates it.

3 The central argument

Here is what we will show. We will pick an arbitrary context c and show that, given our previous indexical assumptions, Negative Introspection governs what the agents in c should be distributively sure of. More precisely, we will prove:

Negative Introspection. $\neg S_c \phi \models S_c \neg S_c \phi$

We've just seen the two main premises of the argument. I will also make two standard assumptions about the logic of full belief, that it is closed under fully believed entailment and that logical truths are known:

Closure. $S_c(\phi \supset \psi), S_c\phi \models S_c\psi$

Omniscience. If $\models \phi$, then $\models S_c\phi$

These four principles allow us to easily derive Negative Introspection. Let $\top$ be some arbitrary tautology. Then:

1. $\neg S_c \phi$ (assumption for conditional proof)
2. $S_c \top$ (Omniscience)
3. $S_c(\neg \phi > \top)$ (Stability)
4. $S_c((\neg \phi > \top) \supset \neg S_c \phi)$ (Omniscience, If-to-might, double negation)

presupposition my result is slightly more nuanced: whenever $\ulcorner \neg S\phi \urcorner$ is true *and* has its presuppositions satisfied, then so does $\ulcorner S \neg S\phi \urcorner$. This does not seem to me to be much of an improvement over the result in the main text.

5. $S_c \neg S_c \phi$ (Closure, 3, 4)

C. $\neg S_c \phi \supset S_c \neg S_c \phi$ (1,5, conditional proof)

In brief, if you're not sure that ϕ, then Stability entails that, for some tautology $\top$, you are sure of $\phi > \top$. From Antecedent Presupposition and Closure, it then follows you are sure the presupposition of the conditional must be satisfied — you are sure that you aren't sure that ϕ.

We will now explore what can be said to this argument, given the normal understanding of information-sensitivity. I think there are five basic moves here:

1. accept Negative Introspection;
2. reject some of the background logic;
3. reject Antecedent Presupposition wholesale;
4. reformulate Antecedent Presupposition;
5. reject Stability.

I'll argue that none of these are satisfactory.

4 Accept Negative Introspection?

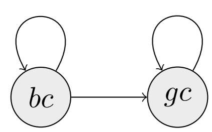
Why not simply accept Negative Introspection? Because it is generally taken to be incompatible with externalism, a dominant approach in contemporary epistemology.

Externalists are motivated here by thinking about *good* cases and *bad* cases. In a good case, things are exactly as they appear. In a bad case, our information-gathering mechanisms fail: though things seem a certain way, they are not in fact that way. For instance, in a good case we may be looking at a red wall; in the corresponding bad case, we might be looking at a white wall that, unbeknownst to us, is lit with red lighting.[7]

[7]The basic idea here comes from Williamson (2000) but the example is Salow (2019)'s. See also Lasonen-Aarnio (2015) for other interesting examples.

What is part of your evidence in each case? On pain of skepticism, we must say that it *is* part of your evidence in the good case that the wall is red. The bad case is a different matter. It's widely accepted by externalists that only truths can be part of your evidence.[8] This means that, in the bad case, it cannot be part of your evidence that the wall is red. But in the bad case, you have no reason to think things have gone awry: it's compatible with your evidence that *the wall is red* is part of your evidence.

Now we can see why, on the externalist approach, such cases are a failure of Negative Introspection. In the bad case, it is not part of your evidence that the wall is red, so you should not be sure of it. But it is compatible with your evidence that you are in the good case and so that it is part of your evidence that the wall is red. In diagram form:



As the reader can verify, Negative Introspection fails in such a configuration.

This is not a knock-down argument against Negative Introspection of course. The principle has some defenders even in the face of arguments like the above, especially in the internalist tradition.[9] Nonetheless, having to accept Negative Introspection here seems to me to be a definite cost of the linking principles: in a conflict between the two, it's not clear which should be the first to go.

One might point out that the conclusion my argument is slightly more limited than full Negative Introspection: it says that Negative Introspection holds in any arbitrary *context*. But externalists will not think this limited principle is any more plausible. Simply modify the counterexample above so that I am the only relevant agent in the context and we have a counterexample to the more restricted version of Negative Introspection. All in all, the conclusion of my argument should be worrying: rejecting Negative Introspection is a core tenet of a dominant approach to epistemology.

[8] See for example Williamson (2000).

[9] See for example Smithies (2016).

5 Reject some of the background logic?

My background principles are subject to some controversy. Most do not accept that we should be sure of *all* logical truths. And some reject closure principles. But I think they are not the ultimate source of the problem. Strong versions of these assumptions bring out the issues nicely; but the problem remains even when we weaken them.

Omniscience in full generality is not needed to push through the argument. We used it at just two places: step 2, where it really amounts to the assumption that the agent is sure of *some* tautology, and step 4, where it amounts to the assumption that you are sure that If-to-might holds. Presumably we should only target the second assumption — perhaps we should not always be sure of the If-to-might principle, even if it is always true.

But this does not address the scale of the problem. It's just not plausible that all rational agents are *necessarily* ignorant of this logical truth. We can imagine logically omniscient agents, ones who know *all* logical truths. And we can imagine logically omniscient agents who find themselves in *bad* cases. The fact that they are logically omniscient should not affect their standing with respect to Negative Introspection: if the arguments against Negative Introspection are good, they should be good for omniscient agents too. I submit that this is not a promising response.

Those who reject Closure, like Dretske (1970) and Nozick (1981), typically reject it for *knowledge* rather than rational sureness. But even putting that aside, my use of Closure is quite minimal. By step 4, we have deduced that I am sure of the material conditional ($\neg\phi > \top \supset S\neg\phi$) and I am sure that $\phi > \top$. Closure allows us to conclude that I am sure that I am not sure that ϕ. Even those who deny full-blown Closure tend to agree some restricted version must be plausible: we must somehow explain why deduction tends to extend our knowledge or justified full beliefs. I know of no plausible restriction that justifies thinking Closure fails at this crucial step of the argument.

6 Reject Antecedent Presupposition Wholesale?

Antecedent Presupposition is probably the most tempting place to push back on the argument. In this section, I'll first explore a possible motivation for rejecting the principle altogether.

One might simply reject Antecedent Presupposition entirely. Because while in general one cannot assert a conditional and rule out its antecedent, this generality seems to have exceptions. Consider the following from Dorst (2019a):

(12) I know that Oswald shot Kennedy. But I also know that if he didn't, then someone else did.

(13) I know that plenty of people exist. But I also know that if no one exists, then I don't exist.

If the above are fine, can we really say conditionals presuppose the openness of their antecedents?

The data here are diverse: Dorst's examples are felicitous, but many others are not. We cannot yet conclude Antecedent Presupposition is thus refuted. We need to ask whether the best overall account of *all* the data assumes something like Antecedent Presupposition.

A good story can be told about these data using Antecedent Presupposition. We are granting that conditionals are interpreted relative to the information we take for granted in context. But what we take for granted is itself a context-sensitive notion and can vary with our standards. Suppose I am about to leave the house for work. I've heard on the radio that the traffic is very light and besides I am leaving 15 minutes earlier than I need to. It seems appropriate for me to say:

(14) I will get to work on time today.

But then you point out to me that there is a small chance I will suffer a fatal heart attack on my way. Now I can no longer flat out assert (14). I must content myself with more hedged claims such as:

(15) It's overwhelmingly likely that I will get to work on time today.

This is because one you have raised the possibility of a heart attack, our standards rise. We can no longer take for granted that I won't have a heart attack, however unlikely it may be.

A natural story about Dorst's examples is that the conditionals are themselves raising the standards. Once we utter the conditional:

(16) If Oswald didn't shoot Kennedy, then someone else did.

we raise to salience the possibility that the assassin was someone else. Now we are no longer taking for granted that Oswald *did* shoot Kennedy and so now the presupposition of (16) is satisfied. Further confirmation for my story comes from an order asymmetry in the examples. It sounds worse to my ear to say either of:

(17) I know that if Oswald didn't shoot Kennedy, then someone else did. # But I know that Oswald shot Kennedy.

(18) I know that if no one exists, then I don't exist. # But I know that plenty of people exist.

This is what my story predicts. It's well known that raising the standards is easier than lowering them. So after asserting (16), the standards should remain high, preventing us from continuing the discourse in either of (17) or (18).

Antecedent Presupposition can account for these apparent counterexamples. And I know of no good story that accounts for all the data without it. The best story I can think of says that our original examples are bad because they involve redundancy. But this fails to capture Dorst's original examples.

Recall we noted in §1 that Or-to-if seems like a good rule of inference. In fact, we can go further than this: indicatives often just seem *interchangeable* with disjunctions in reasoning. For example, the claim:

(19) If it doesn't rain, it will snow.

seems *interchangeable* with the claim:

(20) Either it rains or it snows.

This gives us something like the following as a generally reliable rule of inference:

(21) If ϕ, $\psi \leftrightsquigarrow \neg\phi$ or ψ

Return to our problem. Given (21), when someone who asserts

(22) #If it rains, there won't be a picnic but it isn't going to rain.

what they say is equivalent to:

(23) (Either it isn't going to rain or there'll be a picnic) and it isn't going to rain.

This latter might be thought to involve some problematic redundancy, since the right conjunct entails the left.[10]

But this explanation doesn't explain why Dorst's examples are good. Disjunctions and indicatives seem equally interchangeable under *knowledge* ascriptions:

(24) John knows if it doesn't rain, it will snow. $\leftrightsquigarrow$ John knows either it will rain or snow.

So if the redundancy explanation is correct, in Dorst's examples, the claim

(12) I know that Oswald shot Kennedy. But I also know that if he didn't, then someone else did.

should communicate:

(25) I know that Oswald shot Kennedy. But I also know that either Oswald shot Kennedy or someone else shot Kennedy.

Here the left conjunct entails the right; so (12) too should be redundant and marked, not natural as Dorst observed.

[10]See Hurford (1974) and the subsequent literature on Hurford's Constraint for further discussion of this kind of redundancy.

7 Reformulate Antecedent Presupposition?

Even if we grant that something like Antecedent Presupposition is needed, one might think my specific formulation of the principle is at fault; perhaps other, better formulations don't have the problematic consequence. Here I'll explore two different ways to reformulate Antecedent Presupposition, the first by stating it in terms of an attitude other than sureness, the second by referring *rigidly* to the relevant information. Neither, I'll argue, successfully get us out of the argument.

7.1 First attempt

I assumed that the relevant background information in a context is what the agents fully believe or are sure of. But perhaps that is not right. Perhaps the relevant information is stated in terms of some stronger notion like what the relevant agents *know*; or some weaker notion like what they *weakly believe*.

Even so, I think my argument, or something like it, will go through. I divide this strategy into two cases: one might try to state the principle using some notion entailed by rational sureness; or using some notion *not* entailed by rational sureness. I show that in the first case only a minor adjustment is needed to my original argument. And I argue that in the second case we can derive a slightly more restricted version of Negative Introspection for that attitude, one that will still be unwelcome to externalists.

First, consider the more straightforward case when the relevant notion is weaker. Here we need only add a few lines to my original argument to again derive Negative Introspection for sureness. Suppose that we thought that Antecedent Presupposition should be stated in terms of justified weak belief:

JB Antecedent Presupposition. $[\![\phi > \psi]\!]^{c,w}$ is defined only if $[\![\neg B_c \neg\phi]\!]^{c,w} = 1$.

This gives us a new version of our *If-to-might* principle:

If-to-JB-might. $\models (\phi > \psi) \supset \neg B_c \neg \phi$.

Finally, since we are assuming that sureness entails weak belief, we have the following:

S-to-B. $\models S_c\phi \supset B_c\phi$

The above, together with Stability, allow us to once again establish Negative Introspection:[11]

1. $\neg S_c\phi$ (assumption for conditional proof)
2. $S_c\top$ (Omniscience)
3. $S_c(\neg\phi > \top)$ (Stability)
4. $S_c((\neg\phi > \top) \supset \neg B_c\phi)$ (Omniscience, If-to-JB-might, double negation)
5. $S_c\neg B_c\phi$ (3,4, Closure)
6. $S_c(\neg B_c\phi \supset \neg S_c\phi)$ (Omniscience, and S-to-B, contraposed)
7. $S_c\neg S_c\phi$ (6, 7, Closure)

C. $\neg S_c\phi \supset S_c\neg S_c\phi$ (1,6, conditional proof)

Because we know that sureness entails weak belief, a lack of weak belief entails a lack of sureness. This then allows us to complete the proof of Negative Introspection for rational sureness, despite the fact that Antecedent Presupposition was stated in terms of a different, weaker notion. And even though we are no longer identifying the agent's full beliefs with the relevant information in the context, externalists should still want to resist this conclusion.

Now let us take the second case, where Antecedent Presupposition is stated in terms of some attitude *not* weaker than sureness. For instance, what if the relevant body of information is what the agents are in a position to (distributively) know? Let us introduce another operator, K_c, where $\ulcorner K_c\phi \urcorner$ holds at $\langle c, w\rangle$ iff the relevant agents in c are in a position to know $[\![\phi]\!]^c$. Then we would state Antecedent Presupposition as follows:

Knowledge AP. $[\![\phi > \psi]\!]^{c,w}$ is defined only if $[\![\neg K_c\neg\phi]\!]^{c,w} = 1$

[11] Note that, even if the relevant information is not what the agents are sure of, it seems to me that the statement of Stability in terms of full belief remains justified, both by its connections to the linking principles and the natural language data.

This then gives us a different version of our *If-to-might* principle:

If-to-K-might. $\models (\phi > \psi) \supset \neg K_c \neg \phi$

Does this block the argument? It blocks the original argument, yes. But we can still prove a slightly more limited result that will be almost as objectionable to externalists. Following Dorst (2019a), say that a *diligent* agent is one who should be sure of a proposition just in case they are in a position to know it:

Diligence. $S_c\phi \equiv K_c\phi$

For diligent agents, we can show that a version of Stability stated in terms of knowledge holds:[12]

K Stability. $\neg K_c \neg \phi\ K_c\psi \models K_c(\phi > \psi)$

[12] I focus on diligent agents because K-Stability has counterexamples when it comes to non-diligent agents. (These cases are due to [redacted]; I learned of them from [redacted] (p.c.).) Consider the following case:

> **Some bad testimony.** Suppose Alice tells me her birthday is on November 1 and Billy tells me his is on December 1. Both are generally trustworthy, so I believe what they tell me. But only Alice is telling the truth; Billy is lying.

Stability for knowledge fails here. Presumably I know that Alice's birthday is on November 1. It's true and learned from a reliable informant; and the fact that a totally separate person is unreliable should not defeat Alice's testimony. (If it helps, imagine Alice and Billy don't know of each other's existence.) So:

(i) K(Alice's birthday is November 1)

But I can't know Billy's birthday is on December 1, because it isn't. So it's compatible with my knowledge that *one* of them is wrong:

(ii) $\neg K \neg$(Alice's birthday is not November 1 $\vee$ Billy's birthday is not December 1)

Applying Stability gives us that I know that, if either Alice's birthday isn't November 1 or Billy's isn't December 1, then Alice's birthday is still on November 1:

(iii) K(Alice's birthday is not November 1 $\vee$ Billy's birthday is not December 1> Alice's birthday is November 1)

But it is quite implausible that I know this conditional. Knowing this conditional is tantamount to me knowing that, if one of them is lying, then it is Billy. But I don't know this.

I consign the (simple) proof to a footnote.[13] Given K Stability, it is straightforward to adapt my original proof and prove that, in any given context involving a single diligent agent, they must obey the Negative Introspection principle for knowledge:

(26) $\neg K_c\phi \models K_c\neg K_c\phi$

Limiting Negative Introspection to these kinds of agents should not be of comfort to externalists. Diligent agents are ones whose evidential positions line up exactly with what they are in a position to know. There is no reason that this should insulate them from failures of Negative Introspection. Just like anyone else, a diligent agent may find themselves in a bad case. If we are persuaded by the general reasoning that people in bad cases fail to obey Negative Introspection, then we should think that Negative Introspection fails for diligent agents too. (In fact, given that everyone concedes knowledge is factive, there is far less room to manoeuvre here.)

So while the original argument targeted sureness specifically, we can now see a way to extend it to any attitude A not entailed rational sureness. Suppose we state Antecedent Presupposition in terms of A. Then say that an agent is diligent with respect to A just in case they A that ϕ just in case they should be sure of it. For any such attitude, diligence will allow us to derive the corresponding version of Stability; and so we will be able to construct a proof that A obeys Negative Introspection for agents that are diligent with respect to A. I conjecture that, for any plausible candidate attitude for stating Antecedent Presupposition, externalists should think Negative Introspection does not hold for that attitude.

[13] Again, we use conditional proof, assuming the premises of K Stability:

1. $\neg K_c\neg\phi$ (Assumption)
2. $K_c\psi$ (Assumption)
3. $\neg S_c\neg\phi$ (From 1 and Diligence)
4. $S_c\psi$ (From 2 and Diligence)
5. $S_c(\phi > \psi)$ (from 3,4 and Stability)
6. $K_c(\phi > \psi)$ (from 5 and Diligence)

7.2 Second Attempt

Rather than object to the attitude, we might object to the *way* my formulation of Antecedent Presupposition refers to the relevant body of information. My statement of the principle refers to the relevant body of information *non-rigidly*: as uttered in c, $\ulcorner\phi > \psi\urcorner$ presupposes at a given world that the relevant information at that world leaves open ϕ. But perhaps instead the presupposition rigidly refers to the body of information, so that, as uttered in c, $\ulcorner\phi > \psi\urcorner$ presupposes at that the relevant information in c leaves open ϕ. Put more precisely, this reads:

Rigidified Antecedent Presupposition. $[\![\phi > \psi]\!]^{c,w}$ is defined only if $S_c(w_c) \cap [\![\phi]\!]^c \neq \varnothing$.

On this view, *If-to-might* is merely *diagonally* valid. Say that ϕ diagonally entails ψ ($\phi \models_D \psi$) just in case if $[\![\phi]\!]^{c,w_c} = 1$ then $[\![\psi]\!]^{c,w_c} = 1$; that is, if whenever ϕ is true at a *context*, then so is ψ. Given Rigidified Antecedent Presupposition, we merely have:

Diagonal If-to-might. $\models_D (\phi > \psi) \supset \neg S_c \neg\phi$

This version of Antecedent Presupposition would indeed block our proof, because Omniscience does not hold for merely *diagonally* valid sentences. Modifying Kaplan (1989)'s original example, the fact that

(27) I am here.

is diagonally valid does not allow us to infer

(28) John knows I am here.

Indeed we still cannot infer that, even if we suppose John is logically omniscient. So Rigidified Antecedent Presupposition blocks the argument at the fourth step, where Omniscience is applied to If-to-might.

In important respects, this comes close to my solution based on Conditional Locality. Nonetheless, I think we can dispense with this alternative statement of Antecedent Presupposition pretty quickly. The reason is that we need it to apply in embedded environments. For instance, consider the following:

(29) # John is imagining that I am sure that it's not raining and am sure that if it is raining, I will get wet.

This sentence sounds marked, and it seems that the reason is closely connected to the reasons why sentences like (10), repeated below, are marked:

(10) #If it rains, there won't be a picnic but it isn't going to rain.

Just as by saying (10), I am felt to commit myself to a contradiction, in (37), John is heard to attribute to me a contradiction. However we state it, we should want our version of Antecedent Presupposition to explain why the pattern in (10) persists under embeddings.

Rigidified Antecedent Presupposition does not do this. Return to (37) above and simply imagine that John knows that as a matter of fact that I am not sure whether it is raining. Then, according to Rigidified Antecedent Presupposition, the presupposition of the embedded conditional in (37) is met: even though the conditional is being evaluated at non-actual worlds, all it takes for its presuppositions to be met there is for me to leave open its antecedent in the *actual* world. On the other hand, my original statement of Antecedent Presupposition *does* explain why John is felt to attribute to me a contradiction; but at the cost of imposing Negative Introspection.

8 Reject Stability?

Our final option is to deny Stability. This is not the easiest route out of the argument. Considered on a case by case basis, it looks extremely plausible. But one might worry about its theoretical motivations: (versions of) both of our linking principles have after all been subjected to triviality results. Inspired by this, one might press a similar argument against Stability — perhaps it too should be rejected because it trivialises.

Here is how the objection might go. Stability looks quite similar to another principle, Preservation, due to Bradley (2000) and which says that if you have probability 0 in ψ and non-zero probability in ϕ, then you have probability 0 in $\phi > \psi$:

Preservation. For any probability function Pr, if $Pr_w(\llbracket\psi\rrbracket^c) = 0$ and $Pr_w(\llbracket\phi\rrbracket^c) > 0$ then $Pr_w(\llbracket\phi > \psi\rrbracket^c) = 0$

This similarity should be worrying. Bradley showed that, on any Boolean semantics, Preservation trivialises the conditional in a particular sense: for any non-conditional ϕ and ψ either ϕ entails ψ or $\phi > \psi$ entails ψ, if Preservation holds fully.

For suppose that neither $\phi \models \psi$ nor $\phi > \psi \models \psi$. Then there is some probability measure where $Pr_w(\llbracket\phi\rrbracket^c) > 0$, $Pr_w(\llbracket\phi > \psi\rrbracket^c) > 0$ but $Pr_w(\llbracket\psi\rrbracket^c) = 0$; and this is a counterexample to Preservation. So, given Preservation, either ϕ entails ψ or $\phi > \psi$ entails ψ. This would clearly be absurd: neither of

(30) It will rain.

(31) If it rains, there will be no picnic.

logically entail

(32) There will be no picnic.

For both of (30) and (31), one can imagine a situation in which it holds and but (32) doesn't.

More worrying still, the similarity is more than superficial. Stability entails (a version of) Preservation, assuming some plausible principles about the conditional and about the relation between sureness and probability. First assume that sureness is probability 1:

(33) $\llbracket S(\phi)\rrbracket^{c,w} = 1$ iff $Pr_w(\llbracket\phi\rrbracket^c) = 1$.

And assume Conditional Non-Contradiction, the principle that, when you leave open its antecedent, the contrary of an indicative entails its negation:

CNC. $\neg S\neg\phi, \phi > \neg\psi \models \neg(\phi > \psi)$

These together allow us to derive, where $Pr_c(\phi)$ is the rational probability of ϕ in c:

Local Preservation. For any probability function Pr, if $Pr_{c_w}(\llbracket\psi\rrbracket^c) = 0$ and $Pr_{c_w}(\llbracket\phi\rrbracket^c) > 0$ then $Pr_{c_w}(\llbracket\phi > \psi\rrbracket^c) = 0$

That is, whenever in a given context you are sure that ψ is false but leave open ϕ, you should be sure of the conditional that $\phi > \psi$ expresses in *your* context.

However, I think triviality is not ultimately a threat to my argument for two quite different reasons. The first is that full blown Stability is in fact far stronger than what we actually need for the proof. The only instance of Stability we needed was the following:

Tautological Stability. $\neg S_c \neg\phi, S\top \models S_c(\phi > \top)$

Whatever about full blown Stability, this principle is extremely plausible. Furthermore, triviality in Bradley's sense is harmless here. It would say that, for Tautological Stability to hold, it would have to be either that ϕ entails $\top$ or $(\phi > \top)$ entails $\top$. But both of these facts are already known to hold. Trivialisation in Bradley's sense is only a concern when the consequent of the relevant conditional is not tautological.

The second is that, as Mandelkern and Khoo (2018) have shown, Local Preservation is crucially different from Bradley's original principle.[14] Suppose that the conditional is non-trivial in Bradley's sense, there are non-conditional ϕ and ψ, where neither $\phi \models \psi$ nor $\phi > \psi \models \psi$. It follows that there is some world where Preservation is violated, where $Pr_w(\llbracket\phi\rrbracket^c) > 0$, $Pr_w(\llbracket\phi > \psi\rrbracket^c) > 0$ but $Pr_w(\llbracket\psi\rrbracket^c) = 0$. We do *not* get to assume that the relevant world w is the world of the context. Indeed, given the very idea that conditionals are information-sensitive, it seems sensible to stipulate as an extra condition on probability measure that this never happens.[15]

Why we were only able to derive Local Preservation and not the full blown principle? Because my original statement of Stability was itself local: it said that when you are sure of $\llbracket\psi\rrbracket$ and leave open $\llbracket\phi\rrbracket$, you should be sure of the proposition $\phi > \psi$ expresses in *your* context. Triviality ensues from the stronger principle:

[14] Mandelkern and Khoo (2018) prove a further, stronger triviality result from Preservation; but they also show that Local Preservation does not fall afoul of this result.

[15] Mandelkern and Khoo (2018) show how to derive such a constraint from more basic conditions on conditionals and probability and then prove a tenability result.

Global Stability. $\neg S \neg \phi, S\psi \models S(\phi > \psi)$

This says that, when you are sure of ϕ and ψ, you should be sure of *any* proposition expressed by $\phi > \psi$. But given the information-sensitivity of conditionals, only the weaker, local principle was plausible in the first place; and it suffices to capture the data that motivate Stability.

9 The Way Out

We have exhausted the options available to us, on the traditional understanding of information-sensitivity. I will now show that by endorsing Conditional Locality, the claim that the contribution of the conditional can depend on its *local* informational environment, we can do better. I will present my favoured implementation, also advocated in [redacted], and show how its additional flexibility allows us to state Antecedent Presupposition in a way that avoids Negative Introspection.

9.1 Conditional Locality

Local contexts were first posited to explain facts about presupposition projection. It is assumed that a sentence like

(34) Susie stopped smoking.

presupposes that Susie smoked: it is felicitous only if it is accepted in the global context that Susie smoked in the past. However, a sentence like

(35) If Susie used to smoke, she stopped smoking.

does not as a whole presuppose that Susie used to smoke.

Why the contrast? The now standard explanation is that a sentence's presuppositions need only be satisfied in their *local* context, which can contain information present in the sentence but not in the global context. For instance, the local context for the consequent of a conditional is usually thought to include the information in the antecedent; so the presupposition of (34) is satisfied in its local context, allowing the sentence as a whole to presuppose nothing about Susie's smoking habits.

First posited by Stalnaker (1974) and Karttunen (1977), a general theory of local contexts have been rigorously worked out by Schlenker (2009b).

There are important precedents for putting local contexts in the semantics of epistemic vocabulary. Epistemic modals give rise to *epistemic contradictions*: sentences like

(36) # It's raining but it might not be.

sound marked; but unlike Moorean contradictions, they continue to be marked in *embedded* contexts. A static tradition, including Yalcin (2007) and Mandelkern (2019), adds an information state parameter to the index and claims that sentences like (36) put inconsistent constraints on that parameter. An older dynamic tradition, including Veltman (1996) and Gillies (2001), claims that (36) imposes an inconsistent *update* on a context, whether that context be the global one or some local context introduced by an embedding expression. Either way, both families of views subscribe to something like the following:

Epistemic Locality. The contribution of an epistemic modal is sensitive to its local context.

Epistemic Locality has a byproduct that will be instructive for us. These views about epistemic modals validate their own kind of linking principle, *Transparency*:

Transparency. $\models S(\Diamond\phi) \equiv \neg S\neg\phi$

Transparency says that to be sure of *might* ϕ is just to not be sure whether ϕ. In a standard framework one might expect that relatively strong constraints on the introspection principles would be needed to deliver this. But in fact theories that subscribe to epistemic locality can deliver Transparency without any assumptions about the underlying accessibility relation.[16]

I will defend an analogous view about conditionals:

Conditional Locality. The contribution of a conditional is sensitive to its local context.

[16] See for instance Yalcin (2007, 2011) here.

This will help us for parallel reasons — it allows us to vindicate linking principles and Antecedent Presupposition without placing constraints on the accessibility relation.

9.2 The Local, Shifty Theory of Conditionals

I implement Conditional Locality in a static, variably strict framework for conditionals.[17]

First, I add a local context parameter, κ, to the index. I then build on the variably strict theory of indicatives, defended by Stalnaker (1975) and Kratzer (2012), where $\ulcorner \phi > \psi \urcorner$ is true iff the closest ϕ-worlds are ψ-worlds. Following Stalnaker, I take the closest worlds to be supplied by f, a selection function; but, as well as taking a world and a proposition, the selection function f also takes the local context as an argument:

Local, Shifty Conditionals. $[\![\text{if } \phi, \text{then } \psi]\!]^{c,w,\kappa} = 1$ if and only if: $\forall w' \in f_{c,\kappa}(w, [\![\phi]\!]^{c,\kappa}) : [\![\psi]\!]^{c,w',\kappa} = 1$

On the Local, Shifty theory, $\ulcorner \phi > \psi \urcorner$ is true just in case the selected ϕ-worlds, *relative to the local context*, are all ψ-worlds. This is how the theory implements Conditional Locality: in a sense, the local context tells the semantics which selection function to use for the conditional.

It is generally accepted that attitude operators shift the local context for material embedded within. To capture this, I modify the standard Hintikka semantics for *is sure that*. Following Schlenker (2009a), I assume that the local context introduced by an attitude predicate like *is sure that* at a world w is the set of worlds compatible with what the subject is sure of in w. So, in addition to shifting the world parameter, attitudes now shift the local context parameter too:

Shifty Hintikka Semantics. $[\![S\phi]\!]^{c,\kappa,w} = 1$ if and only if: $\forall w' \in R(w) : [\![\phi]\!]^{c,R(w),w'}$

[17]This is not to say I think only the static approach can resolve my dilemma: a dynamic semantics for conditionals, e.g. Dekker (1993), combined with the standard dynamic semantics for attitudes, e.g. Heim (1992), can capture these data, if a consistency presupposition is added in the natural way.

A solution based on Conditional Locality can also be implemented in the strict approach to indicatives defended by Gillies (2004), Gillies (2009), Rothschild (2013), and Willer (2017), where $\ulcorner \phi > \psi \urcorner$ says that $\phi \supset \psi$ holds throughout some fixed set of closest worlds.

Apart from this extra feature, Shifty Hintikka Semantics treats 'is sure that' as a necessity operator, just as the standard Hintikka semantics does.

Since diagonal validity will again be relevant, we must define truth at a context. We have assumed the relevant information in the global context is what the relevant agents should be sure of; and the local context for an *un*embedded sentence is taken to be just the global context. Putting this together, we get the following definition of truth at a context:

Truth at a context. ϕ is true at a context c iff $[\![\phi]\!]^{c,w_c,R_c(w_c)} = 1$

Now let's return to the premises of my original argument. In this framework, we can understand Antecedent Presupposition as placing a constraint on our local context parameter. This is spelled out as follows:[18]

Local Antecedent Presupposition. $[\![\phi > \psi]\!]^{c,w,\kappa}$ is defined only if $\kappa \cap [\![\phi]\!]^{c,\kappa} \neq \varnothing$

Put roughly, this says that a conditional $\phi > \psi$ gets a truth-value at a point of evaluation $\langle c, w, \kappa \rangle$ only if κ contains some worlds where the antecedent is true. Given the original motivation for the local context parameter, stating the presupposition in this way is extremely natural.

To ensure Stability is validated, we need to add some constraints to the selection function. As well as obeying localised versions of generally accepted principles,[19] I assume selection functions obey a localised version of Stalnaker's Indicative Constraint:[20]

Localized Indicative Constraint. If $\mathbf{A} \cap \kappa \neq \varnothing$, then $\forall w' \in \kappa : f_{c,\kappa}(w', \mathbf{A}) \subseteq \kappa$

[18]Note that I use $[\![\phi]\!]^{c,\kappa}$ as shorthand for $\{w | [\![\phi]\!]^{c,w,\kappa} = 1\}$.

[19]Namely:

Success. $f_\kappa(w, \mathbf{A}) \subseteq \mathbf{A}$

Minimality. If $w \in \mathbf{A}$, then $w \in f_\kappa(w, \mathbf{A})$.

Non-Vacuity. If $\kappa \cap \mathbf{A} \neq \varnothing$ then $f_{c,\kappa}(w, \mathbf{A}) \neq \varnothing$.

[20]See Stalnaker (1975) for the original formulation. Mandelkern (2020) also defends a localised version of this constraint, but states it as a definedness condition on the conditional.

This says that if there are ϕ-worlds in the local context, then the selected ϕ-worlds at any world in the local context must be in the local context. As well as having other important applications, [21] it suffices to establish Stability, one of the main premises of our argument:[22]

Fact 1. $\neg S_c \neg\phi, S_c\psi \models S_c(\phi > \psi)$

For us, the *manner* in which it does so is important. By our use of Conditional Locality, we can validate Stability *without placing any constraints on the accessibility relation*; as the reader can check, the proof of fact 1 makes no assumptions about R. This is not possible in standard frameworks. This feature, the same one that yields Transparency in the epistemic case, is a necessary component of my solution.

9.3 Avoiding Negative Introspection

The Local, Shifty theory vindicates the premises of my argument, or something like them. But it also has the flexibility needed to evade Negative Introspection. To see this, let's first walk through some crucial results about it.

On the Local, Shifty theory, even though we have a version of Antecedent Presupposition, If-to-might is not valid simpliciter. Why not? Consider what happens when there is a mismatch between the local context parameter and the body of information picked out by S_c. κ might pick out the information of someone outside the conversation. Local Antecedent Presupposition imposes a condition on κ, that there are ϕ-worlds in it. However, when there is simply no relationship between S_c and κ this will of course not impose any constraints on S_c. This gives us an important fact:[23]

Fact 2. $\not\models (\phi > \psi) \supset \neg S_c \neg\phi$

[21] Specifically, as shown in [redacted], it suffices to validate:

The Qualitative Thesis. $\neg S \neg\phi \supset (S(\phi > \psi) \equiv S(\phi \supset \psi))$

[22] See the Appendix for proofs of this and the other facts in the main text.

[23] See the Appendix for the full proofs of this and the following facts.

To see how this helps with our arguments, let's extend these considerations to show that $\ulcorner S_c(\text{if } \phi, \psi \supset \neg S_c \neg \phi)\urcorner$, step 4 of my argument, fails too. The crucial feature here is that conditionals and attitudes now behave quite differently under attitude operators, potentially consulting different bodies of information. To see whether $\ulcorner S_c(\phi > \psi)\urcorner$ holds at w_c and c, we check whether $\ulcorner \phi > \psi \urcorner$ holds at all worlds accessible from w_c *relative to the local context* $\ulcorner S_c(w_c)\urcorner$. Since our version of Antecedent Presupposition only puts a constraint on the local context, this only places a constraint on the information in the global context i.e. what the speaker in sure of in the actual context, even in this kind of embedding.

Compare now $\ulcorner S \neg S \neg \phi \urcorner$. To see whether this holds at w and c, we check whether $\ulcorner \neg S \neg \phi \urcorner$ holds at every world accessible from w; that is, we check whether every world w' accessible from w in turn sees some world w'' where ϕ holds.[24] However, when introspection principles fail, then the worlds accessible from w may diverge significantly from the worlds accessible from some world accessible from w.

For precisely this reason, $\ulcorner S_c(\text{if } \phi, \psi \supset \neg S_c \neg \phi)\urcorner$ admits counterexamples. As we saw in our explanation of Fact 1, If-to-might fails when the local context for the conditional diverges from S_c. When introspection principles fail, this can happen again in embedded contexts. When embedded under $\ulcorner S_c \urcorner$ here, the local context for $\ulcorner \phi > \psi \urcorner$ happens to be the global context; so $\ulcorner \phi > \psi \urcorner$ places a constraint on the information in the global context $S_c(w_c)$. When embedded under $\ulcorner S_c \urcorner$, $\ulcorner \neg S_c \neg \phi \urcorner$ checks the information which, for all we are sure of in the global context, might be what we are sure of. When introspection principles fail, these two bodies of information can come apart: your actual information diverges from what your information might be, given your information. And it is in exactly such situations that If-to-might to fail, giving us:[25]

Fact 3. $\not\models S_c((\phi > \psi) \supset \neg S_c \neg \phi)$

Because of these results, the Local, Shifty theory escapes my argument. *If-to-might* is not valid; and step 4, $\ulcorner S_c(\text{if } \phi, \psi \supset \neg S_c \neg \phi)\urcorner$, has counterexamples when

[24] In the non-modal case, at least.

[25] See the proof of Fact 3 for a model where both $\ulcorner S_c((\phi > \psi) \supset \neg S_c \neg \phi)\urcorner$ and Negative Introspection fail.

introspection principles fail.[26] So, on the Local, Shifty theory, my argument fails at the fourth step.

But while it does not validate If-to-might, the Local, Shifty theory still does justice to the intuitions that motivate Antecedent Presupposition. Like the theory in §6.3, If-to-might is *diagonally* valid; it is true in every context. Where now $\phi \models_D \psi$ iff when $[\![\phi]\!]^{c,w_c,R_c(w_c)} = 1$, $[\![\psi]\!]^{c,w_c,R_c(w_c)} = 1$, we have:

Fact 4. $\models_D (\phi > \psi) \supset \neg S_c \neg \phi$

Because neither are embedded under an attitude, both $\ulcorner \phi > \psi \urcorner$ and $\ulcorner \neg S_c \neg \phi \urcorner$ consult the information in the global context, the former because here its local context is the global context; and the latter because it is evaluated at the same world as the entire sentence.

This is enough to capture the basic data that motivate Antecedent Presupposition. Suppose someone asserts:

(1) #If it rains, there won't be a picnic but it isn't going to rain.

Since *If-to-might* is diagonally valid, the left conjunct of (1) communicates that the speaker isn't sure whether it will rain; and the right conjunct communicates that they *are* sure it will rain. So we correctly predict that (1) commits the speaker to a contradiction.

But unlike the theory in §6.3, we also capture the embedding data that sunk Rigidified Antecedent Presupposition. First, notice we validate *simpliciter* a principle closely related to If-to-might, which I'll call Certain-if-to-might:

Certain-if-to-might. $S(\phi > \psi) \supset \neg S \neg \phi$

This principle says that if you should be *sure* of a conditional, then you should leave open its antecedent. This principle is *strictly* valid and not merely diagonally valid:

Fact 5. $\models S(\phi > \psi) \supset \neg S \neg \phi$

Since the conditional is under the sureness operator, its local context are the worlds

[26] In fact, as Fact 6 in the Appendix demonstrates, the latter principle is not even *diagonally* valid.

accessible from the world of evaluation; and since the consequent is not embedded under an operator, it makes a claim about the same body of information.

Certain-if-to-might then allows us to predict the embedding data. Recall:

(37) # John is imagining that I am sure that it's not raining and am sure that if it is raining, I will get wet.

Since Sure-if-to-might holds, the right conjunct of what John is imagining entails that I am not sure whether it is raining, which contradicts the first conjunct. So because I validate Sure-if-to-might, we predict that in a sentence like (37), John is imagining that I am sure of contradictory propositions.

In a way, the additional flexibility of Conditional Locality allows us to combine the virtues of Indexical Antecedent Presupposition and the virtues of Rigidified Antecedent Presupposition. Since local contexts are shifted by attitudes, we validate Certain-if-to-might, which allows us to employ the basic idea of Antecedent Presupposition even in embedded contexts; but the same shifting behaviour merely diagonally validates our original If-to-might principle, which allows us to block the argument for Negative Introspection.

10 Conclusion

Against a relatively weak background logic, I showed that two plausible principles together deliver a form of Negative Introspection, given the standard understanding of information-sensitivity. The first principle, Stability, is a downstream consequence of linking principles like the Ramsey Test and Stalnaker's Thesis; and the second, Antecedent Presupposition, is a well-motivated constraint on the definedness conditions of indicatives.

On reflection, this is not entirely surprising. Linking principles link doxastic attitudes towards conditionals to doxastic attitudes about non-conditionals. But on the traditional understanding of information-sensitivity, doxastic attitudes towards conditionals just are doxastic attitudes about our attitudes. Seen in this way, one might have suspected introspection principles would lie downstream of the linking principles.

I showed that Conditional Locality can resolve this tension between our conditional epistemology and epistemology at large. I gave an example of such a theory, the local, shifty theory of conditionals, and showed how it invalidates the argument. It vindicates the idea that conditionals are *transparent*, that believing or knowing a conditional just amounts to believing or know some first-order claim, rather than some claim about our information. And it does so without placing any constraints on our underlying attitudes: conditionals are still transparent, even if the attitudes do not iterate freely. Assuming Conditional Locality, transparency and iteration principles come apart, allowing us to reconcile linking principles with externalist views about iterated attitudes.

A Proofs

Our language $\mathcal{L}$ is the smallest set of sentences generated by the following grammar:

$$\phi ::= p \mid \neg\phi \mid \phi \wedge \psi \mid \phi > \psi \mid S\phi \mid S_c\phi$$

A *shifty frame* $\mathcal{F}$ for $\mathcal{L}$ is a tuple $\langle W, R, R_c, f_\kappa \rangle$. W is a non-empty set of worlds. R and R_c are serial accessibility relations. f_κ is a shifty selection function, a function from $\mathcal{P}(W)$ to a selection function. Say that a *minimal* shifty frame imposes the following constraints on f_κ:

Success. $f_\kappa(w, \mathbf{A}) \subseteq \mathbf{A}$

Minimality. If $w \in \mathbf{A}$, then $w \in f_\kappa(w, \mathbf{A})$.

Non-Vacuity. If $\kappa \cap \mathbf{A} \neq \varnothing$ then $f_\kappa(w, \mathbf{A}) \neq \varnothing$.

Localized Indicative Constraint. If $\mathbf{A} \cap \kappa \neq \varnothing$, then $\forall w' \in \kappa : f_\kappa(w', \mathbf{A}) \subseteq \kappa$

Our points of evaluation are tuples of a context, represented by a world, worlds and a local context, i.e. a set of worlds in W. We recursively define truth at a point of evaluation as follows:

$[\![p]\!]^{c,w,\kappa} = 1$ iff $w \in V(p)$

$[\![\neg\phi]\!]^{c,w,\kappa} = 1$ iff $[\![\phi]\!]^{c,w,\kappa} = 0$

$[\![\phi \wedge \psi]\!]^{c,w,\kappa} = 1$ iff $[\![\phi]\!]^{c,w,\kappa} = [\![\psi]\!]^{c,w,\kappa} = 1$

$[\![\phi > \psi]\!]^{c,w,\kappa}$ is defined only if $\kappa \cap [\![\phi]\!]^{c,\kappa} \neq \varnothing$
If defined, $[\![\phi > \psi]\!]^{c,w,\kappa} = 1$ iff $f_\kappa(w, [\![\phi]\!]^{c,\kappa}) \subseteq [\![\psi]\!]^{c,\kappa}$

$[\![S\phi]\!]^{c,w,\kappa} = 1$ iff $\forall w' \in R(w) : [\![\phi]\!]^{c,R(w),w'} = 1$

$[\![S_c\phi]\!]^{c,w,\kappa} = 1$ iff $\forall w' \in R_c(w) : [\![\phi]\!]^{c,R_c(w),w'} = 1$

where $[\![\phi]\!]^{\kappa} = \{w : [\![\phi]\!]^{c,w,\kappa} = 1\}$. We define truth at a context as follows:

Truth at a context. ϕ is true at a context c iff $[\![\phi]\!]^{c,c,R_c(c)} = 1$

Finally, we define validity and diagonal validity (i.e. preservation of truth at a context):

$\phi_1,...,\phi_n \models \psi$ just in case if $[\![\phi_1]\!]^{c,w,\kappa} = ... = [\![\phi_n]\!]^{c,w,\kappa} = 1$ then $[\![\psi]\!]^{c,w,\kappa} = 1$

$\models \phi$ iff $\varnothing \models \phi$

$\phi_1,...,\phi_n \models_D \psi$ just in case if $[\![\phi_1]\!]^{c} = ... = [\![\phi_n]\!]^{c} = 1$ then $[\![\psi]\!]^{c} = 1$

$\models_D \phi$ iff $\varnothing \models_D \phi$

We prove the following facts:

Fact 1. Where ϕ and ψ are non-conditional, $\neg S\neg\phi$, $S\psi \models S(\phi > \psi)$.

Proof. Suppose that $[\![\neg S\neg\phi]\!]^{c,w,\kappa} = [\![S\psi]\!]^{c,w,\kappa} = 1$. Then for some w' in $R(w)$: $[\![\phi]\!]^{c,w',R(w)} = 1$ and for all w'' in $R(w)$: $[\![\psi]\!]^{c,w',R(w)} = 1$. For contradiction, suppose $[\![S(\phi > \psi)]\!]^{c,w,\kappa} = 0$. Then for some w''' in $R(w)$: $[\![\phi > \psi]\!]^{c,w''',R(w)} = 0$ i.e. for some w'''' in $f_{R(w)}(w''', [\![\phi]\!]^{c,R(w)})$: $[\![\psi]\!]^{c,w'''',R(w)} = 1$. Since $R(w)$ contains a world where $[\![\phi]\!]^{c,R(w)}$ holds, we know the Local Indicative Constraint applies to $f_{R(w)}(w''', [\![\phi]\!]^{c,R(w)})$ i.e. $f_{R(w)}(w''', [\![\phi]\!]^{c,R(w)}) \subseteq R(w)$. But this then means that for some world w''''' in $R(w)$: $[\![\psi]\!]^{c,w''''',R(w)} = 0$. This contradicts our initial assumptions.

Fact 2. $\not\models (\phi > \psi) \supset \neg S_c\neg\phi$

Proof. Take some frame where $\langle c, w_1, \kappa\rangle$ is such that $\kappa = \{w_1, w_2\}$ and $R_c(w_1) = \{w_1, w_3\}$. Now define the interpretation function V so that $V(\phi) = \{w_2\}$ and $V(\psi) = \kappa$. Here we can see that $[\![\phi > \psi]\!]^{c,w_1,\kappa} = 1$: by the Localized Indicative Constraint and Non-Vacuity, $f_{c,\kappa}(w_1, [\![\phi]\!]^{c,\kappa}) = \{w_2\} \subseteq \kappa$. But clearly $[\![S_c\neg\phi]\!]^{c,w_1,\kappa} = 1$ so $[\![\neg S_c\neg\phi]\!]^{c,w_1,\kappa} = 0$

Fact 3. $\not\models S_c((\phi > \psi) \supset \neg S_c\neg\phi)$

Proof. Take some frame where $\langle c, w_1, \kappa\rangle$ is such that $R_c(w_1) = \{w_1, w_2\}$ and $R_c(w_2) = \{w_2\}$. Now define the interpretation function V so that $V(\phi) = \{w_1\}$ and $V(\psi) = \{w_1, w_2\}$. Here we can see that $[\![S_c(\phi > \psi)]\!]^{c,w_1,R_c(w_1)} = 1$: by the Localized Indicative Constraint, Minimality and Non-Vacuity, $f_{c,R_c(w_1)}(w_1, [\![\phi]\!]^{c,R_c(w_1)}) = f_{c,R_c(w_1)}(w_2, [\![\phi]\!]^{c,R_c(w_1)}) = \{w_1\}$ and $\{w_1\} \subseteq [\![\psi]\!]^{R_c(w_1)}$. But clearly $[\![S_c\neg\phi]\!]^{c,w_2,R_c(w_1)} =$

1 so $[\![\neg S_c \neg p]\!]^{c,w_2,R_c(w_1)} = 0$. So there's some $w \in S_c(w_1)$ such that $[\![(\phi > \psi) \supset \neg S_c \neg p]\!]^{c,w,R_c(w_1)} = 0$. So $[\![S_c((\phi > \psi) \supset \neg S_c \neg p)]\!]^{c,w_1,R_c(w_1)} = 0$.

Fact 4. $\models_D (\phi > \psi) \supset \neg S_c \neg \phi$

Proof. Suppose that, for some arbitrary $\langle c, w_c, R_c(w_c)\rangle$, $[\![\phi > \psi]\!]^{c,w_c,R_c(w_c)} = 1$. By Shifty Antecedent Presupposition $R_c(w_c) \cap [\![\phi]\!]^{c,R_c(w_c)} \neq \varnothing$. So, by Shifty Attitudes, $[\![\neg S_c \neg \phi]\!]^{c,w_c,R_c(w_c)} = 1$. So $[\![(\phi > \psi) \supset \neg S_c \neg \phi]\!]^{c,w_c,R_c(w_c)} = 1$.

Fact 5. For non-modal ϕ, $\models S(\phi > \psi) \supset \neg S \neg \phi$

Proof. Suppose that $[\![S(\phi > \psi)]\!]^{c,w,\kappa} = 1$. Then, for all w' such that $w' \in R(w)$ $[\![\phi > \psi]\!]^{c,w',R(w)} = 1$. However, given Localized Antecedent Presupposition, $[\![\phi > \psi]\!]^{c,w',R(w)} = 1$ only if $[\![\phi]\!]^{c,R(w)} \cap R(w) \neq \varnothing$. So $[\![\neg S \neg \phi]\!]^{c,w,R(w)} = 1$. Since, ϕ is non-modal, it follows that $[\![\neg S \neg \phi]\!]^{c,w,\kappa} = 1$

Fact 6. $\not\models_D S_c((\phi > \psi) \supset \neg S_c \neg \phi)$

Proof. As for Fact 2, with the further assumption that $w_1 = w_c$ and $\kappa = R_c(w_1)$.

References

Andrew Bacon. Stalnaker's thesis in context. *Review of Symbolic Logic*, 8(1): 131–163, 2015. doi: 10.1017/s1755020314000318.

Richard Bradley. A preservation condition for conditionals. *Analysis*, 60(3):219–222, 2000. doi: 10.1111/1467-8284.00228.

Paul Dekker. *Transsentential meditations: Ups and Downs in Dynamic Semantics.* ILLC Dissertation Series, Amsterdam, 1993.

Kevin Dorst. Abominable KK failures. *Mind*, page fzy067, 2019a. doi: 10.1093/mind/fzy067.

Kevin Dorst. Lockeans maximize expected accuracy. *Mind*, 128(509):175–211, 2019b.

Fred I. Dretske. Epistemic operators. *Journal of Philosophy*, 67(24):1007–1023, 1970.

Peter Gärdenfors. Belief revisions and the Ramsey Test for conditionals. *Philosophical Review*, 95(1):81–93, 1986. doi: 10.2307/2185133.

Anthony S. Gillies. A new solution to moore's paradox. *Philosophical Studies*, 105 (3):237–250, 2001.

Anthony S. Gillies. Epistemic conditionals and conditional epistemics. *Noûs*, 38 (4):585–616, 2004. doi: 10.1111/j.0029-4624.2004.00485.x.

Anthony S. Gillies. On truth-conditions for if (but not quite only if). *Philosophical Review*, 118(3):325–349, 2009. doi: 10.1215/00318108-2009-002.

John Hawthorne, Daniel Rothschild, and Levi Spectre. Belief is weak. *Philosophical Studies*, 173(5):1393–1404, 2016.

Irene Heim. Presupposition projection and the semantics of attitude verbs. *Journal of semantics*, 9(3):183–221, 1992.

Hans Herzberger. Dimensions of truth. *Journal of Philosophical Logic*, 4(2):535–556, 1973.

James R Hurford. Exclusive or inclusive disjunction. *Foundations of language*, 11 (3):409–411, 1974.

David Kaplan. Demonstratives: An essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals. In Joseph Almog, John Perry, and Howard Wettstein, editors, *Themes From Kaplan*, pages 481–563. Oxford University Press, Oxford, 1989.

Lauri Karttunen. Syntax and semantics of questions. *Linguistics and philosophy*, 1(1):3–44, 1977.

Lauri Karttunen and Stanley Peters. Conventional implicatures in montague grammar. In *Syntax and semantics, 11: Presupposition*. Academic Press, 1979.

Justin Khoo. The meaning of *If*. Ms, 2019.

Angelika Kratzer. Conditionals. In *Modals and Conditionals*, chapter 4. Oxford University Press, 2012.

Maria Lasonen-Aarnio. New rational reflection and internalism about rationality. In Tamar Szabo Gendler and John Hawthorne, editors, *Oxford Studies in Epistemology, Volume 5*. Oxford University Press, 2015.

Brian Leahy. Presuppositions and antipresuppositions in conditionals. *Proceedings of Semantics and Linguistic Theory*, pages 257–274, 2011.

David Lewis. *Counterfactuals*. Blackwell, 1973.

David Lewis. Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 85(3):297, 1976. doi: 10.2307/2184045.

Matthew Mandelkern. Bounded modality. *Philosophical Review*, 128(1):1–61, 2019.

Matthew Mandelkern. If p, then p! Ms, 2020.

Matthew Mandelkern and Justin Khoo. Against preservation. *Analysis*, 2018. doi: 10.1093/analys/any051.

Robert Nozick. *Philosophical Explanations*. Harvard University Press, 1981.

Frank Plumpton Ramsey. *The Foundations of Mathematics and Other Logical Essays*. Routledge & Kegan Paul, 1925.

Daniel Rothschild. Do indicative conditionals express propositions? *Noûs*, 47(1): 49–68, 2013. doi: 10.1111/j.1468-0068.2010.00825.x.

Daniel Rothschild. What it takes to believe. *Philosophical Studies*, forthcoming.

Bernhard Salow. Elusive externalism. *Mind*, 128(510):397–427, 2019. doi: 10.1093/mind/fzx015.

Philippe Schlenker. Local contexts. *Semantics and pragmatics*, 2:1–78, 2009a.

Philippe Schlenker. Local contexts. *Semantics and pragmatics*, 2:3–1, 2009b.

Declan Smithies. Belief and selfknowledge: Lessons from moore's paradox. *Philosophical Issues*, 26(1):393–421, 2016. doi: 10.1111/phis.12075.

Robert Stalnaker. A theory of conditionals. *American Philosophical Quarterly*, pages 98–112, 1968.

Robert Stalnaker. Probability and conditionals. *Philosophy of Science*, 37(1):64–80, 1970. doi: 10.1086/288280.

Robert Stalnaker. Pragmatic presuppositions. In Robert Stalnaker, editor, *Context and Content*, pages 47–62. Oxford University Press, 1974.

Robert Stalnaker. Indicative conditionals. *Philosophia*, 5(3):269–286, 1975.

Robert Stalnaker. *Inquiry*. MIT Press, 1984.

Bas van Fraassen. Probabilities of conditionals. In W. Harper and C. Hooker, editors, *Foundations of probability theory, statistical inference, and statistical theories of science*. D. Reidel, Dordrecht, 1976.

Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25 (3):221–261, 1996.

Kai von Fintel. The presupposition of subjunctive conditionals. In Uli Sauerland and Orin Percus, editors, *MIT Working Papers in Linguistics 25*, pages 29–44. MITWP, Cambridge, MA, 1998.

Malte Willer. Lessons from sobel sequences. *Semantics and Pragmatics*, 10, 2017.

Timothy Williamson. *Knowledge and its Limits*. Oxford University Press, Oxford, 2000.

Seth Yalcin. Epistemic modals. *Mind*, 116, 2007.

Seth Yalcin. Nonfactualism about epistemic modality. In Andy Egan and Brian Weatherson, editors, *Epistemic Modality*, pages 295–332. Oxford University Press, 2011.