A counterexample to modus ponenses

Matthew Mandelkern[*]

March 10, 2020

Abstract

McGee (1985) argued that modus ponens was invalid for the natural language conditional 'If... then...'. Many subsequent responses have argued that, while McGee's examples show that modus ponens fails to preserve truth, they do not show that modus ponens fails to preserve rational full acceptance, and thus modus ponens may still be valid in the latter *informational* sense. I show that when we turn our attention from indicative conditionals (the focus of most of the literature to date) to subjunctive conditionals, we find that modus ponens does not preserve *either* truth *or* rational full acceptance, and thus is not valid in either sense. In concluding I briefly consider how we can account for these facts.

1 Introduction

McGee (1985) argued that modus ponens was invalid for the natural language conditional 'If... then...'

on the basis of examples like the following:

> Opinion polls taken just before the 1980 election show the Republican Ronald Reagan decisively ahead of the Democrat Jimmy Carter, with the other Republican in the race, John Anderson, a distant third. Those apprised of the poll results believed, with good reason:
>
> (1)   If a Republican wins the election, then if it's not Reagan who wins it will be Anderson.
>
> (2)   A Republican will win the election.
>
> Yet they did not have reason to believe
>
> (3)   If it's not Reagan who wins, it will be Anderson.

A series of responses since have pointed out that, though examples like this certainly show *something* interesting about modus ponens, it is not clear exactly what it is. In particular, these responses have argued that, while cases like this might show that modus ponens is invalid in the sense that it does not preserve *truth*, these cases do not challenge modus ponens as a principle about *preservation of full acceptance* (sometimes called *informational consequence*). Indeed, it looks like, when you *fully accept* (1) and (2), then you are required to fully accept (3), and so examples like this are no counterexample to the informational version of modus ponens. In this paper I will argue that, while this response seems correct for indicative conditionals, the focus of most of the literature, it does not extend to subjunctive conditionals: in the case of subjunctives, *both* kinds of modus ponens, truth-conditional and informational, are invalid.

2   Two kinds of modus ponens

I begin by explaining the two kinds of modus ponens that have been distinguished in response to McGee's examples. My take on the basic idea is this.[1] McGee's counterexamples show that you can have high credence in ⌜p, and if p, then q⌝ without being logically compelled to have high credence in q (since you can have high credence in the conjunction of (1) and (2) without being compelled to have high credence in (3)). This suggests that modus ponens is not *truth-preserving*: if q is true in any case where ⌜p and if p, then q⌝ are both true, then, assuming rational credence can be modeled probabilistically, the laws of probability would rule out this state of affairs.[2] So McGee's counterexample is a good counterexample to the rule that says that the inference from p and ⌜If p, then q⌝ to q is *truth-preserving*. Call this rule *truth-conditional modus ponens*. Truth-conditional modus ponens is one natural way of spelling out what modus ponens amounts to, and McGee's examples indeed appear to show that it is not valid.

Nonetheless, there is a different way of spelling out modus ponens which these examples do not refute: namely, as the principle that if you *fully accept* both p and ⌜If p, then q⌝, then you are rationally

---

[1]The basic idea, in differing formulations, goes back at least to Over 1987. I am here following most closely discussion in Santorio 2020.

[2]I will assume that conjunction has classical properties which allow us to move freely between the two-premise formulation of modus ponens and the corresponding formulation with one conjunctive premise. This simplifies discussion but is not essential.

required to fully accept q. Call this rule *informational modus ponens*. On reflection, McGee's case does not look like a counterexample to informational modus ponens. After all, in McGee's case, if you really fully accept that a Republican will win, in the sense of *fully* ruling out that a Democrat won (suppose a perfectly reliable oracle tells you that a Republican will win, or a perfectly reliable newspaper prints 'Carter loses!'), then you must indeed accept that, if it's not Reagan who wins, it's Anderson. For if you get conclusive evidence that a Republican will win, without learning which Republican it will be, then you leave open just two options: Reagan and Anderson. And so it seems that you ought to accept that, if it's not Reagan, it's Anderson.

This line of response is spelled out in detail in a number of places. Bledin (2015) in particular (building on work in Veltman 1996; Gillies 2004, 2009; Yalcin 2007) proposes a conception of logic on which it is 'a science that is fundamentally concerned not with the preservation of truth but rather with the preservation of structural properties of the bodies of information that we generate, encounter, absorb, and exchange as we interact with one another and learn about our world'. On that *informational* conception, 'an argument is logically valid if and only if any body of information that incorporates all of its premises also incorporates its conclusion by virtue of logical form'. When we think in these terms, McGee's examples indeed do not threaten the validity of modus ponens, for the reason just given: if you fully incorporate the premises into your information, then you ought to incorporate the conclusion as well.[3]

In sum, then, it seems that there are (at least) two reasonable ways of saying precisely what modus ponens amounts to:

1. *Truth-conditional modus ponens*: the inference from p together with ⌜If p, then q⌝ to q always preserves truth.

2. *Informational modus ponens*: the inference from p together with ⌜If p, then q⌝ to q always preserves rational full acceptance.

McGee's counterexamples show the first principle is invalid, but they do not show that the second principle is invalid.

---

[3]There are different ways that the notion of informational consequence could be spelled out, but differences in detail won't matter for our purposes. Cf. Norlin 2019 for related discussion; see also Stern and Hartmann 2018 for a slightly different, but related, distinction between two kinds of modus ponens.

3   The problem with subjunctives

The literature on modus ponens has focused on indicative conditionals (like the conditional in (1)), rather than subjunctives (like 'If a Republican had won, then if it hadn't been Reagan, it would have been Anderson'). There is a natural reason for this: as McGee points out, it is difficult to assess modus ponens for conditionals in the subjunctive mood, because subjunctives tend to communicate that their antecedents are false (e.g. 'If the picnic had been cancelled, Susie would have been upset' generally communicates that the picnic *wasn't* cancelled).[4] Levi (1996, p. 106), Brogaard and Salerno (2008), and Briggs (2012) have attempted to formulate counterexamples to modus ponens in the subjunctive mood, but the results are not compelling, to my ear, for precisely the reason McGee points out. Here is Briggs's case, for illustration:

> A firing squad (consisting of two executioners, X and Y) is about to shoot a prisoner. If the court orders the execution, this will cause the captain to signal to the executioners. Each executioner is posed to fire if, and only if, the captain signals. Either executioner's shot by itself will be enough to kill the prisoner. As a matter of fact, the court does order the execution, so that the captain signals, the executioners shoot, and the prisoner dies.

Briggs then considers this conditional:

(4)   If executioner X had fired, then (even) if the captain had not signalled, the prisoner would have died.

The thought then seems to be that (4) and 'Executioner X fired' can both be true, while '(even) if the captain had not signalled, the prisoner would have died' is not. It feels hard to assess this claim, however, because it is hard to assess (4) in a context like the one given, where we know that Executioner X fired. In that case—when we know that the antecedent is true—my intuitions about (4) feel very unclear. So, while I am sympathetic to Briggs's aims, I do not think Briggs's case yields very clear judgments. Similar points go for Levi's and Brogaard and Salerno's cases, which have the same structural problem.[5]

---

[4]McGee speculates that modus ponens is likely invalid for subjunctives, since import-export appears to be valid for subjunctives, and there is a tension between modus ponens and import-export drawn out by Gibbard (1981); however, see Etlin 2008 for potential counterexamples to import-export for subjunctives which make this line of argument less convincing.

[5]A complication about evaluating modus ponens for subjunctives concerns mood. In English, there will often be morpho-

4 Subjunctive modus ponens

Importantly, however, not all subjunctives suggest that their antecedents are false. Anderson (1951) famously made this point with (5):

(5)   If Jones had taken arsenic, he would have shown just exactly those symptoms which he does in fact show.

As Anderson observes, (5) can be used as an argument that Jones has in fact taken arsenic, and thus does not always communicate that Jones hasn't taken arsenic. We can build on Anderson's paradigm to overcome the confound that faces examples like Briggs's. In this section, I will give two cases along these lines, and show that in these cases, neither truth-conditional nor informational modus ponens appears to be valid.

First, suppose that a doctor and a nurse are observing a patient Jones. Jones is displaying symptoms characteristic of people who have bronchitis and are in genotype A, B, or C (say Jones has a cough and fatigue, but no fever). The doctor is trying to convince the nurse that Jones has bronchitis. In doing so, she asserts (6):

(6)   If Jones had had bronchitis, then if he had been in genotype A, he would be showing the symptoms he in fact is showing.

She might carry on with:

(7)   And, if Jones had had bronchitis, then if he had been in genotype B or C, he would also be showing the symptoms he in fact is showing.

---

logical mismatches between the non-conditional premise of a modus ponens inference and the antecedent of the relevant subjunctive conditional. For instance, consider the conditional 'If Jones had now had bronchitis, then he would be coughing'; the relevant non-conditional premise has the form 'Jones now has bronchitis', rather than 'Jones had now had bronchitis'. The problem with insisting on exact identity between the conditional antecedent and the non-conditional premise is that 'Jones had now had bronchitis' is not felicitous on its own without making salient a past time, and does not intuitively express what the antecedent of this conditional does. By contrast, 'Jones now has bronchitis' is felicitous and seems to express the proposition which is expressed by the antecedent of the conditional in question. So it seems we are better off focusing on inference patterns which adjust the morphology of the conditional antecedent so that it is well-formed and intuitively expresses the same proposition as the conditional antecedent. Indeed, this is the standard practice in the study of indicative modus ponens, where we find similar mismatches. We could alternately study modus ponens for subjunctives while keeping the non-conditional premise in the subjunctive mood, but I will not explore that topic here.

Given the stipulated medical background, these conditionals are all plainly true. But now suppose further that genotype A is negatively correlated with bronchitis: although people in genotype A can get bronchitis—in which case they display Jones's actual symptoms—most people in genotype A are immune to bronchitis. So, given his symptoms, it's most plausible that Jones has bronchitis and is in genotype B or C; and that, if he *had* been in genotype A, he would not have gotten bronchitis in the first place.

Now focus on the conditional in (6). We are, again, certain that this conditional is true in the scenario. Moreover, its antecedent is highly plausible: given Jones's symptoms, we have every reason to have high credence in the proposition that Jones has bronchitis. But we do not have reason to have high credence in (6)'s consequent, repeated in (8):

(8)   If Jones had been in genotype A, he would be showing the symptoms he in fact is showing.

Again, given his symptoms—cough, fatigue but no fever—it is most likely that Jones has bronchitis but is in genotype B or C, and that if Jones *had* been in genotype A, he would not have gotten bronchitis in the first place. So we have high credence in the conjunction of (6) with its antecedent (since we are certain of (6), and very confident that Jones has bronchitis); but we do not have high credence in its consequent (8). Thus, in parallel with McGee's counterexamples to truth-conditional modus ponens for indicatives, we have a counterexample to truth-conditional modus ponens for subjunctives.

More interestingly, however, an elaboration of the present case constitutes a counterexample to *informational* modus ponens for subjunctives. For suppose we now get decisive evidence, by way of a conclusive bacterial assay, that Jones has bronchitis. We now *fully accept* the antecedent of (6). We also still fully accept (6), which is established as a matter of medical fact; if asked to review the relevant facts which first convinced us that the patient had bronchitis, we could perfectly well repeat (6) or (7). But we are not, it seems to me, logically compelled to accept the consequent of (6), even though we fully accept its antecedent. For even though we fully accept that Jones has bronchitis, we still think it is very likely he is in genotype B or C, and that, if he *had* been in genotype A, he would not have gotten bronchitis in the first place, and so would not be showing the symptoms he in fact is showing. And so we still think it is likely *false* that, if Jones had been in genotype A, he would be

showing the symptoms he in fact is showing. Thus we can fully accept p and ⌜If p, then q⌝, without being logically required to fully accept q. And so modus ponens does not look informationally valid for subjunctive conditionals.

This is in stark contrast to indicative conditionals. There, as we saw above, fully accepting p and ⌜If p, then q⌝ does appear to compel you to accept q. And indeed, if we consider an indicative variant of the present case, modus ponens again seems informationally valid: 'If Jones has bronchitis, then if he is in genotype A, he is showing the symptoms he in fact is showing' seems trivially true, and so does its consequent, 'if Jones is in genotype A, he is showing the symptoms he in fact is showing'; so whenever we fully accept the former and its antecedent, we must (trivially) fully accept the latter. By contrast, the present case suggests that the inference from p and ⌜If p, then q⌝ to q does not seem to be truth-conditionally *or* informationally valid when the conditional is in the subjunctive mood, specifically when q itself is a conditional.

It is worth noting that, as Paolo Santorio helpfully points out, it seems possible to bring out *a* reading of (8) which seems true and accepted in this situation. This is a reading that intuitively holds fixed the fact that Jones has bronchitis when we consider what would have happened if he had been in genotype A. The existence of such a reading, however, does not undermine the force of this case as a counterexample to modus ponens. Instead, it provides support for the widely accepted thesis that conditionals are context-sensitive. For modus ponens to be (truth-conditionally or informationally) valid if conditionals are context sensitive, it would have to be the case that, *for any context*, if p and ⌜If p, then q⌝ are both true/accepted, as interpreted in that context, then so is q, as interpreted in that context. In the case above, this appears not to hold: on the most contextually salient reading of (6), we accept (6) and its antecedent, but not its consequent. The existence of such a reading is enough to show that modus ponens is not (informationally or truth-conditionally) valid, even if there is *another* reading of (6) on which we are inclined to accept (6), its antecedent, and its consequent.[6]

Here is another case which illustrates the same point. Suppose the Baroness has been found dead. The two suspects are the gardener and the plumber. The clues at the crime scene suggest the gardener

---

[6]A different response, in a similar vein, would hold that there is some kind of *context-shift* between the premises and conclusion of the argument in the purported counterexample. But there is not much *prima facie* evidence for this kind of response, since there is apparently *one context*—the one we are in after we learn that Jones has bronchitis—where we have all the relevant information, accept (6) and its antecedent, and do not accept its consequent.

did it (freshly mown grass, mud), and are consistent with the gardener killing the Baroness with either a shovel or a wrench (the cause of death was blunt force trauma, but it's impossible to tell how big the weapon was). So the detective fully accepts (9) and (10), which she might assert by way of convincing her colleague that it was the gardener:

(9)   If the Baroness had been killed by the gardener, then if the Baroness had been killed with the wrench, then the crime scene would have looked exactly as it does.

(10)   If the Baroness had been killed by the gardener, then if the Baroness had been killed with the shovel, then the crime scene would have looked exactly as it does.

But given the gardener's profession, it is far more likely that, if the Baroness was killed by the gardener, then it was with the shovel rather than the wrench; whereas the plumber would have most likely done it with the wrench. And it is likely that, if the plumber (instead of the gardener) had killed the Baroness, the crime scene would look quite different than it in fact does (there would probably have been no grass or mud; instead, perhaps grease stains). So, in sum, the detective has high credence in the conjunction of (11) and (12) (since she is certain of (12) and confident in (11)), but not in (13):

(11)   The Baroness was killed by the gardener.

(12)   If the Baroness had been killed by the gardener, then if the Baroness had been killed with the wrench, then the crime scene would have looked exactly as it does.

(13)   If the Baroness had been killed with the wrench, then the crime scene would have looked exactly as it does.

Once again, then, we have a counterexample to truth-conditional modus ponens. And again, this can be converted to a counterexample to informational modus ponens. For suppose now that the detective finds evidence which decisively shows that the gardener did it (say, a credible confession). She doesn't yet know for certain what the murder weapon was: it's still possible that the gardener did it with the wrench to throw the detectives off her path, but it is much more likely that it was with the shovel, given that the gardener can access the latter more easily. Thus the detective still has every reason

to fully accept (12), and might, again, repeat (12) by way of explaining her preliminary conclusion that the Baroness was killed by the gardener. She also now fully accepts its antecedent in (11): she is certain that the gardener killed the Baroness. But she is not by any means now required to accept its consequent, in (13). After all, it is unlikely that the gardener would have used the wrench rather than the shovel to kill the Baroness. So it is likely that, had the Baroness been killed with the wrench, the plumber would have been the culprit, not the gardener; and then the crime scene might have looked very different than it in fact does look (grease instead of mud and grass). So the detective is under no obligation to now accept that, if the Baroness had been killed with the wrench, then the crime scene would have looked exactly as it does.

And so here, again, it looks like it can be rational to fully accept a subjunctive conditional and its antecedent, without coming anywhere near to fully accepting its consequent.

5 Capturing these facts

McGee's counterexamples taught us that modus ponens is not truth-conditionally valid. Much work has since observed that modus ponens still appears to be informationally valid. But that work has focused on indicative conditionals. The present cases appear to show that, when it comes to subjunctive conditionals, modus ponens is *neither* truth-conditionally *nor* informationally valid.

That modus ponens is not informationally valid for subjunctives does not, of course, undermine the insight that it *does* appear to be informationally valid for indicatives. But it shows that things are more complicated than they first appeared. An adequate theory of the conditional will have to predict the divergence between indicatives and subjunctives vis-à-vis modus ponens: that is, it will have to predict that modus ponens is informationally valid for indicatives, but not for subjunctives, whereas modus ponens is truth-conditionally invalid for both.

My primary aim here is not to defend a positive theory, but rather to make clearer the contours of the logic of the conditional. In this final section, however, I will briefly sketch a potential positive theory, to give a sense of how we might capture these logical facts.

Present considerations suggest that, if we want a broadly unified theory of the conditional, the most natural way to find it is to start with a baseline theory which invalidates both informational

and truth-conditional modus ponens, and then augment it in such a way that it validates informational modus ponens for indicatives but not for subjunctives. We can do this by starting with McGee (1985)'s theory. McGee's theory is a variant on Stalnaker (1968)'s theory which, unlike Stalnaker's theory, invalidates modus ponens.[7] I will work with a simple formal language comprising atoms $A, B, C \ldots$ closed under '$\wedge$' ('and'), '$\neg$' ('not'), and '$>$' ('if...then'). I use lower-case italics to range over sentences of the language (corner quotes are omitted for readability). On McGee's theory, sentences are evaluated relative to four parameters. The first is a Stalnakerian *selection function* f (Stalnaker, 1968; Stalnaker and Thomason, 1970) which takes a consistent proposition and a world to a world (intuitively, the "closest" one which makes the proposition in question true),[8] and which satisfies three constraints (where φ and ψ are any sets of worlds): (i) *Success*: $f(\varphi, w) \in \varphi$ (the selected world is in the proposition in question); (ii) *CSO*: $f(\varphi, w) = f(\psi, w)$ if $f(\varphi, w) \in \psi$ and $f(\psi, w) \in \varphi$ (this ensures that f can be interpreted in terms of a closeness ordering on worlds); and (iii) *Strong Centering*: $f(\varphi, w) = w$ if $w \in \varphi$ (the closest world to any world is itself). The second parameter is a set of sentences Γ (the *hypothesis set*), whose role is to keep track of conditional antecedents. The third parameter is a possible world w. The fourth is an atomic valuation function $\mathfrak{I}$ from atoms to sets of worlds. I write $[\![p]\!]^{\Gamma}$ for $\{w : [\![p]\!]^{\Gamma,w} = 1\}$; I leave relativization to $\mathfrak{I}$ and f implicit. Then we have:

- $[\![p]\!]^{\Gamma,w} = 1$ if $\bigcap_{r \in \Gamma} [\![r]\!]^{\varnothing} = \varnothing$; else

- $[\![A]\!]^{\Gamma,w} = 1$ iff $f(\bigcap_{p \in \Gamma} [\![p]\!]^{\varnothing}, w) \in \mathfrak{I}(A)$

- $[\![\neg p]\!]^{\Gamma,w} = 1$ iff $[\![p]\!]^{\Gamma,w} = 0$

- $[\![p \wedge q]\!]^{\Gamma,w} = 1$ iff $[\![p]\!]^{\Gamma,w} = 1$ and $[\![q]\!]^{\Gamma,w} = 1$

- $[\![p > q]\!]^{\Gamma,w} = [\![q]\!]^{\Gamma \cup \{p\},w}$

These clauses say, first, that any sentence is true if the hypothesis set is inconsistent.[9] Otherwise, an atomic sentence is true in w under the set of hypotheses Γ iff it is true in the closest world to w in

---

[7]Another natural starting point here could be Kratzer (1981)'s theory, which, while not explicitly motivated on the basis of counterexamples to modus ponens, invalidates it in very similar ways to McGee's theory.

[8]f is undefined on inconsistent propositions.

[9]McGee relativizes this condition to an accessibility relation, which I suppress here for simplicity since it does not matter for our purposes (so to be precise, the theory I present here is McGee's under the assumption that the relevant accessibility relation is universal).

which all the members of Γ are true. A negation is true at $\langle \Gamma, w \rangle$ iff the negatum is false at $\langle \Gamma, w \rangle$; a conjunction is true at $\langle \Gamma, w \rangle$ iff both conjuncts are. Finally, a conditional $p > q$ is true at $\langle \Gamma, w \rangle$ iff q is true at $\langle \Gamma \cup \{p\}, w \rangle$. We say a sentence is *simply true* at w iff it is true at w under the empty set of hypotheses.

To get a sense of how the theory works, consider a sentence with the form $A > B$. To evaluate this sentence at $\langle \varnothing, w \rangle$, we evaluate B at $\langle \{A\}, w \rangle$; B is true at $\langle \{A\}, w \rangle$ iff the closest A-world to w is a B-world. So $A > B$ ends up having the same truth conditions as it does in Stalnaker's more familiar theory. More generally, this theory is equivalent to Stalnaker's for any conditional whose consequent and antecedent do not themselves contain a conditional. So, in particular, whenever q does not contain a conditional, then, if $p > q$ and p are simply true at a given world on McGee's theory, q is also true at that world. So we validate truth-conditional modus ponens for conditionals whose consequents do not contain conditionals, which is as desired: the literature so far has turned up apparent counterexamples to modus ponens only with conditionals whose consequents contain conditionals (or modal language of some other kind). But when it comes to conditionals with conditional consequents, modus ponens is not truth-conditionally valid in this system—unlike in Stalnaker's. Consider a conditional with the form $A > (B > C)$. Suppose that this conditional is true at $\langle \varnothing, w \rangle$ and that A is true but B false at w. Then it follows that the closest $A \wedge B$-world to w is a C-world. But it does not follow that $B > C$ is true at $\langle \varnothing, w \rangle$. That would require the closest B-world to w to be a C-world; but it may well be that the closest B-world to w is a $\neg A$-world and not a C-world. So, as desired, we invalidate truth-conditional modus ponens for conditionals whose consequents themselves contain conditionals; the trick is using the hypothesis set to record successive conditional antecedents and using them all at once to find the closest world, rather than taking the antecedents each in turn, as in Stalnaker's theory.

Now, how can we predict that modus ponens is still *informationally* valid for indicatives? The trick is to combine McGee's theory with a proposal from Stalnaker 1975; von Fintel 1998 which says that indicative conditionals differ from subjunctives in the following respect: an indicative conditional presupposes that its consequent is evaluated at a world compatible with what is accepted. Sloganistically: indicative selection functions can't take us beyond what is accepted. The basic observation motivating this comes from contrasts like the following:

(14)   Sue's party was cancelled.

    a.   If it hadn't been cancelled, it would have been fun.

    b. #If it wasn't cancelled, it was fun.

The subjunctive continuation in (14-a) is fine, but the corresponding indicative in (14-b) is infelicitous. In general, there is something weird about using an indicative conditional $p >_i q$ once p has been ruled out—whereas the subjunctive $p >_s q$ remains fine. We account for this if we assume that indicatives, but not subjunctives, presuppose that the selected world is compatible with the context. Then, in a context where p has been ruled out, $p >_i q$ cannot have its presupposition satisfied (assuming p is consistent), since (by *Success*) the selected world must be a p-world. (In addition to accounting for cases like this, the other main motivation for this assumption is accounting for the felt goodness of the 'or'-to-'if' inference, on which see Stalnaker 1975.[10])

To implement this kind of presupposition—which I call the *indicative constraint*—in the present framework, let's use '$>_i$' and '$>_s$' to denote the indicative and subjunctive conditionals, respectively. We assume that '$>_i$' and '$>_s$' have exactly the same truth-conditions, namely those of McGee's '$>$'. '$>_i$' and '$>_s$' differ only in that '$>_i$' will also be associated with a presupposition as follows. Let c be a context: a set of worlds representing what is fully accepted in the conversation.[11] Then we say that $p >_i q$ as assessed at $\langle c, \Gamma, w \rangle$ presupposes that, for all $w \in c$, $f(\bigcap_{r \in (\Gamma \cup \{p\})} [\![r]\!]^{\varnothing}, w) \in c$.[12] In other words, $p >_i q$ presupposes that the closest world to any context world where p and everything in Γ is true is also a context world. By contrast, there is no similar constraint on the subjunctive conditional.

What is significant for present purposes is that marrying the indicative constraint with McGee's semantics for the conditional yields a logical profile which is very close to what we are looking for. In this framework, we can say that an inference from a set of sentences Φ to q is informationally valid just in case any context c which accepts all the elements of Φ also accepts q, provided every world in

---

[10]See also e.g. Gillies 2009; Leahy 2011; Dorr and Hawthorne 2018; Khoo 2019; Holguín 2018; Mandelkern 2019b; Boylan and Schultheis 2019 for more recent motivation and discussion.

[11]The details of how this is spelled out do not matter for us, though I will assume that the relevant notion is factive, to avoid conflicts with *Strong Centering*.

[12]I assume that the indicative constraint is checked in the course of computation, so that, e.g. in a conditional with the form $p >_i (q >_i r)$ assessed at Γ, the indicative constraint on $q >_i r$ is evaluated relative to *its* hypothesis set, i.e. $\Gamma \cup \{p\}$, not relative to Γ. To implement this formally, we simply need to integrate undefinedness into our system in a weak Kleene manner; I leave this implicit.

c satisfies the presuppositions of all the elements of Φ and of q, relative to c and the empty hypothesis set. And we say c accepts a sentence p just in case p is simply true at every world in c.[13] In the present system, modus ponens is truth-conditionally invalid for both indicatives and subjunctives. It is also informationally invalid for subjunctives. But it is informationally *valid* for (almost all) indicative conditionals.

The basic reasoning behind this is as follows. When the indicative constraint is satisfied, indicative antecedents always keep us inside the context. Now suppose a context c accepts $A >_i (B >_i C)$ and A, and the indicative constraint of both $A >_i (B >_i C)$ and of $B >_i C$ are satisfied throughout c relative to $\varnothing$. Does c also necessarily accept $B >_i C$? Consider an arbitrary world w in c. We know the closest $A \wedge B$-world to w must be a C-world, since c accepts $A >_i (B >_i C)$. By the indicative constraint on $B >_i C$, the closest B-world to w must be in c, and so must be an A-world, since c accepts A. But, by *CSO*, if the closest B-world to w is an A-world, then it must be the same world as the closest $A \wedge B$-world; and we know the closest $A \wedge B$-world is a C-world. So, since C is true at the closest B-world to w, $B >_i C$ is simply true at w. Since w was chosen arbitrarily, $B >_i C$ is simply true throughout c.

This reasoning generalizes to arbitrary indicative conditionals, with one caveat. Whenever a context accepts $p >_i q$ and p, then it must accept q as well, provided that all of $p >_i q$, p, and q have their indicative constraints satisfied throughout c, and provided that q does not contain a subjunctive conditional. (I will not explore here whether this latter caveat can be independently motivated or is a bug of this system.) I spell out this reasoning in a footnote.[14]

---

[13]This notion of informational validity follows Bledin 2018, and is the natural way to generalize informational validity to a system with presuppositions: it combines Yalcin (2007)'s notion of informational consequence as preservation of acceptance with von Fintel (1999)'s notion of Strawson entailment as preservation of truth in cases where premises and conclusion all have their presuppositions satisfied.

[14]For arbitrary p and q, with q not containing a subjunctive conditional, suppose that $p >_i q$ and p are accepted at c, relative to a selection function f and valuation function $\mathfrak{I}$. And suppose that the indicative constraints of $p >_i q$ and of q are satisfied throughout c. We show that q is accepted at c by showing that for arbitrary $w \in c$, q is true at $\langle c, \varnothing, w \rangle$ iff q is true at $\langle c, \{p\}, w \rangle$; since we know that $p >_i q$ is true at $\langle c, \varnothing, w \rangle$, we know q is true at $\langle c, \{p\}, w \rangle$, and so we will have that q is true at $\langle c, \varnothing, w \rangle$. We can reduce the question of the truth of q to two sets of questions: first, whether the hypothesis set remains consistent each time it is updated with a conditional antecedent in the course of computation; second, whether each atomic sentence which appears in q outside the antecedent of a conditional is true or false, relative to a corresponding hypothesis set. In our case, however, we can assume that the hypothesis set always remains consistent, since the indicative constraints of $p >_i q$ (where q does not contain a subjunctive conditional) are satisfied only if all the relevant hypothesis sets are consistent. So, where $\mathcal{A}_q$ is the sequence of atoms $A_1, A_2, \ldots A_n$ (in linear order) which appear in q outside of conditional antecedents, we can reduce the question of q's truth to the question of the truth or falsity of each $A_i \in \mathcal{A}_q$ relative to its corresponding hypothesis set Γ_i, which comprises all the antecedents under which A_i is embedded in q. For

By contrast, subjunctive conditionals lack the indicative constraint, so this reasoning won't go through: we can accept $p >_s q$ and p without accepting q, when q itself contains a subjunctive conditional. For instance, $A >_s (B >_s C)$ says that the closest $A \wedge B$-world is a C-world; while $B >_s C$ says the closest B-world is an C-world. But these two worlds need not be the same, even if we accept A, since the closest B-world can be outside the context and hence a $\neg A$-world.

To make all this reasoning more concrete, let's reconsider two examples from above, indicative and subjunctive, to see why modus ponens is informationally valid in the first case but not the second. Start with McGee's indicative example from the introduction. Suppose we accept 'If a Republican wins the election, then if it's not Reagan who wins it will be Anderson' and 'A Republican will win the election'. It follows that we accept 'If it's not Reagan who wins, it will be Anderson', provided all the relevant indicative constraints are satisfied. Let c be the context. Consider an arbitrary world w in c. 'If it's not Reagan who wins, it will be Anderson' is true at w just in case the closest world to w where Reagan doesn't win—call it w'—is one where Anderson does win. Given the indicative constraint, w' must be in c, and so must be a world where a Republican wins. w' also makes 'If a Republican wins the election, then if it's not Reagan who wins it will be Anderson' true, since we accept this by assumption. That means that the closest world to w' where a Republican wins the election and it's not Reagan is one where Anderson wins. But the closest world to w' where a Republican wins and it's not Reagan is w' itself, since these are both true at w'; so it follows that Anderson wins at w'. And so 'If it's not Reagan who wins, it will be Anderson' is true at w; since w was chosen arbitrarily, it is simply true throughout c.

Now compare the situation with subjunctive conditionals by returning to the Baroness example from above. Suppose that we accept 'The Baroness was killed by the gardener' and we accept 'If the Baroness had been killed by the gardener, then if the Baroness had been killed with the wrench, then

---

arbitrary A_i, by our clause for atoms and the consistency of Γ_i and $\Gamma_i \cup \{p\}$ (which are both guaranteed by the satisfaction of our indicative constraints) we have $[\![A_i]\!]^{c,\Gamma_i,w} = 1$ iff $f(\bigcap_{r\in\Gamma_i} [\![r]\!]^{c,\varnothing}, w) \in \mathfrak{I}(A_i)$. Likewise, $[\![A_i]\!]^{c,\Gamma_i \cup \{p\},w} = 1$ iff $f(\bigcap_{r\in(\Gamma_i \cup \{p\})} [\![r]\!]^{c,\varnothing}, w) \in \mathfrak{I}(A_i)$. Hence we show $f(\bigcap_{r\in(\Gamma_i \cup \{p\})} [\![r]\!]^{c,\varnothing}, w) = f(\bigcap_{r\in\Gamma_i} [\![r]\!]^{c,\varnothing}, w)$. Let b be the right-most conditional antecedent which embeds A_i in q. By the indicative constraint on the conditional with antecedent b, we know $f(\bigcap_{r\in\Gamma_i} [\![r]\!]^{c,\varnothing}, w) \in c$ and so $f(\bigcap_{r\in\Gamma_i} [\![r]\!]^{c,\varnothing}, w) \in [\![p]\!]^{c,\varnothing}$ (since p is accepted in c) and so by CSO $f(\bigcap_{r\in\Gamma_i} [\![r]\!]^{c,\varnothing}, w) = f(\bigcap_{r\in(\Gamma_i \cup \{p\})} [\![r]\!]^{c,\varnothing}, w)$. So $[\![A_i]\!]^{c,\Gamma_i,w} = [\![A_i]\!]^{c,\Gamma_i \cup \{p\},w}$. Since A_i was chosen arbitrarily from $\mathcal{A}_q$, this holds for all $A_i \in \mathcal{A}_q$. Since these values together determine the value of q, it follows that q is true at $\langle c, \{p\}, w\rangle$ iff q is true at $\langle c, \varnothing, w\rangle$, as desired.

the crime scene would have looked exactly as it does'. Unlike in the parallel situation for indicatives, it does *not* follow that we accept 'If the Baroness had been killed with the wrench, then the crime scene would have looked exactly as it does'. Let s be the context. Consider a world w in s where the Baroness was killed by the gardener with the shovel (which, recall, is compatible with what we accept). 'If the Baroness had been killed with the wrench, then the crime scene would have looked exactly as it does' is true at w just in case the closest world to w where the Baroness was killed with the wrench—call it w'—is one where the crime scene looks exactly as it does at w. Since our conditional is subjunctive, w' does not need to be in s. Indeed, plausibly the closest world where the Baroness is killed with a wrench is one where she was killed by the plumber, not the gardener—and thus is *outside* of s. We can further suppose that at w', the crime scene does not look as it does at w, since, at w', the crime was committed by the plumber, not the gardener. So 'If the Baroness had been killed with the wrench, then the crime scene would have looked exactly as it does' is false at w, and thus not accepted in s.

In sum, if we couple McGee's semantics for the conditional with the indicative constraint, we get something very like the logical profile which I have argued the conditional has vis-à-vis modus ponens. There are, of course, other ways we could capture this. In fact, I am pessimistic about McGee's theory for reasons independent from present concerns. Namely: in addition to invalidating modus ponens, it invalidates the identity principle, which says that ⌜If p, then p⌝ is a logical truth (Mandelkern, 2019a,b). In future work I hope to explore how to account for the facts about modus ponens sketched here without invalidating identity. But I will not explore alternate approaches here. My main goal in the paper is not to defend a positive theory, but rather, first, to argue that modus ponens is not informationally valid for subjunctive conditionals; and, second, to sketch one possible theory which captures both what is the same and what is different about the logics of indicatives and subjunctives as concerns modus ponens. Whatever its downsides, the theory I have given is conservative insofar as it simply combines two extant theories—McGee's theory of the conditional, together with the Stalnaker/von Fintel theory of the difference between indicative and subjunctive conditionals. Moreover, the general approach I have taken here—combining the indicative constraint with a unified baseline theory of the conditional which invalidates modus ponens—seems like it will be broadly applicable

across a variety of theories.

In closing, let me step back from questions about the semantics of conditionals to make a broader point about logic. In response to McGee's counterexamples to modus ponens, theorists have distinguished truth-conditional from informational modus ponens, arguing, correctly, that his counterexamples were only a problem for the first. Some went further, arguing that informational entailment in general is the correct notion for studying inference patterns, and that, once we see this, we can see that modus ponens remains valid simpliciter, i.e. valid on the notion of entailment appropriate for the study of inference in general. As Gillies (2017) puts it: 'There is no way to say what counts as "modus ponens" without saying what counts as "entailment". And where some might see modus ponens failures I'm more inclined to see a poor choice for a story about entailment'. The present considerations, I think, tell against this response. If we focus just on indicative conditionals, then it does indeed look plausible that we can save modus ponens by holding that informational consequence is, very generally, the correct notion of logical consequence for the study of inference. But once we see that modus ponens is informationally invalid for subjunctive conditionals, this route looks much less appealing, since modus ponens is not valid, even in the informational sense, for subjunctives. At least as a strategy for rescuing modus ponens, then, focusing exclusively on informational validity thus seems undermotivated. Of course, there might be independent reasons for thinking that informational consequence is in general the correct notion to study, though I am inclined to think that we must study both truth-conditional and informational consequence, and account both for cases where these notions converge and for cases where they diverge.

References

Anderson, A. (1951). A note on subjunctive and counterfactual conditionals. *Analysis*, 12(2):35–38.

Bledin, J. (2015). *Modus Ponens* defended. *Journal of Philosophy*, CXII(2):57–83.

Bledin, J. (2018). Fatalism and the logic of unconditionals. *Noûs*.

Boylan, D. and Schultheis, G. (2019). Attitudes, conditionals, and the qualitative thesis. Manuscript, University of Chicago and Rutgers, https://www.gingerschultheis.com/uploads/8/2/5/7/82578678/version_without_diagonal_strategy.pdf.

Briggs, R. (2012). Interventionist counterfactuals. *Philosophical Studies*, 160:139–166.

Brogaard, B. and Salerno, J. (2008). Counterfactuals and context. *Analysis*, 68(1):39–46.

Dorr, C. and Hawthorne, J. (2018). If...: A theory of conditionals. Manuscript, NYU and USC.

Etlin, D. (2008). Modus ponens revisited. Manuscript, MIT.

von Fintel, K. (1998). The presupposition of subjunctive conditionals. In Percus, O. and Sauerland, U., editors, *The Interpretive Tract*, volume 25, pages 29–44. MIT Working Papers in Linguistics, Cambridge, Massachusetts.

von Fintel, K. (1999). NPI licensing, Strawson entailment, and context dependency. *Journal of Semantics*, 16(2):97–148.

Gibbard, A. (1981). Two recent theories of conditionals. In Harper, W. L., Stalnaker, R., and Pearce, G., editors, *Ifs: Conditionals, Beliefs, Decision, Chance, and Time*, pages 211–247. Reidel, Dordrecht.

Gillies, A. (2004). Epistemic conditionals and conditional epistemics. *Noûs*, 38(4):585–616.

Gillies, A. (2009). On truth conditions for *If* (but not quite only *If*). *Philosophical Review*, 118(3):325–349.

Gillies, A. S. (2017). Conditionals. In Hale, B., Miller, A., and Wrighta, C., editors, *Blackwell Companion to the Philosophy of Language*. Blackwell.

Holguín, B. (2018). Knowledge in the face of conspiracy conditionals. Manuscript, NYU, `https://www.benholguin.com/uploads/1/1/3/6/113613527/k_cc_dec_2018.pdf`.

Khoo, J. (2019). The meaning of *If*. MS, MIT.

Kratzer, A. (1981). The notional category of modality. In Eikmeyer, H. and Rieser, H., editors, *Words, Worlds, and Contexts: New Approaches in Word Semantics*, pages 38–74. de Gruyter.

Leahy, B. (2011). Presuppositions and antipresuppositions in conditionals. In *Semantics and Linguistic Theory (SALT)*, volume 21, pages 257–274.

Levi, I. (1996). *For the Sake of Argument*. Cambridge University Press.

Mandelkern, M. (2019a). Crises of identity. In Schlöder, J. J., McHugh, D., and Roelofsen, F., editors, *Proceedings of the 22nd Amsterdam Colloquium*, pages 279–288. University of Amsterdam.

Mandelkern, M. (2019b). If p, then p! Manuscript, All Souls College, Oxford, `https://philpapers.org/rec/MANIPT-2`.

McGee, V. (1985). A counterexample to modus ponens. *The Journal of Philosophy*, 82(9):462–471.

Norlin, K. (2019). Surprises from hook and arrow in the logic of certainty. MS.

Over, D. (1987). Assumptions and the supposed counterexamples to modus ponens. *Analysis*, 47(3):142–146.

Santorio, P. (2020). Trivializing informational consequence. Manuscript, University of Maryland, College Park.

Stalnaker, R. (1968). A theory of conditionals. In Rescher, N., editor, *Studies in Logical Theory*, pages 98–112.

Oxford: Blackwell.

Stalnaker, R. (1975). Indicative conditionals. *Philosophia*, 5(3):269–86.

Stalnaker, R. C. and Thomason, R. H. (1970). A semantic analysis of conditional logic. *Theoria*, 36(1):23–42.

Stern, R. and Hartmann, S. (2018). Two sides of modus ponens. *The Journal of Philosophy*, 115(11):605–621.

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

Yalcin, S. (2007). Epistemic modals. *Mind*, 116(464):983–1026.