Counterfactuals, Infinity and Paradox

Andrew Bacon*

October 14, 2020

Abstract

In this paper I discuss two paradoxes of infinity, and draw out their implications for the logic of counterfactuals. I suggest the paradoxes may be resolved in a conditional logic that is independently motivated by considerations relating to the probabilities of conditionals. I compare the resulting view with alternative resolutions suggested by the theories of David Lewis and Kit Fine.

In this paper two paradoxes of infinity are considered through the lens of counterfactual logic, drawing heavily on a result of Kit Fine (Fine, 2012a). I will argue that a satisfactory resolution of these paradoxes will have wide ranging implications for the logic of counterfactuals. I then situate these puzzles in the context of the wider role of counterfactuals, connecting them to indicative conditionals, probabilities, rationality and the direction of causation, and compare my own resolution of the paradoxes to alternatives inspired by the theories of Lewis and Fine.

Here is a quick overview of the paper. Sections 1 and 2 introduce two paradoxes of infinity that rest on certain principles concerning the logic of counterfactuals. Section 3 considers three possible ways to weaken the counterfactual logic that would resolve these paradoxes, and examines three existing theories of conditionals of each sort: (Lewis, 1973), (Fine, 2012b) and (Bacon, 2015). With the first two theories found wanting, section 4 develops my preferred solution in (Bacon, 2015) a little further, and provides some simple models to show that the paradoxes can indeed be resolved in that framework.

1 Yablo's Button

Consider the following scenario, adapted from (Bacon, 2011).[1] Suppose that time has no beginning and that, in particular, there is no first day. On each

*I'd like to thank Kit Fine, both for giving me some helpful insights on how he is currently thinking about these issues, as well as providing comments on an earlier draft of the paper. I would also like to thank two anonymous referees and Frederik Van Der Putte for their very detailed comments on this paper, which resulted in many improvements in both presentation and substance.

[1] In (Bacon, 2011) the puzzle is presented as a supertask, but this is not necessary for the puzzle as it is presently stated.

day, a man must choose whether or not to press *Yablo's Button*. Yablo's Button is rigged so that it dispenses a chocolate on its first pressing, which the man will immediately convert into positive utils, and a painful zapping on subsequent pressings.[2] (More generally, the button has some mechanism that records whether it has been pressed before, and will zap if it has: there needn't be a first pressing.) For the man's convenience, a display is positioned above the button which reads 'Chocolate' if the button has never been pressed before, and 'Zap' if it has.

One might have thought that if the man is rational on day n he would behave as follows:

1. Press the button if it has not been pressed on any earlier day (i.e. if the display reads 'Chocolate').

2. Leave the button alone if it has been pressed on an earlier day (i.e. if the display reads 'Zap').

But a variant of Yablo's paradox demonstrates that it is impossible for someone to behave like this on every day. For convenience, suppose the days have been associated with integers in such a way that day $n-1$ immediately precedes day n. Suppose that the man follows the rules 1 and 2 on every day. It cannot be that he has never pressed the button, for then he has acted irrationally on day 0, say, for this is a day before which he has never pressed the button, and so he has forgone the chance to receive a chocolate on that day by pressing the button (violating 1).[3] So he must have have pressed the button on some day. If he was acting rationally then that day would be the first day he pressed the button, for otherwise he would be knowingly pressing a button that will zap him (violating 2). But then the preceding day is a day in which he forewent a chocolate, since it is a day on which the button went unpressed on every preceding day (violating 1 again).

It follows that no one can follow rules 1 and 2 on every day, and thus anyone finding themselves in such a situation must find themselves acting irrationally on some day or other. In (Bacon, 2011) I claimed that, while no rational being could exist and face such a sequence of decisions, this does not undermine the possibility of a rational being altogether. For provided one is not in fact presented with an infinity of decisions like this, one can nonetheless have the disposition to behave rationally with respect to any particular decision in the sequence. That is, the following counterfactuals are apparently consistent, and arguably true provided you are not in fact in the man's position:[4]

1′. If you were in the man's position on day n and the button had not been pressed on an earlier day, you would press the button.

[2] It is important that the rewards and punishments here are things that can happen immediately, as opposed rewards that may accumulate without being spent, such as winning money or amassing debts.

[3] Of course, by similar reasoning he has acted irrationally on any other day.

[4] If you are in the man's position you may derive the material conditionals 1 and 2, and reason to a contradiction as above.

$2'$. If you were in the man's position on day n and the button had been pressed on an earlier day, you would not press the button.

The joint consistency of $1'$ and $2'$ turned essentially on a similarity based semantics for the counterfactual in which the limit assumption failed. The informal idea is as follows. Suppose that w and w' are two worlds in which I face Yablo's Button on every day.[5] w is ranked as at least as close to actuality as a world w', written $w \preceq w'$, iff the set of days on which rules 1 or 2 are violated in w is a subset of the days on which they are violated in w'.[6] We then use Lewis's (Lewis, 1973) account of counterfactuals: a counterfactual, $A \Box\!\!\rightarrow B$ is true iff for some A-world, w, B is true at every A-world at least as close to actuality as w, or A is true at no worlds. $1'$ is true in this model for the following reason: there clearly is a metaphysically possible world where I'm stuck playing Yablo's button on day n, the button has been pressed on an earlier day, and I act rationally by not pressing the button on day n. Moreover, any world closer to actuality will have fewer instances of irrationality than this world does, so they will also be worlds where I do not press the button on day n. Thus the counterfactual $1'$ is true according to the Lewisian semantics. A parallel argument establishes the truth of $2'$ in this model.

To make this idea a little more explicit, let us regiment a little more. We will work in a propositional language containing, in addition to the standard truth functional connectives, a binary counterfactual connective $\Box\!\!\rightarrow$ and infinitary analogues of conjunction and disjunction.[7] The truth functional connectives, and their infinitary variants, will be assumed to be governed by classical logic.[8] We shall write Y for the proposition that I am playing Yablo's Button: that is, I face the choice of whether to press Yablo's Button on every day, the display is working properly, I can read it, and so on. Let D_n be the proposition that I decline to press the button on day n, and let $D_{\leq n}$ be the infinite conjunction $D_n \wedge D_{n-1} \wedge \ldots$: i.e. the proposition that I decline on every day up until day n. If I am rational then I ought satisfy the counterfactuals stating that I *would* choose the rational action on day n, if I were in the man's position on day n. In particular, if I were in his position and had declined on every previous day, I would press the button, and if I were in his position and had pressed the button on a previous day, I would decline:

[5] The ranking of worlds where you are not in the mans position are not important for evaluating 1' and 2'.

[6] The resulting similarity ordering allows for incomparable worlds, and thus follows (Pollock, 1976), rather than (Lewis, 1973) who posits a total ordering. A Lewisian semantics which secures the counterfactuals $1'$ and $2'$ is also possible: say that $w \preceq w'$ iff the longest string of violation-free days counting backwards from day 0 in w is is at least as long as in w'.

[7] More formally, we suppose that there are infinitely many sentence letters, $P_1, P_2, ...$ that each count as sentences, and that whenever $A_1, A_2, A_3, ...$ are sentences, so are $\neg A_1$, $(A_1 \wedge A_2)$, $(A_1 \vee A_2)$, $(A_1 \Box\!\!\rightarrow A_2)$, $\bigwedge_n A_n$ and $\bigvee_n A_n$.

[8] The notion of a valuation of the $\Box\!\!\rightarrow$-free fragment of the language is a mapping from sentences to truth values subject to the usual clauses for the truth functional connectives — e.g. that a conjunction (finitary or not) is true iff every conjunct is true. Classical entailment may be defined in the usual way. Any substitution instance of a classical entailment in the full language will also be counted as a classical entailment.

Chocolate Preference $Y \wedge D_{\leq n} \boxright \neg D_{n+1}$

Zap Avoidance $Y \wedge \neg D_{\leq n} \boxright D_{n+1}$

In (Fine, 2012a), Kit Fine presents a puzzling result in the logic of counterfactuals. One may interpret this result as a challenge to the truth of Chocolate Preference in conjunction with some reasonable sounding principles of counterfactual logic, and the assumption that it is counterfactually consistent that I be in the man's position and decline the button on an infinite set of days:[9]

Consistency $\neg(Y \wedge D_{\leq n} \boxright \neg(Y \wedge D_{\leq n}))$

In order to tailor the result to the present discussion, I will simplify Fine's argument in a couple of ways. First, I will slightly strengthen one of the premises Fine uses in his derivation: Disjunction below. Second, apart from this change, I will only use a smaller set of premises in my derivation. Because of this minor strengthening of Disjunction, and the redundancy of some of Fine's other premises, my proof will be significantly shorter. The principles of counterfactual logic are:

Identity $\vdash A \boxright A$

Substitution $A \boxright B \vdash A' \boxright B$ when A and A' are classically equivalent.[10]

Weakening $A \boxright B \vdash A \boxright B'$ when B entails B'.

Disjunction $A \boxright C, B \boxright C \vdash A \vee B \boxright C$

Infinite Conjunction $A \boxright B_1, A \boxright B_2, ... \vdash A \boxright \bigwedge_n B_n$

The additional premises in Fine's proof are: Finite Conjunction (a finitary version of Infinite Conjunction) and a principle sometimes called Restricted Transitivity (which Fine calls Transitivity), a statement and discussion of which may be found in section 3.4.[11] In (Fine, 2012a) Fine also uses a weaker version of Disjunction which has the added restriction that it may only be applied when A and B are logically exclusive, however Fine now accepts the unrestricted version of the rule.[12]

Identity and Weakening straightforwardly imply a principle I will follow Fine in calling Entailment, which states that the counterfactual $A \boxright B$ is true whenever A entails B. The result may be stated as follows:

Theorem 1.1. *Chocolate Preference and Consistency are inconsistent with the listed principles of counterfactual logic.*

[9] Counterfactual consistency, $\Diamond A$ is defined by $\neg(A \boxright \neg A)$, and counterfactual necessity, $\Box A$, by $(\neg A \boxright A)$.

[10] Where classical equivalence is spelled out in the sense of footnote 8.

[11] Transitivity is usually reserved for the extremely contentious rule $A \boxright B, B \boxright C \vdash A \boxright C$. Consider, e.g.: if I were to put on my coat I'd get warm, if I were to go to the arctic I'd put on my coat; but it doesn't seem true that if I were to go to the arctic I'd get warm.

[12] Fine p.c. See also (Fine, 2018a), (Fine, 2018b) and (Fine) for the truth-maker semantics in which Disjunction is valid.

Here is the proof:

1. $(Y \wedge D_{\leq -2}) \Box\!\!\rightarrow \neg D_{\leq -1}$ (from Chocolate Preference by Weakening)

2. $Y \wedge (D_{\leq -1} \vee D_{\leq -2}) \Box\!\!\rightarrow \neg D_{\leq -1}$ (from 1 by Substitution)

3. $(Y \wedge \neg D_{\leq -1} \wedge (D_{\leq -3} \vee D_{\leq -4} \vee ...)) \Box\!\!\rightarrow \neg D_{\leq -1}$ (by Entailment)

4. $(Y \wedge (D_{\leq -1} \vee D_{\leq -2}) \vee (Y \wedge \neg D_{\leq -1} \wedge (D_{\leq -3} \vee D_{\leq -4} \vee ...))) \Box\!\!\rightarrow \neg D_{\leq -1}$ (from 2 and 3 by Disjunction)

5. $Y \wedge (D_{\leq -1} \vee D_{\leq -2} \vee ...) \Box\!\!\rightarrow \neg D_{\leq -1}$ (by Substitution)

Note that in the step from 1 to 2, we are substituting a sentence of the form A for a classically equivalent sentence of the form $(B \wedge A) \vee A$, where $A = D_{\leq -2}$ and $B = D_{-1}$ (recall that by definition $D_{\leq -1} = D_{-1} \wedge D_{\leq -2}$).

By completely parallel reasoning we get $Y \wedge (D_{\leq -2} \vee D_{\leq -3} \vee ...) \Box\!\!\rightarrow \neg D_{\leq -2}$. But $(D_{\leq -2} \vee D_{\leq -3} \vee ...)$ is logically equivalent to $(D_{\leq -1} \vee D_{\leq -2} \vee ...)$ (every disjunct of the latter entails a disjunct of the former, and conversely). So we get $(Y \wedge (D_{\leq -1} \vee D_{\leq -2} \vee ...)) \Box\!\!\rightarrow \neg D_{\leq -2}$. And by parallel reasoning we get $Y \wedge (D_{\leq -1} \vee D_{\leq -2} \vee ...) \Box\!\!\rightarrow \neg D_{\leq n}$ for every n, and by Entailment we get $Y \wedge (D_{\leq -1} \vee D_{\leq -2} \vee ...) \Box\!\!\rightarrow Y$. So by Infinite Conjunction we get $Y \wedge (D_{\leq -1} \vee D_{\leq -2} \vee ...) \Box\!\!\rightarrow Y \wedge (\neg D_{\leq -1} \wedge \neg D_{\leq -2} \wedge ...)$, and by Weakening $Y \wedge (D_{\leq -1} \vee D_{\leq -2} \vee ...) \Box\!\!\rightarrow \neg(Y \wedge (D_{\leq -1} \vee D_{\leq -2} \vee ...))$, contradicting Consistency.

Before interpreting this result, let me turn to another paradox of infinity on which counterfactual logic has some bearing.

2 Benardete's paradox

In (Benardete, 1964) José Benardete presents us with the following paradox. A man decides that he will walk between two points, A and B, a mile apart. But in a stroke of poor luck, an infinite pantheon of gods lie in wait to thwart him. The first god resolves to build a wall at the $\frac{1}{2}$-mile mark if the man makes it that far. The second god likewise resolves to build a wall at the $\frac{1}{4}$-mile mark if the man makes it that far; more generally the nth god will build a wall at $(\frac{1}{2})^n$ mile mark if the man makes it that far. It may be seen that the man will not be able to pass point A. Suppose, for contradiction, he travels (continuously) some distance past point A. But if he got this far, he would have passed the $(\frac{1}{2})^n$th mile mark for some n, contradicting the assumption that the nth god would have halted him if he had gotten that far. So he cannot proceed past point A. The puzzle is this: if the man does not pass the $(\frac{1}{2})^n$th mile mark, for any n, then no god will have built their wall, for they have resolved only to build the wall if the man has made it as far as the $(\frac{1}{2})^n$-mile mark. He will thus not be able to pass beyond point A, even though no walls block his path: it is as though there is 'a strange field of force blocking his passage forward' (Benardete, 1964) p255.

The received opinion on Benardete's paradox, as expressed by Benardete himself, is that it is indeed metaphysically possible for the gods to form the described intentions, but in any such world the man does not pass point A.[13] But Fine's result in the logic of counterfactuals would, I claim, challenge this diagnosis, provided we help ourselves to the same *prima facie* plausible principles of counterfactual logic, underpinned by the established possible world semantics for counterfactuals of Lewis, Stalnaker and others. Indeed, this conclusion has been arrived at independently by Michael Caie (Caie, 2018), who similarly argues that the morals usually drawn from Benardete's paradox cannot hold if they are combined with the orthodox approach to counterfactuals.

Let us formalize the above a little. Let us write P_n to denote the proposition that the man has passed the $(\frac{1}{2})^n$th mile mark, and write $P_{\geq n}$ for the infinite conjunction $P_n \wedge P_{n+1} \wedge ....$ More plainly, $P_{\geq n}$ means that the man has passed all of the points $(\frac{1}{2})^k$ for $k \geq n$. Given that the man travels continuously, it's clear that he has passed the $(\frac{1}{2})^n$th mile mark iff he has passed the $(\frac{1}{2})^k$th mile mark for all $k \geq n$. Thus the laws of physics guarantee that P_n and $P_{\geq n}$ stand or fall together. But logic itself does not, so for the purposes of exploring the logical consequences of various assumptions we will continue to distinguish them. The first god is disposed to create a barrier if the man succeeds in passing all of the $\frac{1}{4}$-mile mark, $\frac{1}{8}$-mile mark, $\frac{1}{16}$-mile mark, and so on. Thus if the man had gotten this far (if the other gods hadn't formed similar intentions, for example), he wouldn't succeed in passing the $\frac{1}{2}$-mile mark. Employing our abbreviations, we have $P_{\geq 2} \mathrel{\Box\!\!\rightarrow} \neg P_1$, using $\mathrel{\Box\!\!\rightarrow}$ to represent the counterfactual conditional. The nth god forms a parallel intention, securing more generally:

Divine Dispositions $P_{\geq n+1} \mathrel{\Box\!\!\rightarrow} \neg P_n$

Of course (*pace* Zeno of Elea) it is entirely *possible* for a man to walk continuously between two points A and B, as corroborated by our daily movements. It's certainly possible that there be no gods intending to thwart the man's progress, no barriers, or anything else of that sort. In which case, the hypothesis that he passes at least one of the $(\frac{1}{2})^n$th-mile marks is *counterfactually consistent*: the hypothesis does not lead, counterfactually, to an absurdity.[14]

Anti-Zenoism $\neg(\bigvee_n P_{\geq n} \mathrel{\Box\!\!\rightarrow} \neg \bigvee_n P_{\geq n})$

So far I have said nothing about what the man will or won't try to do: I have not, for example, said that the man will travel past the $(\frac{1}{2})^n$-mile mark if there are no barriers impeding his passage. I have just reaffirmed the anti-Zenoean orthodoxy that it is consistent that he be able to move, and a premise representing the fact that the gods have formed the relevant intentions, and are

[13] (Yablo, 2000) considers a variant set of intentions which are jointly inconsistent, but stops short of concluding that the gods intentions in Benardete's paradox are themselves inconsistent. He writes 'But maybe the reason for contradiction is that the gauntlet [of demons] is in itself incoherent', but then softens his position to 'Or, rather, it is saved from incoherence only by the assumption that [the man] stops at [A].' p150.

[14] Note that the disjunction $\bigvee_n P_{\geq n}$ is stronger than the disjunction $\bigvee_n P_n$. The stronger disjunction states that it's possible that, for some n, the man travel the the $(\frac{1}{2})^n$-mile mark, and all of the preceding markers. This is clearly supported by the anti-Zenoean though, for it is clearly possible that the man move *continuously* to the $(\frac{1}{2})^n$-mile marker (as opposed to by teleportation), hitting all the intermediate points.

consequently disposed to block the man *if* he were to make it a certain distance. For all I've said, the man may be intending to just stand still — to not even attempt to reach B.

According to conventional wisdom, we do not have an inconsistency yet. To get the actual inconsistency we need a premise to the effect that the man will walk to B if the are no barriers blocking his path: i.e. that he's trying to walk to B and there are no 'invisible force-fields'. This is, of course, eminently plausible given Benardete's description. But it is a paradox, after all, and something plausible must be rejected: according to orthodoxy, it is exactly this premise which must be rejected — the man will attempt to walk to B but be blocked by an invisible force-field. The alternative is to embrace Zeno's conclusion, and reject the idea that it is even possible to move (Anti-Zenoism), or to reject the hypothesis that the gods could possess the relevant dispositions (Dispositions). While the Zenoean route has rightly found no adherents, I think the second diagnosis — that the gods simply cannot have these dispositions — has been unjustly neglected.[15] This is substantiated by the fact that Dispositions and Anti-Zenoism are inconsistent on their own, given a modicum of background counterfactual logic. This suggests there is a paradox even if the man decides to stand still, or walk in the opposite direction. Given this counterfactual logic, the assumption that the gods can have the relevant dispositions is inconsistent on its own, and the assumption that there are 'no invisible force-fields' — that the man will walk toward B absent any barriers — is not needed.

Theorem 2.1. *Divine Dispositions and Anti-Zenoism are inconsistent with the principles of counterfactual logic listed in section 1.*

We argue similarly:

1. $P_{\geq 2} \mathrel{\Box\!\!\rightarrow} \neg P_{\geq 1}$ (from Dispositions by Weakening)

2. $(P_{\geq 1} \vee P_{\geq 2}) \mathrel{\Box\!\!\rightarrow} \neg P_{\geq 1}$ (from 1 by Substitution)

3. $(\neg P_{\geq 1} \wedge (P_{\geq 3} \vee P_{\geq 4} ...)) \mathrel{\Box\!\!\rightarrow} \neg P_{\geq 1}$ (by Entailment)

4. $(P_{\geq 1} \vee P_{\geq 2} \vee (\neg P_{\geq 1} \wedge (P_{\geq 3} \vee P_{\geq 4} ...))) \mathrel{\Box\!\!\rightarrow} \neg P_{\geq 1}$ (from 2 and 3 by Disjunction)

5. $(P_{\geq 1} \vee P_{\geq 2} \vee ...) \mathrel{\Box\!\!\rightarrow} \neg P_{\geq 1}$ (by Substitution)

As before, we may generalize the above reasoning to derive $(P_{\geq 1} \vee P_{\geq 2} \vee ...) \mathrel{\Box\!\!\rightarrow} \neg P_{\geq n}$ for every n, and so by Infinite Conjunction and Weakening contradict Anti-Zenoism.[16]

[15] With the exception of Caie.

[16] Michael Caie similarly derives a contradiction from something analogous to Divine Dispositions and Anti-Zenoism with the following infinitary inference of counterfactual logic:

$$\bigvee_n A_n \mathrel{\Box\!\!\rightarrow} \bigvee_n (A_n \wedge \neg B_n), \bigwedge_n (A_n \mathrel{\Box\!\!\rightarrow} B_n) \vdash \bigvee (A_n \mathrel{\Box\!\!\rightarrow} \bot)$$

(Caie actually formulates it as the inconsistency of a triad of sentences. For comparison with principles like Infinitary Conjunction, I have reformulated it as a rule. The reformulation requires some modest principles of counterfactual logic that are not at issue

Granting, for the sake of argument, the listed principles of counterfactual logic, must we reject the possibility of the gods having the dispositions described? One might resist this conclusion in a couple of ways, neither of which I think stand up to scrutiny. Firstly, one might object that Anti-Zenoism, as it is presently stated, doesn't really capture the idea that it is consistent that we move, for it is stated in terms of *counterfactual* consistency, and this might be a lot narrower than commonly supposed. One view on which counterfactual possibility is narrower than, say, metaphysical possibility, is a view in which it coincides with physical necessity, so that counterfactuals with physically impossible antecedents are vacuously true. But then Anti-Zenoism states the still eminently plausible assumption that movement is physically possible. On a yet narrower understanding of counterfactual possibility, it is a counterfactual necessity that the gods have the intentions they in fact do, and each instance of Divine Dispositions becomes a vacuous truth. Of course, the most salient interpretation with this feature is that the counterfactual necessities are exactly the truths, in which case the counterfactual collapses into the material conditional — a view that has been repeatedly refuted, and as far as I know has no proponents.

Secondly, one might attempt to maintain that the gods have the relevant dispositions, but do not satisfy the counterfactuals stated in Divine Dispositions. For instance, a glass securely wrapped in bubble wrap has the disposition to break — it is still fragile — even though it would not break if dropped: this disposition has been *masked*.[17] But what is playing the role of bubble wrap in this case? It is not as though the gods have their hands tied behind their back, or there is something else preventing their dispositions from manifesting as counterfactuals. If the first god were the only god to exist, and he were to form the relevant disposition, then we surely would want to assert the corresponding counterfactual. Likewise, if he were accompanied by finitely many similarly dispositioned gods. What is it about having infinitely many similarly dispositioned gods that masks the dispositions?

There is also a sense in which this response misses the point. For Benardete's

here.) Unlike Infinite Conjunction, which appears to me to be unassailable (more on that later), this infinitary principle does not strike me as particularly compelling. It furthermore mixes into the infinitary rule aspects of the finitary fragment of the Lewis-Stalnaker logic I think are contentious and ought to be separated. (Indeed, I will ultimately recommend a logic that does not contain even the finitary versions of Caie's rule, such as the inference $A \mathrel{\Box\!\!\rightarrow} B, C \mathrel{\Box\!\!\rightarrow} D, (A \vee B) \mathrel{\Box\!\!\rightarrow} (A \wedge \neg B) \vee (C \wedge \neg D) \vdash (A \mathrel{\Box\!\!\rightarrow} \bot) \vee (C \mathrel{\Box\!\!\rightarrow} \bot)$.) Moreover, even if one grants the finitary principle it is not straightforward to extend the justification to the infinite case; Lewis, for instance, accepts the finitary version of the inference, but not the infinitary version. Fine's result helpfully separates the infinitary principle, Infinite Conjunction, from principles distinctive to possible worlds semantics in general, like Substitution, and these from principles that are specific to the similarity semantics of Lewis and Stalnaker, like Disjunction. Also implicit in Caie's formulation is a substantive principle about the connection between metaphysical and counterfactual necessity (defined as $\neg A \mathrel{\Box\!\!\rightarrow} \bot$): my formulation simply replaces occurrences of the former with the latter. This also makes for easier comparison with the systems being used here. So my focus, in what follows, will be Fine's result; although the relevance of what I have to say on Caie's argument should be evident throughout.

[17] See (Johnston, 1992).

paradox, we do not care about the possibility of the gods being disposed to stop him, but in such a way that the corresponding counterfactuals are false. Of course, one *might* maintain that in spite of our result, it is metaphysically possible for there to be infinitely many gods disposed to block the man, because it is metaphysically possible that they be so disposed while they all have their hands tied behind their backs (and so lack the corresponding counterfactual properties — instances of Divine Dispositions). But such a victory would be pyrrhic at best: it is the idea that there could be infinitely many gods with the counterfactual properties — gods with untied hands, and the actual potential to stop the man — that seems evidently possible; concessions such as the possibility that they could have the dispositional properties, but only if masked, are poor spoils.

3 Revising counterfactual logic

If these logical assumptions governing counterfactuals are all true, then the gods simply cannot have the dispositions described. This runs against some pretty robust modal judgments. For example, it surely seems possible that, had the other gods not decided to thwart the man's progress, the first god could have decided to block the man if he reaches the $\frac{1}{2}$-mile mark and secured the counterfactual $P_{\geq 2} \mathrel{\Box\!\!\rightarrow} \neg P_1$; similar remarks apply to the other gods. But if each god individually could have the relevant disposition, why couldn't they together? After all, the gods' intentions appear to be independent of one another.

Drawing a parallel moral in the first puzzle seems to undermine the possibility of a rational being. For granting the counterfactual consistency of facing Yablo's Button on every day since eternity, it follows that Chocolate Preference is false for some day n: i.e. for some n, it's not the case that if you were in the man's position and the button had never been pressed before, you would press the button to receive the chocolate. Now, of course, it's clearly not the case that to be rational one must do the rational thing under every counterfactual supposition. For example, even a rational person would make poor choices if they were hit sufficiently hard on the head. But the counterfactuals considered above do not seem to have this flavor.

It is thus natural to revisit the logic of counterfactuals in light of these judgments. In doing so I also want to bring into the discussion some wider considerations involving conditionals that bear on their logic. Thus we will take into account the logic of indicative conditionals, the relationship between probabilities and conditionals, and other factors, to determine whether there might be independent reasons to adopt or reject particular logical principles governing counterfactuals. In the following subsections I examine three particular logical principles: (i) Infinite Conjunction, (ii) Substitution and (iii) Disjunction. We will also examine various accounts of the counterfactual that illustrate them: the similarity semantics of Lewis, Stalnaker and others, and the truth-maker semantics of Fine ((Fine, 2012b), (Fine)), and my own account in section 4. N.B.: I use the term *similarity semantics* to cover a wide range of theories in which the truth conditions of a counterfactual are stated in terms of an ordering of worlds — an ordering which may or may not bear much resemblance to the ordinary notion of 'similarity'; see the precise definition in the next subsection.[18]

3.1 Infinite Conjunction

As we saw in section 1, it is possible to validate Chocolate Preference and Zap Avoidance using a Lewisian semantics. Lewis's semantics famously eschews what is sometimes called the limit assumption. This assumption is often presented as the thesis that similarity to a given world is a well-founded relation: for any set of worlds, A, and world w there is always an A-world maximally similar to w. As a principle about similarity, it sounds implausible, for one can easily imagine an infinite sequence of worlds $w_1, w_2, ...$ each successively more similar to actuality than the previous, and so $\{w_1, w_2, ...\}$ has no $\prec_w$ minimal elements. It is often seen to be an advantage of Lewis's semantics that he doesn't make this posit. But taken at face value, as a principle about similarity, it implies nothing about counterfactuals unless we take the similarity analysis of counterfactuals for granted. A better strategy is to isolate a principle that corresponds (granting the similarity analysis) to the limit assumption, and evaluate that directly. Fine offers the following infinitary principle:[19]

Infinite Conjunction $A \Box\!\!\rightarrow B_1, A \Box\!\!\rightarrow B_2, ... \vdash A \Box\!\!\rightarrow \bigwedge_n B_n$

Roughly, any conjunction of things that would have been the case if A, would also have been the case if A.

Before continuing to investigate the role of this principle in our paradoxes, let me digress a little into a lesser known corner of counterfactual logic that will prove useful to our discussion.

Firstly, Infinite Conjunction does indeed correspond to the limit assumption, *provided we assume the similarity semantics.* More specifically, provided we assume a very general similarity semantics, which subsumes many of the standard proposals as special cases.[20] According to this semantics, a preorder on worlds, $\preceq_w$, is associated with each world w (often, but not always, glossed as ordering worlds by how 'similar' they are to w). The truth conditions for counterfactuals are then specified as follows:

[18] Lewis's notion, for instance, is highly theoretical, and in Stalnaker's later work the connection with ordinary similarity is even more distant. However, in virtue of sharing this same abstract analysis in terms of an ordering, these analyses share certain common logical principles which will be the focus of this paper. I have defined a similarity theory in a way that is neutral about whether $\leq_w$ is a total order preorder or not. For instance, (Pollock, 1976), (Veltman, 1976), (Kratzer, 1977) and (Lewis, 1981) treat the ordering as merely partial — see (Swanson, 2011) for a helpful overview.

[19] This principle is stated in a language which contains the ordinary truth functional connectives, the counterfactual condition $\Box\!\!\rightarrow$, closed under the usual formation rules and a further rule: if $A_1, A_2, A_3...$ is a countable sequence of sentences of the language, so is the countably infinite conjunction, $\bigwedge_n A_n$.

[20] Including (Lewis, 1973), (Lewis, 1981), (Stalnaker, 1968), (Kratzer, 1977), (Veltman, 1976) and (Pollock, 1976) among others.

Similarity Semantics $A \mathrel{\Box\!\!\rightarrow} B$ is true at w iff for every A-world x there is some A-world $y \preceq_w x$ such that every A-world $z \preceq_w y$ is a B-world

Given suitable (and routine) definitions of a frame, validity, and so forth, the validity of Infinite Conjunction corresponds exactly to the claim that $\preceq_w$ is well-founded for each w.[21]

Secondly, note that, just as Infinite Conjunction concerns countable conjunctions, there are apparent strengthenings involving longer conjunctions that may be formulated in a suitable larger infinitary language:

Infinite Conjunction$_\kappa$ $\{A \mathrel{\Box\!\!\rightarrow} B_\alpha \mid \alpha < \kappa\} \vdash A \mathrel{\Box\!\!\rightarrow} \bigwedge_{\alpha<\kappa} B_\alpha$

If we restrict our attention to the similarity semantics, nothing is gained by including the uncountable versions of this principle: if a similarity model validates the countable Infinite Conjunction principle, $\preceq_w$ must be well-founded for each w, and so that model must also validate Infinite Conjunction$_\kappa$ for any infinite κ. That said, there is a good sense in which the uncountable variants of Infinite Conjunction are stronger than Infinite Conjunction: we just have to move beyond the similarity semantics to see that (see the discussion of filter and ultrafilter semantics to come).

Fine proceeds to argue, convincingly in my view, that Infinite Conjunction is valid. To begin with, the finitary version of it is a part of almost every logic of conditionals on the market, and rightly so: given that I would have gotten wet if it had rained, and that I would have gotten cold if it had rained, then it just seems to follow, as a matter of logic, that I would have gotten wet and cold if it had rained. But Fine notes, citing (Pollock, 1976), that it is hard to motivate the finitary version without also motivating the infinitary version: 'what makes the finitary rule plausible is the more general principle that the logical consequences of the counterfactual consequences of a counterfactual supposition should also be counterfactual consequences of the supposition. But if this is the justification of the finitary rule, then it serves equally well to justify the infinitary rule.' (Fine, 2012a), p39.

A more theoretical argument for Infinite Conjunction can be given, founded on *prima facie* plausible principles connecting the probability of conditionals to conditional probabilities. For indicative conditionals this often takes the form of a constraint saying that a rational credence in a conditional, if A then B, should be one's conditional credence in B on A: this thesis is sometimes called 'Stalnaker's thesis'.[22] Of course, probabilistic constraints on conditionals like these are subject to well-known limitative results.[23] But the idea is a powerful one,

[21] If $\preceq_w$ is not well-founded at w, there is a countable infinite descending chain, $...x_2 \prec x_1 \prec x_0$. Letting A_n be true at $\{x_n, x_{n+1}, x_{n+2}, ...\}$, then $A_0 \mathrel{\Box\!\!\rightarrow} A_n$ is true at w for each n, but $A_0 \mathrel{\Box\!\!\rightarrow} \bigwedge_n A_n$ is false at w. Conversely, if $\preceq_w$ is well-founded, let X be the set of $\preceq_w$ minimal A-worlds. $A \mathrel{\Box\!\!\rightarrow} B_n$ is true iff X is a subset of the B_n worlds. If this is true for every n, then X is a subset of the $\bigwedge_n B_n$ worlds.

[22] See (Stalnaker, 1970).

[23] Although I think their significance has in general been overstated as I have argued elsewhere (cf. (Bacon, 2015)).

and moreover, there are many restrictions of the connection which are perfectly consistent, and suffice to draw conclusions about the logic of conditionals.[24]

Given this and the assumption of countable additivity, Infinite Conjunction is a probabilistically valid rule in the sense that if the premises are all certain, then so is the conclusion, for any rational probability function.[25] (It is interesting to note that this argument does not extend to Infinite Conjunction$_\kappa$ for uncountable κ, unless we made the far less plausible assumption that rational credences are κ-additive, an assumption which entails that all probability originates from at most a countable set of worlds.)

Indicative and counterfactual logic do not necessarily have to coincide, but coincidence is a worthy aspiration, and if the indicative version of Infinite Conjunction was valid that would certainly be suggestive. Versions of the probability conditional link also exist between counterfactuals and chances, so there is a more direct argument for Infinite Conjunction along similar lines.[26]

So we have some *prima facie* good reasons to accept Infinite Conjunction. Infinite Conjunction also screens off certain paradoxical situations, for without Infinite Conjunction we are open to the possibility of *counterfactual pathologies*: cases where a collection of jointly inconsistent propositions are all true if A had been the case, even when A itself is consistent.[27] Indeed, (Herzberger, 1979) points out that Lewis's particular way of invalidating Infinite Conjunction is susceptible to exactly these sorts of worries. Suppose that, as it happens I am shorter than 6 foot tall, and that, other things being equal, a world in which I am closer to my actual height is closer to actuality than a world in which I am not. Then Lewis's analysis seems to predict that the following would all have been the case had I been more than 6 foot tall. (i) I would have certainly been taller than 6 foot. (ii) I would have been less than 6 and a $\frac{1}{2}$ feet tall. (iii) I would have been less than 6 and a $\frac{1}{4}$ feet tall, more generally, I would have been less than 6 and a $\frac{1}{2}^n$ feet tall for any n. So the collection of propositions that would have been true had I been more than 6 feet tall is ω-inconsistent in the sense that, while any finite subset of the propositions is consistent, the propositions cannot all be true together.

The invalidity of Infinite Conjunction on its own does not *guarantee* that there will be counterfactual pathologies. In the presence of another principle of counterfactual logic, however, it does:

Conditional Excluded Middle $(A \mathrel{\Box\!\!\rightarrow} B) \vee (A \mathrel{\Box\!\!\rightarrow} \neg B)$

The debate over Conditional Excluded Middle (henceforth CEM) is intricate, and I will not attempt to reproduce or summarize it here.[28] I will note, however,

[24] I have in mind, particularly, van Fraassen (van Fraassen, 1976) and Bacon (Bacon, 2015).

[25] For if $Cr(A \rightarrow B_n) = Cr(B_n \mid A) = 1$ for every n, then applying countable additivity to $Cr(\cdot \mid A)$ we may conclude that $Cr(\bigwedge_n B_n \mid A) = Cr(A \rightarrow \bigwedge_n B_n) = 1$.

[26] See also (Skyrms, 1980) and (Moss, 2013) for a discussion of some related thoughts, and necessary qualifications of the idea.

[27] The consistency I have in mind is simple counterfactual consistency: A does not counterfactually imply absurdity $(\neg(A \mathrel{\Box\!\!\rightarrow} \neg A))$.

[28] (Lewis, 1973), (Stalnaker, 1981), (Williams, 2010), (Mandelkern, forthcoming).

that, like Infinite Conjunction, it can also be given a probabilistic underpinning, for given the aforementioned connection to probabilities its probability must be 1 when A has non-zero probability.[29]

Suppose, then, that Infinite Conjunction is invalid. So we should expect to find A and $B_1, B_2, ...$ such that $A \mathrel{\Box\!\!\rightarrow} B_n$ is true for each n, and $A \mathrel{\Box\!\!\rightarrow} \bigwedge_n B_n$ is false. By CEM it follows that $A \mathrel{\Box\!\!\rightarrow} \neg \bigwedge_n B_n$, since CEM ensures that one of $A \mathrel{\Box\!\!\rightarrow} \bigwedge_n B_n$ or $A \mathrel{\Box\!\!\rightarrow} \neg \bigwedge_n B_n$ is true. But then A counterfactually implies an ω-inconsistent collection of propositions: B_n for each n, and $\neg \bigwedge_n B_n$.

Conditional Excluded Middle is germane to our present discussion because, like Infinite Conjunction, it also secures the limit assumption relative to the possible world semantics sketched above. Indeed, the validity of CEM corresponds to $\preceq_w$ not merely being a well-founded preorder, but a *well-order*: worlds will additionally be totally ordered by $\preceq_w$. CEM, unlike Infinite Conjunction, is a finitary principle.

Let me now make a brief digression about the relation between CEM and Infinite Conjunction. It is sometimes supposed, for presumably these reasons, that there is a difficulty devising a semantics that validates CEM without Infinite Conjunction: the above, for instance, demonstrates that this is impossible in a similarity semantics — for in that semantics CEM semantically implies Infinite Conjunction. Eric Swanson (Swanson, 2012), for example, proposes a (loosely speaking) supervaluationist account which validates CEM but allows for ω-inconsistent collections of sentences to be individually true in the scope of a counterfactual supposition. But there is a problem with his account: he achieves this at the expense of permitting actual ω-inconsistencies. An ω-inconsistent set of sentences can be true simpliciter, not merely true within the scope of a counterfactual supposition. For an infinite conjunction of supertruths can be superfalse in his semantics, and conversely an infinite disjunction of superfalsehoods supertrue.[30] The semantics also fails to validate logical rules not involving the counterfactual conditionals, such as infinitary conjunction introduction, which, as we have effectively just established, does not preserve supertruth. Discussions of these features do not appear in (Swanson, 2012), but they seem to me to diminish the interest in the semantics significantly.

A less revisionary approach would be to modify the familiar selection function semantics for counterfactuals. The orthodox selection function semantics for CEM consists of a function, $f : P(W) \times W \to P(W)$, that maps a set A and

[29] We must assume also that the probability of $(A \mathrel{\Box\!\!\rightarrow} B) \wedge (A \mathrel{\Box\!\!\rightarrow} \neg B)$ is 0 when A has positive probability (this seems plausible, since given the finitary version of Infinite Conjunction, and its converse, it is equivalent to $A \mathrel{\Box\!\!\rightarrow} (B \wedge \neg B)$, and by the probabilitistic connection, has chance $Ch(B \wedge \neg B \mid A) = 0$. The reason CEM must have chance 1 is because $Ch((A \mathrel{\Box\!\!\rightarrow} B) \vee (A \mathrel{\Box\!\!\rightarrow} \neg B)) = Ch(B \mid A) + Ch(\neg B \mid A) - Ch((A \mathrel{\Box\!\!\rightarrow} B) \wedge (A \mathrel{\Box\!\!\rightarrow} \neg B)) = x + (1-x) + 0 = 1$.

[30] For Swanson, precisifications are ordered by how 'good' they are (for instance, how well they approximate usage). A sentence is supertrue if, for some precisification, it is true at every precisification at least as good as it. Thus, $A_1, A_2, A_3 ...$ etc may all be supertrue, even when there are no precisifications at which they are all true. This would happen, for example, if the precisifications were enumerable in sequence strictly increasing in order of goodness, $v_1, v_2, v_3, ...$ (i.e. v_n is better than v_m exactly when $n > m$) and when A_n is true at exactly $v_n, v_{n+1}, v_{n+2} ....$

a world w to the singleton of the world that would have been the case (relative to w) had A been the case (and maps it to the emptyset if A is counterfactually inconsistent at w). Alternatively, we might instead theorize in terms of a function $f : P(W) \times W \to P(P(W))$. $f(A, w)$ can be thought of as the set of propositions that would have been true if A had been true, and we can specify the semantic value of a counterfactual, $[\![A \mathrel{\Box\!\!\rightarrow} B]\!]$, as the set of worlds w where $[\![B]\!] \in f([\![A]\!], w)$. CEM corresponds to the constraint that $f(A, w)$ contains any set of worlds or its complement. If we additionally want to validate Weakening and Finite Conjunction, it must also be closed under supersets and finite intersections: i.e. it must be an ultrafilter.[31] Infinite Conjunction, by contrast, corresponds to the condition that $f(A, w)$ is closed under countable intersections (and, more generally, Infinite Conjunction$_\kappa$ to closure under κ sized intersections). It is a well-known fact that, given the axiom of choice and infinite W, there are non-degenerate ultrafilters over W that are not closed under countable intersections. We can leverage this to show that Conditional Excluded Middle does not entail Infinite Conjunction, even in an infinitary logic with the usual rules, including infinite conjunction introduction (a rule that Swanson's system lacks). Thus, while the combination of CEM without Infinite Conjunction is not particularly attractive — it commits one to counterfactual pathologies — our deviant semantics shows that it is not formally inconsistent. (Moreover, unlike Swanson's semantics, this establishes the consistency of this package with all the classical inference rules, including their infinitary analogues for infinite conjunction.)

It's worth noting that we can also show that Infinite Conjunction$_\omega$ does not entail Infinite Conjunction$_\kappa$ for $\kappa > \omega$ via similar means, even in the presence of CEM, although one has to go beyond the standard axioms of set theory to do so. (What one needs is an ultrafilter that is closed under countable intersections but not arbitrary intersections. This can be secured by assuming the existence of a measurable cardinal.)

We may summarize our findings thus. Infinite Conjunction is a plausible principle; the limit assumption — that every set contains a world maximally similar to actuality — is not. Yet, given the similarity semantics, they are equivalent. Many commentators have employed these observations narrowly, to adjudicate between Lewis and Stalnaker, usually in favour of Lewis. But implicit in the above is really a problem with the similarity semantics for counterfactuals. For once we liberate the counterfactual facts from the similarity constraint — that a counterfactual is true if its consequent is true in some antecedent world, and any antecedent world more similar to actuality than it — the tension between the plausibility of Infinite Conjunction and the implausibility of the limit assumption evaporates. We have also explored the connections between CEM and Infinite Conjunction. It is consistent, without Infinite Conjunction, for there to be counterfactual pathologies, and if we additionally grant CEM all failures of Infinite Conjunction are pathological.

[31] In what follows we count the degenerate ultrafilter $P(W)$, containing every proposition, as an ultrafilter. (It will play the role of the impossible world.)

Infinite Conjunction was one of the principles used in our derivation of the two paradoxes of infinity we opened with. Denying Infinite Conjunction, perhaps along Lewisian lines, thus might offer us a way to resolve those paradoxes. So what must the Infinite Conjunction denier say about our two paradoxes? Even though the denial of Infinite Conjunction does not, on its own, commit one to counterfactual pathologies, I will now show that one cannot resolve either of the two paradoxes without accepting a counterfactual pathology (unless one gives up some other finitary principles used in our proof, making the rejection of Infinite Conjunction unnecessary). Firstly, note that Infinite Conjunction is used once in both derivations. I'll focus on Benardete's paradox first. If the Infinite Conjunction denier accepts the argument until the first and only use of Infinite Conjunction, they will accept $\bigvee_n P_{\geq n} \mathrel{\Box\!\!\rightarrow} \neg P_{\geq k}$, for each k, and retain consistency with Anti-Zenoism by resisting this inference to $\bigvee_n P_{\geq n} \mathrel{\Box\!\!\rightarrow} \bigwedge_n \neg P_{\geq n}$. Paraphrase $P_{\geq n}$ as 'the man traveled continuously as far as the $\frac{1}{2}^n$-mile marker'.[32] In this informal mode, this amounts to accepting all of the following:

1. If the man had made it any distance past A, then he wouldn't have gotten past the $\frac{1}{2}$-mile marker.

2. If the man had made it any distance past A, then he wouldn't have gotten past the $\frac{1}{4}$-mile marker.

$\vdots$

n. If the man had made it any distance past A, then he wouldn't have gotten past the $\frac{1}{2}^n$-mile marker.

$\vdots$

Thus, we see that the set of propositions that would have been the case had the man made it any distance past A is ω-inconsistent. He would have made it past A, but he wouldn't have made it $\frac{1}{2}^n$-miles past A for any n.

A similar moral must be drawn about Yablo's Button. If you accept the reasoning up until the step where Infinite Conjunction is applied, we have:

1. If I had been in the man's position and declined to press the button on some day and all days preceding it, then I would have pressed the button on the last day (day 0).

2. If I had been in the man's position and declined to press the button on some day and all days preceding it, then I would have pressed the button on the penultimate day (day -1).

[32] The continuity parenthetical is there to emphasize that he got to the $\frac{1}{2}^n$-mile marker by way of the $\frac{1}{2}^k$-mile markers for $k > n$.

3. If I had been in the man's position and declined to press the button on some day and all days preceding it, then I would have pressed the button on the day before that (day -2).

4. $\vdots$

In this case, the individual counterfactuals themselves sound completely implausible, indicating that an error must have occurred earlier in the proof, before one has applied Infinite Conjunction. To make matters worse, they also constitute a counterfactual pathology, since the propositions that I pressed the button on days $0, -1, -2, ...$, and that I declined to press the button until some some day are jointly inconsistent (even if any finite subset is consistent). So I think a deeper analysis of reasoning reveals one is already committed to some sort of pathology before one has even applied Infinite Conjunction.

Things get even weirder if we bring Conditional Excluded Middle into the mix. For example, as above we must conclude, for every n, that, had the man made it past A, the man would have been stopped before the $\frac{1}{2}^n$-mile mark. But despite this, one can show that there also must be a point where the man stopped, had he made it past A. That is, there is an n such that he made it to the $\frac{1}{2}^{n+1}$-mile mark, but not the $\frac{1}{2}^n$-mile mark, or he made it all the way to B. The existence of this point is captured by the disjunction $P_{\geq 0} \vee \bigvee_{n>0}(P_{\geq n+1} \wedge \neg P_{\geq n})$. So formalized, the existence of a stopping point is a logical consequence of $\bigvee_n P_{\geq n}$, so it is counterfactually implied by it. We can begin to ask embarrassing questions about this point, such as whether it would have been an odd or even n. For, given CEM, even though we cannot say that the stopping point would have been $\frac{1}{2}^n$ for any particular n, we can capture the claim that the stopping point would have been even with the counterfactual

$$\bigvee_n P_{\geq n} \mathrel{\Box\!\!\rightarrow} P_{\geq 0} \vee \bigvee_{n \in Odds}(P_{\geq n+1} \wedge \neg P_{\geq n})$$

and the claim that it would have been odd with

$$\bigvee_n P_{\geq n} \mathrel{\Box\!\!\rightarrow} \bigvee_{n \in Evens}(P_{\geq n+1} \wedge \neg P_{\geq n}).$$

But Conditional Excluded Middle and the principle that you can substitute logical equivalents in the consequent of a counterfactual guarantees that one of these counterfactuals must be true. The same weirdness is apparent in the first puzzle. If I had declined to press the button up until some day or other, there would have been a first day I pressed the button. It wouldn't have been the 0th day, or the -1th, or the -2th, and so on. But it either would have been even, or it would have been odd. (Putting it heuristically, in the first case it is as though the man would travel an infinitesimal distance past A, and in the second as though the first pressing of the button would be a non-standard number. Thought of this way, they will have many of the sorts of first-order definable properties that non-standard reals and natural numbers share with their standard cousins.[33])

[33] The use of non-principal ultrafilters in constructing these non-standard mathematical objects, and in our way of modelling such counterfactuals, is presumably not a coincidence.

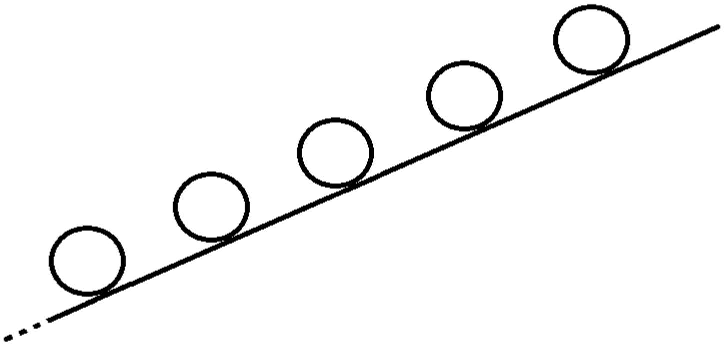


Figure 1: The infinite slope.

3.2 Substitution

Fine's own logic of counterfactuals is motivated by an infinite puzzle with a slightly different flavor. We are invited to imagine an infinite slope, with an infinity of balls precariously balanced as in figure 1. We assume that ball n is numbered so that if it toppled it would knock ball $n+1$. Fine then presents an argument, formally very similar to the two arguments given above, that standard approaches to counterfactual logic cannot accommodate natural verdicts about what would happen if a given ball were to topple.[34] Formalizing 'the nth ball toppled' as B_n, Fine's assumptions may be stated:

Positive Effect $B_n \Box\!\!\rightarrow B_{n+1}$

Negative Effect $B_{n+1} \Box\!\!\rightarrow \neg B_n$

Consistency $\neg(\bigvee_n B_n \Box\!\!\rightarrow \bot)$

The most contentious of these assumptions is surely Negative Effect: it captures the idea that, if a given ball, n, were to topple, it would be the first ball to topple. In other words, it wouldn't be because some ball further up the slope toppled and knocked it. We will return to this assumption in section 3.4.

While these three counterfactuals are inconsistent given the counterfactual assumptions we have outlined above, they are jointly consistent in Fine's logic. Given the similarity between the reasoning in both cases it is worth seeing how it might be applied to the two puzzles we have looked at above. We see that, in the present derivation, Fine's 2012 semantics invalidates two of the principles used: Disjunction and Substitution. More recent versions of Fine's truth-maker semantics, however, validate Disjunction ((Fine, 2018a), (Fine, 2018b), (Fine))

[34] See the discussion in section 1 for the exact points of dissimilarity.

and thus only invalidate Substitution. And, even his 2012 semantics validates the following weakening of Disjunction:

Weak Disjunction $A \mathrel{\Box\!\!\rightarrow} C, B \mathrel{\Box\!\!\rightarrow} C \vdash A \vee B \mathrel{\Box\!\!\rightarrow} C$ provided A and B are logically incompatible.[35]

Fine has a variant way of deriving a paradox to the one in section 1 and 2 that only appeals to this weakening. So it's natural to look to failures of Substitution to explain these paradoxes; and indeed, this is the exact diagnosis Fine offers in the case of the balls.

For concreteness, I will focus on the account in (Fine, 2012a) and (Fine, 2012b), since it is explicitly an account of counterfactual conditionals, and is developed with similar paradoxes of infinity in mind. Many of my remarks will extend to other possible accounts of counterfactuals within this framework, however. In (Fine, 2012a) Fine claims that Substitution is to blame for the inconsistency derivable from Positive Effect, Negative Effect and Consistency. To substantiate this claim, Fine offers a model theory which invalidates the inference he rejects, Substitution, whilst validating the other principles required in his derivation. But a model *theory* only settles which inferences and principles are valid, and leaves unsettled many questions of truth. Unlike the Lewis-Stalnaker semantics, it is not a mechanical matter to apply the technology to particular cases, and in this case Fine does not offer even a heuristic model of the slope, from which we could simply read off answers to questions like these.[36] Applying the model theory to our two paradoxes of infinity is similarly not a mechanical matter, but we can make some principled modelling decisions.

In Fine's semantics, propositions are verified or falsified by partial states of affairs, or just *states*, which unlike worlds, do not settle the truth of all propositions, and are consequently ordered by how complete or incomplete they are. A world is thus just a special kind of state: a maximal one. Verification is exact, in the sense that a proposition need not be verified by states that are more complete than states that verify it. We say that a proposition is *inexactly* verified by a state if it is exactly verified by some less complete state. The clause for the counterfactual in (Fine, 2012b) is given in term of a ternary accessibility relation, $t \rightarrow_w u$, understood informally as saying that u is the state that results from w by making the change t.

> $A \mathrel{\Box\!\!\rightarrow} B$ is true at a world w if and only if, for every state t that verifies A, if $t \rightarrow_w u$, B is inexactly verified by u.

If you blur your eyes somewhat and focus attention on the order of the quantifiers, this looks a bit like the clause for a strict conditional.[37] The high-level

[35] His semantics also validates another weaker rule, that allows one to infer counterfactuals with compatible disjunctive antecedents: $A \mathrel{\Box\!\!\rightarrow} C, B \mathrel{\Box\!\!\rightarrow} C, A \wedge B \mathrel{\Box\!\!\rightarrow} C \vdash A \vee B \mathrel{\Box\!\!\rightarrow} C$.

[36] The primary difference, it seems to me, is that the theoretically central notion of imposing a change to a world by a state to produce another state is not antecedently intelligible, and is not spelled out to the degree that, for example, the notion of closeness is in Lewis's philosophy.

[37] Or better, the clause for a conditional in relevance logic.

moral to take away from this is that being counterfactually implied by an assumption is a relatively demanding condition. Compare this, for instance, to the proponent of Conditional Excluded Middle, who will say, of an unflipped coin, that it either would have landed heads if it had been flipped, or that it would have landed tails. In order for the former (latter) counterfactual to be true on Fine's semantics, every state that verifies the coin being flipped would, if we imposed that change to the actual world, result in a state that inexactly verifies the coin landing heads (tails).

Let's start with Benardete's paradox. What would have happened if the man had made it past A? Would he have made it to the $\frac{1}{2}$-mile mark? To the $\frac{1}{4}$-mile mark? Or is there a false presupposition to this question — like asking, of an unflipped coin, which way it would have landed if it had been flipped? According to the deniers of Conditional Excluded Middle, questions like these have the false presupposition that the coin either would have landed heads or would have landed tails.

I think the thought that goes along most naturally with Fine's semantics is that it's simply not the case that, had the man gotten past A, he would have gotten to the $\frac{1}{2}$-mile mark. Similarly for the $\frac{1}{4}$-mile mark, $\frac{1}{8}$-mile mark, and so on. But it's *also* false that, had the man gotten past A, he *wouldn't* have gotten to the $\frac{1}{2}$-mile mark (violating CEM). And similarly for the $\frac{1}{4}$-mile mark, the $\frac{1}{8}$-mile mark, and so on. This resolves the paradox, since we do not have the premises required to apply Infinite Conjunction — that the man wouldn't have gotten to the $\frac{1}{2}$-mile mark, that he wouldn't have gotten to the $\frac{1}{4}$-mile mark, and so on — and contradict Anti-Zenoism.

Let's focus on a different question: what would have happened if the man had either gotten as far as the $\frac{1}{2}$-mile mark, or the $\frac{1}{4}$-mile mark? To start us off, ask what would have happened if the man had traveled to the $\frac{1}{4}$-mile mark. Then the first god would have built his wall, and he wouldn't have made it past the $\frac{1}{2}$-mile mark. This is just what Divine Dispositions tells us:

$$P_{\geq 2} \mathrel{\Box\!\!\rightarrow} \neg P_1$$

On the other hand, if the man had gotten to the $\frac{1}{4}$-mile mark *or* the $\frac{1}{2}$-mile mark, we are less inclined to say that he wouldn't have gotten to the $\frac{1}{2}$-mile mark. Taking this judgment at face value, we might therefore wish to deny the following counterfactual:

$$P_{\geq 1} \vee P_{\geq 2} \mathrel{\Box\!\!\rightarrow} \neg P_1$$

But this is exactly the move from step 1 to step 2 in our argument. The step was licensed by Substitution, because the antecedents of these two conditionals are logically equivalent. The equivalence may be proven formally, given our definition of 'reaching the $\frac{1}{2}^n$-mile mark', but informally, it's clear that the claim that the man reached the $\frac{1}{4}$-mile mark entails that he either reached the $\frac{1}{4}$-mile mark or the $\frac{1}{2}$-mile mark. And conversely, suppose he reached the $\frac{1}{4}$-mile mark or the $\frac{1}{2}$-mile mark; then either way he must have reached the $\frac{1}{4}$-mile mark, and the converse entailment holds as well.

A similar diagnosis of Yablo's Button may also be made. Had the display read 'Chocolate' on day 0 (i.e. the button has not yet been pressed), then you would press the button on day 0. Why? Because you would get a free chocolate. But, it seems, it is not the case that, had the display read 'Chocolate' on day 0 *or* day -1, you would press the button on day 0. Why? Because if it read 'Chocolate' on day -1 and you had pressed the button on day -1, you would get a zap if you pressed it on day 0.

Yet, as before, the claim that the display reads 'Chocolate' on day -1 is equivalent, given the specification of the display, to reading 'Chocolate' on day -1 or day 0. (Clearly if it read 'Chocolate' on day -1, it did so on either day -1 or day 0. Conversely, if it read 'Chocolate' on day -1 or day 0, the button must not have been pressed until at least day -1, and so it reads 'Chocolate' on day -1.) And given that reading 'Chocolate' on day 0 is really just a convenient substitute for the claim that the button wasn't pressed on any day prior to day 0, $D_{\leq 1}$, they are logically equivalence in the stricter sense spelled out in section 1.

3.3 Substitution and Disjunctive Antecedents

The above judgments are in fact instances of a wider pattern of judgments concerning counterfactuals with disjunctive antecedents. Consider the following two counterfactuals:

✓ If I went to the beach, I would have a good time.

✗ If I went to the beach or went to the beach and got attacked by a shark, I would have a good time.

The first sounds good, but the second does not. Yet the antecedents, of the form A and $A \vee (A \wedge B)$ respectively, are logically equivalent. Indeed, this is the exact same form as the judgments we made above in relation to Yablo's Button and Benardete's paradox (recalling that according to our conventions $P_{\geq n}$ is short for a conjunction).

However, it is contentious whether these sorts of judgments are in good standing: the semantics and pragmatics of counterfactuals with disjunctive antecedents is extremely delicate. It will be instructive, then, to take a little digression though this vexed issue in the philosophy of conditionals. Disjunctive antecedents will also be at the heart of my own account.

Prior to Fine there were roughly two lines of thinking on this issue, exemplified respectively by strict theories of the counterfactual conditional, such as defended by (von Fintel, 2001) and (Gillies, 2007), and the 'variably strict' similarity based theories of (Lewis, 1973) and (Stalnaker, 1968).

The strict theory is formulated in terms of a contextually supplied accessibility relation between worlds, and states that a counterfactual $A \mathrel{\Box\!\!\rightarrow} B$ is true when every accessible A-world is a B-world. The strict theory gets to maintain a highly attractive theory of disjunctive antecedents, in which a counterfactual with a disjunctive antecedent $(A \vee B) \mathrel{\Box\!\!\rightarrow} C$ is just equivalent to the conjunction of two counterfactuals with non-disjunctive antecedents $(A \boxright C) \wedge (B \boxright C)$. I shall call this the simple account of disjunctive antecedents:

The Simple Account $(A \vee B) \boxright C \dashv\vdash (A \boxright C) \wedge (B \boxright C)$

The right-to-left direction of this inference we have called Disjunction, and Fine calls the left-to-right Simplification.

The simple account is not only simple but accords with our intuitions in many cases.[38] For example, if we know that the party would have been a success if Alice or Bob had come, then we can infer that it would have been a success if Alice had come, and that it would have been a success if Bob had come. The converse inference also seems to be valid: if the party would have been a success if Alice had come, and it would have been a success if Bob had come, then it would have been a success if either of them had come. We can also appeal to the inference to explain why the ✗ed counterfactual appears false: it entails the false counterfactual *if I had gone to the beach and gotten attacked by a shark, I would have had a good time.*

The strict theory, however, validates the following highly controversial principle:

Antecedent Strengthening $A \boxright C \vdash A \wedge B \boxright C$

Antecedent Strengthening has many apparent counterexamples, such as:[39]

✓ If I went to the beach, I would have a good time.

✗ If I went to the beach and got attacked by a shark, I would have a good time.

Defenders of the strict counterfactual have, of course, made various contextualist moves to explain away these counterexamples. But they inevitably involve denying the simplest explanation for its felt invalidity: that it is in fact invalid.[40]

This brings us to the other main option: philosophers subscribing to the variably strict theory exemplified by Lewis and Stalnaker, have taken these counterexamples at face value, and provided accounts in which Antecedent Strengthening is semantically invalid. These theories have a straightforward explanation of the above judgments concerning Antecedent Strengthening, but there is a cost: they do not validate the simple account of disjunctive antecedents. $(A \vee B) \boxright C$ does not entail $(A \boxright C) \wedge (B \boxright C)$, despite the appearance to the contrary.

[38] But not all: see the discussion in Nute and Cross §1.8 (Cross and Nute, 1997).

[39] This is not the end of the story: see (von Fintel, 2001) and (Gillies, 2007). But see (Moss, 2012), for some replies.

[40] Some contextualist and dynamic semanticists have introduced deviant notions of 'validity' in which antecedent strengthening is not valid. Roughly $A_1 ... A_n \vdash B$ means that if you were to say $A_1 ... A_n$ in that order, you would be in a position to say B in the resulting context. But very little is valid in this deviant sense — for instance, not even $A, B \vdash A$ is valid if saying B changes the context — and so it has little to do with the notion logicians call by the same name.

Indeed this dilemma is no accident: given the possible worlds framework common to both theories, logically equivalent sentences can be substituted *salve veritate*. One can then show that the simple account of disjunctive antecedents entails Antecedent Strengthening. Suppose that the simple account is correct, and that $A \mathrel{\Box\!\!\rightarrow} C$. Then substituting A for the logically equivalent $A \vee (A \wedge B)$ we may conclude $(A \vee (A \wedge B)) \mathrel{\Box\!\!\rightarrow} C$, and thus $(A \wedge B) \mathrel{\Box\!\!\rightarrow} C$ by the simple account.

Having seen the necessary tradeoffs between Fine's two competitors, we are now in a position to appreciate a powerful reason to adopt a hyperintensional theory of counterfactuals (such as a truth-maker account): one can simultaneously accept the simple account of disjunctive antecedents, whilst rejecting the contentious principle of Antecedent Strengthening.

Fine's truth-maker semantics can indeed validate the simple account whilst invalidating Antecedent Strengthening, but I would like to frame my discussion at a broader question: to what extent can hyperintensional theories of conditionals *in general* accord with our naïve judgments, regarding the validity of the Simple inferences, and the invalidity of Antecedent Strengthening. I will present a limitative result concerning what is possible in this direction, that will apply to truth-maker theories of conditionals and other hyperintensional theories of conditionals alike. If no plausible hyperintensional semantics can vindicate our naïve judgments involving counterfactuals with disjunctive antecedents, then I think we should be much more skeptical about the judgments involving disjunctive antecedents we began with — and thus, much more skeptical of the judgments being wielded to block our paradoxical reasoning.

I will not focus on Antecedent Strengthening, but on the following equally unwelcome variant:

Weak Antecedent Strengthening $A \mathrel{\Box\!\!\rightarrow} C, B \mathrel{\Box\!\!\rightarrow} C \vdash (A \wedge B) \mathrel{\Box\!\!\rightarrow} C$

Unwelcome, since we should affirm the first and second counterfactual, but not the third.

- ✓ If I were to drink this hot beverage, I'd have a pleasant time.
- ✓ If I were to ride this roller-coaster, I'd have a pleasant time.
- ✗ If I were to drink this hot beverage and ride this roller-coaster, I'd have a pleasant time.

While a hyperintensional theory will not in general permit the substitution of arbitrary logical equivalents, some such substitutions are hard to deny: it would be implausible to think, for instance, that a counterfactual of the form $(A \vee B) \mathrel{\Box\!\!\rightarrow} C$ could by true while $(B \vee A) \mathrel{\Box\!\!\rightarrow} C$ is false. Indeed Fine's truth-maker semantics does allow the substitution of these two sorts of sentences, as they have exactly the same truth-makers. The following substitutions all look like could plausibly be made in the antecedent of a counterfactual.

Idempotence $(A \wedge A) \mathrel{\Box\!\!\rightarrow} C \dashv\vdash A \mathrel{\Box\!\!\rightarrow} C \dashv\vdash (A \vee A) \mathrel{\Box\!\!\rightarrow} C$

Commutivity $(A \vee B) \boxright C \dashv\vdash (B \vee A) \boxright C$

Distributivity $A \wedge (B \vee C) \boxright D \dashv\vdash ((A \wedge B) \vee (A \wedge C)) \boxright D$

Associativity $(A \vee B) \vee C \boxright D \dashv\vdash A \vee (B \vee C) \boxright D$

Indeed, in the truth-maker semantics of (Fine, 2012b), these limited forms of Substitution are valid: the antecedents of Idempotence, Commutativity, Distributivity and Associativity have the same truth-makers by that account. Moreover, the clause for the truth of a counterfactual is only sensitive to the truth-makers of the antecedent, so making any of the above substitutions in the antecedent of a true counterfactual (a counterfactual with at least one truth-maker) will result in a true counterfactual.[41]

At any rate, the intersubstitutivity of sentences equivalent according to this limited set of rules is independently plausible, regardless of whether Fine's semantics validate it. Moreover, while there are apparent counterexamples to the general principle Substitution, involving the substitution of A for $A \vee (A \wedge B)$ (as was implicit in the proof of Antecedent Strengthening from Simplification), there do not appear to be any similar counterexamples involving these more limited applications of Substitution.

The problematic result is this:

No-Go Theorem Every theory of conditionals that permits the four limited applications of Substitution listed and validates "The Simple Account", also contains Weak Antecedent Strengthening.

The derivation of Weak Antecedent Strengthening from these assumptions goes as follows:

1. $A \boxright C$ (assumption)
2. $B \boxright C$ (assumption)
3. $A \vee B \boxright C$ (Simple Theory)
4. $(A \vee B) \wedge (A \vee B) \boxright C$ (Idempotence)
5. $((A \vee B) \wedge A) \vee ((A \vee B) \wedge B) \boxright C$ (Distributivity)
6. $(A \wedge (A \vee B)) \vee (B \wedge (A \vee B)) \boxright C$ (Commutativity)
7. $((A \wedge A) \vee (A \wedge B)) \vee ((B \wedge A) \vee (B \wedge B)) \boxright C$ (Distributivity)
8. $(A \vee (A \wedge B)) \vee ((B \wedge A) \vee B) \boxright C$ (Idempotence)

[41] The case of Distributivity is a little more subtle because while both sides have the same truth-makers, they have different falsemakers. So while distributive equivalents can be substituted in the antecedent of a counterfactual, in accordance with Limited Substitution, they cannot always be substituted within a negated context. However, the coincidence in truth-makers is enough to validate the limited forms of Substitution, and that is all that is needed for the following discussion.

9. $(A \vee B) \vee ((A \wedge B) \vee (A \wedge B))) \mathrel{\Box\!\!\rightarrow} C$ (Associativity, Commutivity)

10. $(A \vee B) \vee (A \wedge B) \mathrel{\Box\!\!\rightarrow} C$ (Idempotence)

11. $A \wedge B \mathrel{\Box\!\!\rightarrow} C$ (Simple Theory)

The No-Go theorem then tells us what shape a hyperintensional theory of counterfactuals must have if it is to maintain the Simple Theory concerning disjunctive antecedents, while rejecting the likes of Antecedent Strengthening. Either Idempotence, Commutativity, Distributivity or Associativity must be relinquished.

We can now categorise different truth-maker approaches to counterfactuals by this result. Inspection of the semantics in (Fine, 2012b) reveals that all of Idempotence, Commutativity, Distributivity and Associativity hold. In that account the Simple Theory is valid only with a caveat: the right-to-left portion of the inference (i.e. Disjunction) is only valid when A and B are logically incompatible — the principle we have called Weak Disjunction. This differential treatment of Disjunction and Weak Disjunction seems a little unprincipled. (To lay my cards on the table, I am skeptical of both principles, and indeed think that, by rejecting them, we may provide a satisfying resolution to our paradoxes. But I think this makes the status of Disjunction on Fine's semantics even more urgent, for if one can indeed avoid paradox by relinquishing Disjunction whilst also keeping Substitution, this would undermine a central part of Fine's case for overturning orthodoxy.)

In more recent work Fine has advocated for a simpler truth-maker semantics in which the truth-makers for A and for $A \wedge A$ can be different. Indeed, Fine's present position is that Idempotence is to be rejected, and that the Simple Theory, and the other limited versions of Substitution hold.[42] However, it's worth emphasizing how radical this response is. It is one thing to propose a more fine-grained theory of propositions — one that distinguishes the proposition that A from the proposition that $A \wedge A$. In the grand scheme of things, this move is not that radical: structured theories of propositions, for instance, will distinguish propositions by the number of times a constituent like A occurs, as well as taking a myriad of other things into account, such as constituent order. So along that dimension, Fine's truth-maker theory is a lot less radical than the structured theory. But distinguishing the proposition A from $A \wedge A$ is not the same as denying the counterfactual inference of Idempotence: the inference from $A \mathrel{\Box\!\!\rightarrow} C$ to $(A \wedge A) \mathrel{\Box\!\!\rightarrow} C$ and conversely (which, of course, structured and other fine-grained theories of propositions have no special reason to reject). We have encountered putative counterexamples to inferences of the form $A \mathrel{\Box\!\!\rightarrow} C$ to $(A \vee B) \mathrel{\Box\!\!\rightarrow} C$, when A and $A \vee B$ are logically equivalent, but we do not have any similarly compelling case against Idempotence. Indeed, it seems eminently plausible.

There are independent reasons to adopt Substitution and attempt to explain

[42] In ((Fine, 2018a), (Fine, 2018b) and (Fine) a truth-maker semantics with these features is described.

the putative counterexamples using pragmatic technology.[43] For instance, we have a perfectly good pragmatic explanation for why the counterfactual 'If I had gone the beach or gone to the beach and got attacked by a shark, I would have had a good time' seems false, even when we judge the counterfactual 'If I had gone to the beach, I would have had a good time' to be true. The unnecessary disjunct in the former raises to salience the possibility of shark attacks — something one wouldn't normally consider when evaluating the latter counterfactual. This sort of maneuver is central to the strict theory of counterfactuals, but can also be invoked by the variably strict theorist.

There are also many positive cases for Substitution. One might simply be swayed by the theoretical virtues of the Boolean theory of propositions, in which logically equivalent propositions are identified. If the proposition A and $A \vee (A \wedge B)$ are identical, no compositional semantics can be given for any operator that does not respect their intersubstitutivity.[44] Similar puzzles involving free choice permission and modals exhibit similar phenomena, but do not seem to be easily explained semantically.[45] There are also metaphysical puzzles for the idea that logically equivalent propositions can have different counterfactual implications. For counterfactuals are connected to a large range of philosophical concepts, including causation, chance, and rational action. What should we make of the hypothesis that causes, seemingly being just described in logically equivalent ways, have different effects? Or that we should value logically equivalent things differently, even absent any logical confusions, as would be the case if the value of an action is given by its counterfactual consequences.[46] We have seen that chance-theoretic concerns endow various counterfactual principles with a certain sort of positive status. The same can be extended to Substitution: the probability calculus ensures that the chance of B conditional on A and on A' are the same when A and A' are logically equivalent, so the chances of $A \mathrel{\Box\!\!\rightarrow} B$ and $A' \mathrel{\Box\!\!\rightarrow} B$ must be the same if they are identical to the conditional chances.

3.4 Disjunction

In this section we will give a couple of more direct arguments against the similarity semantics. They will trade on exactly the sorts of 'first ball to fall' intuitions that Fine appeals to to motivate Negative Effect in his argument against the similarity semantics. These alternative arguments are interesting, however, because: (i) they do not involve Substitution, (ii) they do not involve disjunctive antecedents, and (iii) they do not involve infinity. If this is right, I think this undermines some of the morals that Fine draws from the case of the infinite slope discussed in section 3.2. Indeed, we will see that some of the principles that Fine

[43] See for example (Fox).

[44] Of course, we must be careful when it comes to attitude reports, and other connectives that are sensitive to the mode under which the proposition is presented. Counterfactuals, presumably, are not among these connectives.

[45] See (Goldstein, forthcoming) for a helpful characterization of the problem, and (Anglberger et al., 2016) for a truth-maker inspired approach to the problem.

[46] As is the case in causal decision. See, for instance, Gibbard and Harper (Gibbard and Harper, 1978).

himself endorses, and uses in his derivation, are undermined by these intuitions. Since the examples are finitary, diagnoses involving the limit assumption and Infinite Conjunction are also off the table.

It should of course be noted that on all plausible versions of the similarity semantics, counterfactuals are highly context sensitive, and so there is plenty of room for the similarity theorist to resist these putative counterexamples by positing some sort of shift in context. These responses will effectively amount to the denial of the truth of Negative Effect, and variants, in some contexts, and so will be available as responses to Fine's argument as well. My goal is primarily to establish that *if* we take the Finean judgments at face value, certain other things follow (some of which are in tension with Fine's own discussion). With that said, I am myself inclined to take these judgments at face value, and thus take us to have *prima facie* reasons to revise the similarity semantics, albeit not in the way Fine envisions. I will develop an alternative in which it is Disjunction, not Substitution, that gets revised. I will then apply it to our puzzles of infinity.

I think there are two main motivations for accepting Disjunction. The first is that many instances of it just seem to be primitively compelling, for instance 3 seems to follow from 1 and 2.

1. If John went to the party, everyone would have a good time.

2. If Mary went to the party, everyone would have a good time.

3. If John or Mary went to the party, everyone would have a good time.

The example in 1-3 is representative of the sort of cases one might use to motivate Disjunction. However, one must take care not to overgeneralize from particular examples: even an invalid inference may have instances in which the premises necessitate the conclusion. And even the judgment of validity in this instance is fragile. For example, suppose that both John and Mary are the life of any party, but are hostile exes. Because they wish to avoid each other, neither intend to come to the party. Because Mary isn't intending to go to the party, it would have been the case, had John gone, that everyone would have had a good time. Similarly, because John isn't intending to come to the party, had Mary gone, everyone would have had a good time. But had Mary or John gone to the party, they might have both gone. In which case, the party would have been a disaster. So we should reject 3, despite accepting 1 and 2. The intuition against 3 can be made more vivid, if we consider the logically equivalent counterfactual

3′. If either John or Mary or both went to the party, everyone would have had a good time.

3 and 3' are equivalent given Limited Substitution, which even the Finean semantics permits.[47] (To see this equivalence, note that steps 4-10 of our above derivation of Weak Antecedent Strengthening are reversible.)

[47] Of course, Fine should be perfectly happy with this result, since it trades on the possibility of John and Mary both going to the party — exactly the sorts of instances Disjunction he rejects.

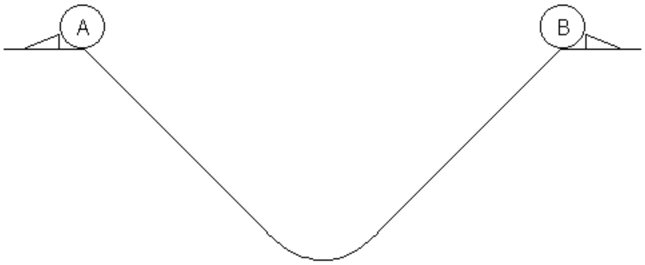


Figure 2: Two balls A and B balanced on opposing slopes.

The second consideration in favour of Disjunction is that its validity is predicted by the now dominant account of counterfactuals, prominently defended by Lewis and Stalnaker, based on similarity semantics. For roughly, if the closest A worlds are C worlds, and the closest B worlds are C worlds, then the closest $A \vee B$ worlds are C worlds. I will now turn to some further arguments against the similarity semantics. These will target principles validated by the similarity semantics that are, relative to a general class of possible worlds semantics, stronger than Disjunction, but equivalent to it given Conditional Excluded Middle (we will make these remarks precise in the next section).

Let us begin with a principle that is sometimes known as CSO. Putting it roughly, it says that if A and B counterfactually imply one another, they are *counterfactually equivalent* in the sense that they counterfactually imply the same propositions:

Counterfactual Equivalence $A \Box\!\!\rightarrow B, B \Box\!\!\rightarrow A, A \Box\!\!\rightarrow C \vdash B \Box\!\!\rightarrow C$

Counterfactual Equivalence is validated by every version of similarity semantics on the market; indeed it it hard to see how one could invalidate it without relinquishing the central insights of the similarity semantics.[48] A simplified argument is available if we make the limit assumption: if the closest A-worlds are B-worlds and the closest B-worlds are A-worlds, then the closest A-worlds must be the same as the closest B-worlds. So A and B must be counterfactually equivalent in the sense that they counterfactually imply the same propositions. A very similar argument for the validity of Counterfactual Equivalence is available if we do not assume the limit assumption.

Now consider the following scenario, in which balls A and B are balanced at opposing ends of a half-pipe, as in figure 2. If A toppled, it would roll to the other side and topple B, and similarly, if B toppled it would roll up to

[48] Fine considers a version of the similarity semantics in which a counterfactual $A \Box\!\!\rightarrow B$ is true iff B is true in the closest A-worlds and in the A-worlds that are not closer to actuality than a closest A-world. This hypothetical theory invalidates Counterfactual Equivalence, and is stated in terms of similarity, although it has not found any proponents as far as I know.

the other side and topple A. (We may suppose, by contrast, that if they both toppled they would collide in the middle.) Since both balls are in fact balanced, and never fall, there's a natural intuition that if one of the balls — A say — were to topple, it would knock B and not the other way around. This is of a piece with Fine's case for Negative Effect, in which it is assumed that if a ball were to topple it would not be the case that some higher ball knocked it: it would be the first ball to fall. I'll appeal to this sort of intuition repeatedly: for convenience lets call it the 'first ball to fall' intuition — it is of a piece with the more general thought that counterfactuals have something to do with the direction of causation. Suppose that A and B are on two pressure plates, set up so that if A's plate is released while B's plate is depressed a buzzer goes off (but not conversely). Then we would want to assert:

✓ If A were to topple, the buzzer would go off.

and deny

✗ If B were to topple, the buzzer would go off.

We also have:

✓ If A were to topple, B would topple.

✓ If B were to topple, A would topple.

These judgments jointly contradict Counterfactual Equivalence. Since Substitution was not involved in the 'derivation' of this contradiction, it cannot be this aspect of the possible worlds semantics that is to blame. Indeed, we do not need a special argument to see this, as with the infinite slope.

Counterfactual Equivalence is closely related to principles of counterfactual logic that Fine considers uncontentious. Here is a weaker principle that Fine himself explicitly endorses:[49]

Restricted Transitivity $A \mathrel{\Box\!\!\rightarrow} B, A \wedge B \mathrel{\Box\!\!\rightarrow} C \vdash A \mathrel{\Box\!\!\rightarrow} C$

Restricted Transitivity is also validated in every version of the similarity semantics on the market, and is valid in Fine's framework.[50] But it also appears to be refuted by similar judgments:

✓ If A were to topple, then B would topple.

[49] Counterfactual Equivalence entails Restricted Transitivity given Identity, Finite Conjunction and Weakening. Suppose that $A \mathrel{\Box\!\!\rightarrow} B$ and $A \wedge B \mathrel{\Box\!\!\rightarrow} C$. Given Entailment we have $A \wedge B \mathrel{\Box\!\!\rightarrow} A$ and from the first assumption, Identity and Finite Conjunction we have $A \mathrel{\Box\!\!\rightarrow} (A \wedge B)$. So A and $A \wedge B$ are counterfactually equivalent, and since $A \wedge B \mathrel{\Box\!\!\rightarrow} C$ we can conclude $A \mathrel{\Box\!\!\rightarrow} C$.

[50] It is easy remove the constraint in Fine's framework that secures Restricted Transitivity, but Fine endorses Restricted Transitivity, writing of Restricted Transitivity, Identity, Weakening and Weak Disjunction that 'there appears to be no plausible counterexamples to them, and their use in counterfactual reasoning is, on the face of it, completely unproblematic' p37 (Fine, 2012a).

✓ If A and B were to topple, A and B would collide somewhere in the middle.

✗ If A were to topple, A and B would collide somewhere in the middle.

Again, the judgment that the second counterfactual is true and the final one false seem to be of a piece with Fine's own justification for Negative Effect — what we have called the 'first ball to fall' intuition. But this reveals an apparent tension in Fine's own argument, because it undermines Restricted Transitivity — a principle Fine uses, in conjunction with Negative Effect, to refute Substitution.

These two principles are part of a wider class of principles of counterfactual logic governing the properties of counterfactual containment: the relation between propositions that holds when the counterfactual consequences of one or more antecedents is contained in the counterfactual consequences of others. Counterfactual Equivalence may be thought of as the limiting case of containment in both directions. These include:

Disjunction $A \mathrel{\Box\!\!\to} C, B \mathrel{\Box\!\!\to} C \vdash A \vee B \mathrel{\Box\!\!\to} C$

Reverse Disjunction $A \vee B \mathrel{\Box\!\!\to} C \vdash (A \mathrel{\Box\!\!\to} C) \vee (B \mathrel{\Box\!\!\to} C)$

Restricted Antecedent Strengthening $A \mathrel{\Box\!\!\to} B, A \mathrel{\Box\!\!\to} C \vdash A \wedge B \mathrel{\Box\!\!\to} C$

Restricted Antecedent Strengthening′ $\neg(A \mathrel{\Box\!\!\to} \neg B), A \mathrel{\Box\!\!\to} C \vdash A \wedge B \mathrel{\Box\!\!\to} C$

Restricted Transitivity $A \mathrel{\Box\!\!\to} B, A \wedge B \mathrel{\Box\!\!\to} C \vdash A \mathrel{\Box\!\!\to} C$

Counterfactual Equivalence $A \mathrel{\Box\!\!\to} B, B \mathrel{\Box\!\!\to} A, A \mathrel{\Box\!\!\to} C \vdash B \mathrel{\Box\!\!\to} C$

Since Disjunction and Reverse Disjunction both involve disjunctive antecedents, and therefore are distracting for reasons we have mentioned already, let us set them aside for a minute. Apart from those two, each of these principles is easily seen to be in conflict with the 'first ball to fall' intuition.[51] They are also all principles that are distinctive to the similarity semantics in the following senses. First, they are all validated by the similarity semantics, but are not a commitment of possible world semantics in general (we will describe some alternatives shortly).[52] Second, if we moreover assume Conditional Excluded Middle, and some other principles of counterfactual logic that ought to be part of any reasonable possible worlds semantics, they are all equivalent (see the next section).

In the case of Counterfactual Equivalence and some of the other principles listed above, we have targeted principles of counterfactual logic validated by the similarity semantics, but not explicitly identified by Fine as being one of

[51] For example, in our original counterexample to Counterfactual Equivalence it seemed true that if A were to fall, B would fall, and that if A were to fall, the buzzer would go off, but not true that if A and B were to fall, the buzzer would go off (since the buzzer does not go off when both pressure plates are released). This contradicts Restricted Antecedent Strengthening and Restricted Antecedent Strengthening′.

[52] They are all invalid in the random selection style semantics defended for counterfactuals by Schulz (Schulz, 2017), and for indicatives Bacon (Bacon, 2015), for example.

the theses that similarity semantics gets right (unlike, say, Restricted Transitivity which is explicitly identified this way). But I think these counterexamples nonetheless indicate that Fine has misdiagnosed the problem that the ‘first ball to fall’ intuition poses for the standard similarity semantics. It’s helpful to distinguish two different facets of that semantics. The first is that it is a form of *possible worlds* semantics, and thus licenses the substitution of logically equivalent sentences in all contexts. The second is that it is a possible world semantics formulated in terms of a *similarity ordering*, which ensures the validity of principles like Counterfactual Equivalence, Restricted Transitivity, Disjunction and the other principles listed that other possible worlds semantics do not. Fine’s argument appealed to both facets of the similarity semantics, but he blamed the possible worlds aspect of the semantics (namely its commitment to Substitution) rather than the similarity aspect (Weak Disjunction and Restricted Transitivity). The fact that these counterexamples generalizing the Negative Effect intuition target the latter class of principles (and only these principles) strongly suggests a different diagnosis: a diagnosis in which it is the principles distinctive to the similarity semantics, but not possible world semantics more generally, that are responsible for the conflict.

Are there any direct counterexamples to Disjunction? Instead of Disjunction consider the following variant:

Disjunction′ $A \mathrel{\Box\!\!\rightarrow} C, B \mathrel{\Box\!\!\rightarrow} C \vdash (A \vee B \vee (A \wedge B)) \mathrel{\Box\!\!\rightarrow} C$

Repurposing our earlier remark about $3'$, it can be seen that Disjunction′ is equivalent to Disjunction given an application of Substitution which even Fine’s (Fine, 2012b) account accepts. Let us suppose that if exactly one plate is released, a buzzer goes off, but if both are depressed or both released it doesn’t. In line with Negative Effect, it seems we are in a position to assert that if A were to topple the buzzer would go off, that if B were to topple the buzzer would go off, and that if both were to topple the buzzer wouldn’t go off. But it does not seem that we should assert that if A or B or both were to topple the buzzer would go off, because in particular if both were to topple it wouldn’t. The analogous judgment with respect to Disjunction is not as clear: if either A or B were to topple, would the buzzer go off? I think we are sometimes tempted to say *yes*, but this divergent reaction is readily explained by the fact that we often interpret disjunctions in English as exclusive. When we rephrase the conclusion so that the disjunction is explicitly an inclusive one, as we have done in Disjunction′, we do not have the same judgment.

Given the unclear status of Disjunction, a possible worlds theorist might wish to be able to retain it whilst also accommodating the ‘first ball to fall’ judgments, by selectively rejecting Restricted Transitivity, Counterfactual Equivalence and any other principles less equivocally in conflict with the ‘first ball to fall’ intuition. But the prospects for such a strategy are dim, since given principles that I think are not contentious in this context, and Substitution — a feature of any compositional possible world semantics — Disjunction (indeed Weak Disjunction) entails Restricted Transitivity, so that it is not possible to have the former without the latter:

1. $(A \wedge B) \boxright C$ (assumption)
2. $(A \wedge B) \boxright (B \supset C)$ (Weakening)
3. $(A \wedge \neg B) \boxright (B \supset C)$ (Entailment)
4. $(A \wedge \neg B) \vee (A \wedge B) \boxright (B \supset C)$ (Weak Disjunction from 2 and 3)
5. $A \boxright (B \supset C)$ (Substitution)
6. $A \boxright B$ (assumption)
7. $A \boxright B \wedge (B \supset C)$ (Finite Conjunction)
8. $A \boxright C$ (Weakening)

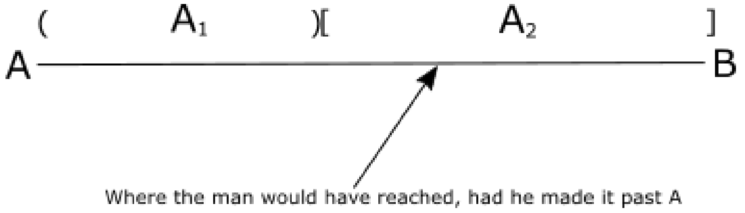
4 Solving the paradox by denying Disjunction

In this section I shall be advocating that we reject Disjunction both as a way of resolving our two paradoxes of infinity, and also of maintaining the 'first ball to fall' intuition in Fine's example of the infinite slope, and the finite examples considered above. There are two versions of the Disjunction denying view: one that combines it with Conditional Excluded Middle, and another that does not. Both views can be given a possible worlds semantics, and will validate Substitution.

Both these views are part of a more general strategy for dealing with cases that, in Lewis's semantics, would give rise to counterfactual pathologies. Let's start with the version that accepts Conditional Excluded Middle, and let us apply it to Benardete's paradox. Recall that A and B are two points a mile apart, and the xth-mile mark refers to the point x miles past A on the straight line from A to B.

What would have happened if the man had made it *some* non-zero distance past A? The picture that goes along with Conditional Excluded Middle is this: there is a particular distance past A such that, had the man made it past A, he would have been stopped there. But it is a chancy (or possibly indeterminate) matter which distance that would have been. (Perhaps the smaller distances are more likely to be the counterfactually selected distance than larger — I won't pass judgment on this issue here.) Informally: if he had gotten some distance past A then infinitely many gods in a row will have failed to build their walls, contrary to their resolutions, but it's a completely indeterminate or chancy matter which infinite sequence of gods it would be. It's possible (even if unlikely) that the man would have gotten all the way to B, had he gotten some distance past A. In which case we we are not in a position to even assert the following:

> If the man had made it some distance past A, he wouldn't have made it all the way to B.



I think the idea that, for all we know, 3 is true is independently appealing: if infinitely many gods failed to build their walls, but we're told no more, then all bets are off, as it were — perhaps they *all* failed to build their walls.

Given Divine Dispositions we may see how the resulting view must invalidate Disjunction. Firstly note that there's some particular point between A and B, x, such that, had the man made it some way past A, he would have stopped at x: see the point indicated in figure 4. Now we may pick a point $\frac{1}{2}^n$ miles past A that is before x. For convenience we will identify points on the line AB with numbers from 0 to 1, representing the distance past A in miles. We can split the line AB into two intervals: $A_1 = (0, \frac{1}{2}^n)$ and $A_2 = [\frac{1}{2}^n, 1]$. Overloading our notation in the obvious way, we will write A for the proposition that the man stopped in $(0, 1]$, A_1 for the proposition that was stopped in $(0, \frac{1}{2}^n)$ and A_2 for the proposition that hes was stopped in $[\frac{1}{2}^n, 1]$. Finally, let C be $(0, \frac{1}{2}^n]$. By Divine Dispositions we know that if the man got past the $\frac{1}{2}^{n+1}$ mark, he was stopped by the $\frac{1}{2}^n$ mark, which lies in C. Thus $A_2 \mathrel{\Box\!\!\rightarrow} C$ is true. If the man stopped in the interval $(0, \frac{1}{2}^n)$ then he certainly stopped in the interval $(0, \frac{1}{2}^n]$ which contained it, thus $A_1 \mathrel{\Box\!\!\rightarrow} C$. But $(A_1 \vee A_2 \mathrel{\Box\!\!\rightarrow} C)$ is false: we began by assuming that if the man was stopped in the interval $(0, 1]$ he would stop at the point x which is after $\frac{1}{2}^n$.

Notice that we do not know in advance what our counterexample to Disjunction will be: it is a chancy matter where the many would have stopped, had he made it past A, and since our construction of A_1 and A_2 depended on where he would have stopped, it is a chancy matter which partition of A yields our counterexample to Disjunction. Thus we make the welcome prediction that negations of particular instances of Disjunction are unassertable, even if the principle itself is not valid.

The version of the view that rejects Conditional Excluded Middle simply replaces the invocation of indeterminacy or chanciness with falsity. What would have happened if the man had made it past A? As before, all bets are off, but this means that the relevant counterfactuals are simply false:

> It's false that, if the man had made it some distance past A, he would have made it to B.
>
> It's false case that, if the man had made it some distance past A, he wouldn't have made it to B.

By modifying the above reasoning, one can also see that this version of the view also renounces Disjunction.

What of Yablo's Button? Again, we wish to maintain that had the display read 'chocolate' on day 0 (i.e. he had declined on every previous day), he would press the button on day 0. And of course, we also must maintain that had it both been the case that he pressed the button on day 0 and there had been an infinite sequence of days on which he declined, then he would have pressed the button on day 0. (The antecedent here entails the consequent.) But on the other hand, if it had been the case that there had been some infinite sequence of days on which he declined, then (according to the CEM version) it's completely indeterminate whether he declined on every day, or every day until day 0, or every day until day -1, and so on. In which case, it's true, for all we know, that had there been an infinite sequence of days on which he declined, he would have declined on every day. This similarly conflicts with Disjunction and Substitution. The attributions of indeterminacy are replaced by attributions of falsity in the anti-CEM variant. Finally, there is the case of the infinite slope: the diagnosis is similar, although by now it should be evident how to apply the picture.

What sort of theory of counterfactuals could substantiate these verdicts? Once Disjunction, and the other contentious principles identified in the last section, are ceded, are any substantive principles of counterfactual logic left? One might worry that any such logic will be overly weak, or *ad hoc* and tailored to solve these particular paradoxes. In order to assuage this worry, I will outline my favored logic of conditionals, indicative and subjunctive alike, and explain why it is natural and independently motivated. I will then show it has models corroborating the sort of picture of the paradoxes just outlined. The *theory* determined by a list of axioms A and rules R will mean the smallest set containing the axioms in A and the propositional tautologies which is also closed under the rules in R, the rule of uniform substitution, and the rule of modus ponens for the material conditional (notated as $\supset$ below).

Necessitation If $\vdash B$ then $\vdash A \mathrel{\Box\!\!\rightarrow} B$

Substitution If $\vdash A \equiv B$ then $\vdash (A \mathrel{\Box\!\!\rightarrow} C) \equiv (B \mathrel{\Box\!\!\rightarrow} C)$

Normality $(A \mathrel{\Box\!\!\rightarrow} (B \supset C)) \supset ((A \mathrel{\Box\!\!\rightarrow} B) \supset (A \mathrel{\Box\!\!\rightarrow} C))$

Identity $A \mathrel{\Box\!\!\rightarrow} A$

Modus Ponens $(A \mathrel{\Box\!\!\rightarrow} B) \supset (A \supset B)$

Conditional Excluded Middle $(A \mathrel{\Box\!\!\rightarrow} B) \vee (A \mathrel{\Box\!\!\rightarrow} \neg B)$

Absurdity $(A \mathrel{\Box\!\!\rightarrow} B) \supset ((B \mathrel{\Box\!\!\rightarrow} \bot) \supset (A \mathrel{\Box\!\!\rightarrow} \bot))$

I will call this logic LC. LC can be extended to an infinitary logic, and further principles like Infinite Conjunction may be added. Since issues of completeness become more complex, I will not investigate these systems thoroughly here.

Each of the principles listed may be given a probabilistic motivation analogous to the ones we gave for Infinite Conjunction and Conditional Excluded Middle, with the exception of Modus Ponens, which we will take as a primitive assumption.[53] Note that even with the caveat about Modus Ponens, the *rule* of counterfactual modus ponens, $A \mathrel{\Box\!\!\rightarrow} B, A \vdash B$ is probabilistically valid since if $Pr(B \mid A) = Pr(A) = 1$ then $Pr(B) = 1$. The justification of the axiom Absurdity is a little more subtle, as it depends on how one treats conditional probabilities on propositions with zero probability, but it is delivered by the two most natural choices.[54]

None of the principles listed in the last section, including Disjunction, Counterfactual Equivalence and Restricted Transitivity, are derivable in this logic. Moreover, we have the following nice characterization of these principles:

Theorem 4.1. *Stalnaker's logic is equivalent to the result of adding any of the following principles to* LC*:*

- *Disjunction*
- *Restricted Transitivity*
- *Counterfactual Equivalence*
- *Restricted Antecedent Strengthening*
- *Restricted Antecedent Strengthening′*
- *Reverse Disjunction and Absurdity*$^+$

In the above Absurdity$^+$ refers to the following strengthening of Absurdity:

Absurdity$^+$ $(A \mathrel{\Box\!\!\rightarrow} \bot) \supset ((B \mathrel{\Box\!\!\rightarrow} C) \equiv (A \vee B \mathrel{\Box\!\!\rightarrow} C))$

It follows that, at least from the perspective of LC, all of the principles listed above are equivalent.[55]

It is also worth emphasizing that just as there are probabilistic considerations in favour of the logic LC, there are probabilistic arguments *against* all of these further principles that are part of Lewis and Stalnaker's logic. For example, Stalnaker (Stalnaker, 1976) has shown that any probability function and interpretation of the conditional jointly satisfying Stalnaker's thesis and Stalnaker's logic of conditionals will be one in which there are at most two disjoint

[53] See (van Fraassen, 1976). Van Fraassen uses an assumption that amounts to the claim that if A and B are guaranteed to have the same probability, given the conditional probability constraint, then they denote the same proposition. In particular, any pair of sentences results from the other from such a substitution have the same probability relative to *any* probability function (whether it satisfies the conditional probability constraint or not). Van Fraassen uses this principle to justify CE, a weaker logic than the one presented here, and he uses a slightly stronger assumption than the claim that Modus Ponens has probabiliy 1.

[54] According to one convention $Pr(B \mid A) = 1$ when $Pr(A) = 0$. Another approach is to take conditional probabilities as primitive. It should be noted that some of these ideas can be used to motivated a principle that is stronger than Absurdity, namely $(A \mathrel{\Box\!\!\rightarrow} \bot) \supset ((B \mathrel{\Box\!\!\rightarrow} C) \equiv (A \vee B \mathrel{\Box\!\!\rightarrow} C))$. I consider this principle to be invalid, even though it must have probability 1 given the link between conditional probabilities and probabilities of conditionals, although I will have to defer discussion of these issues to other work.

[55] I have obtained object language proofs of these equivalences, but they are rather lengthy, so I will not reproduced them here. See (Bacon).

propositions with positive probability. And given the probabilistic validity of LC, we have that any of the listed principles, including Disjunction, will similarly trivialize the probability space.

There is a question of the completeness of LC. Since we have not provided a semantics we are not in a position to ask a precise version of this question. But since we have already suggested a probabilistic criterion of validity, the following question presents itself. Abstractly we might consider the class of conditional algebras (c, W, Pr, Σ) where (Pr, Σ) is a probability measure on $P(W)$, and $c : P(W) \times P(W) \to P(W)$ a binary operation subject to the constraint that $Pr(c(A, B)) = Pr(B \mid A)$ whenever $A, B \in P(W)$ and $Pr(A) > 0$, and $c(A, B) \cap A \subseteq B$. A sentence is valid if, when interpreted in such an algebra in the obvious way, it denotes a proposition with probability 1. LC is incomplete for this notion of validity, since, for example, the following principles are probabilistically valid but not provable:

B $A \supset ((A \mathrel{\Box\!\!\to} \bot) \mathrel{\Box\!\!\to} \bot)$

S4 $(A \mathrel{\Box\!\!\to} \bot) \supset (B \mathrel{\Box\!\!\to} (A \mathrel{\Box\!\!\to} \bot))$

If we were to define a notion of counterfactual necessity as $\Box A := \neg A \mathrel{\Box\!\!\to} \bot$ then these principles are equivalent in LC to the B and $\mathsf{S4}$ axioms for $\Box$ respectively.[56]

Let me instead turn to a more traditional possible world semantics for which LC is complete, and in which we can find concrete models of Divine Dispositions, Anti-Zenoism, Negative Effect and other principles that cannot be modeled in a similarity framework. It is a relatively standard semantics based on selection functions. A selection function is a function, $f : P(W) \times W \to P(W)$, that take a proposition represented by a set of worlds, A, and a world, w and, roughly speaking, outputs the set of worlds that, at w, might have been the case if A (see (Chellas, 1975)). A model additionally assigns sets of worlds, $[\![P]\!]$, to sentences letters P, and is extended recursively to other sentences, with $[\![A \wedge B]\!] = [\![A]\!] \cap [\![B]\!]$ and $[\![\neg A]\!] = W \setminus [\![A]\!]$ as usual, and $[\![A \mathrel{\Box\!\!\to} B]\!] = \{x \mid f([\![A]\!], x) \subseteq [\![B]\!]\}$. A sentence A is valid in a model if $[\![A]\!] = W$, and valid in a class of models if it is valid in each member of the class. LC is complete with respect to models whose selection functions are subject to the following constraints:[57]

MP $x \in f(A, x)$ whenever $x \in A$

ID $f(A, x) \subseteq A$

CEM $|f(A, x)| \leq 1$

AB If $f(A, x) \subseteq B$ and $f(B, x) = \emptyset$ then $f(A, x) = \emptyset$

[56] I do not presently know what the logic of these conditional algebras is, or if it is even recursively axiomatizable.

[57] The completeness follows straightforwardly from the canonical model construction in (Chellas, 1975).

CEM guarantees that $f(A,x)$ is always either empty or a singleton. In the latter case, we may think of the unique member of $f(A,x)$ as the world that *would have obtained* had A been true.

To model infinitary conjunctions we must add the obvious clause for infinitary conjunctions: $[\![\bigwedge_n A_n]\!] = \bigcap_n [\![A_n]\!]$. In the infinitary language the above semantics also validates Infinite Conjunction$_\kappa$ for any κ.

4.1 Models of the Paradoxes

We may obtain models of Divine Dispositions and Anti-Zenoism, Chocolate Preference and Consistency, and Negative and Positive Effect within this model theory as follows. We will begin with Benardete's paradox. Our worlds will consist of the natural numbers, adjoined with a maximum element, representing the actual world, which should be thought of as lying above all the natural numbers. n will represent the world in which the man is halted at the $\frac{1}{2}^n$-mile mark (but gets as far as the $\frac{1}{2}^{n+1}$-mile mark). In the actual world he does not make it any distance past A.

- $W = \mathbb{N} \cup \{@\}$
- $[\![P_n]\!] = \{1, ..., n\}$

Since we only need to validate Divine Dispositions and Anti-Zenoism at the actual world, when $n \neq @$ we may define $f(A,n)$ arbitrarily subject to the constraints MP, ID, CEM and AB (e.g. $f(A,n) = \{min(\{x \in A \mid x \geq n\})\}$, understanding this as empty when $\{x \in A \mid x \geq n\}$ is empty). At the actual world:

- $f(A,@) = \emptyset$ if $A = \emptyset$
- $f(A,@) = \{max(A)\}$ if A has a maximum.
- $f(A,@)$ is some arbitrary singleton lying in A if A has no maximum.

Recall that we are treating @ as lying above every natural number. Thus a non-empty set, A, has a maximum if and only if it is either finite, or infinite and contains @ (in which case its maximum is @). Clearly our constraints, MP, ID, CEM and AB are satisfied, so it is a model of LC. Moreover, Divine Dispositions is true at the actual world. It is part of our modeling assumption that the man travels continously, so in fact $[\![P_n]\!]$ and $[\![P_{\geq n}]\!]$ have the same semantic value: $\{1, ..., n\}$. To validate Divine Dispositions, we must show that $@ \in [\![P_{\geq n+1} \Box\!\!\rightarrow \neg P_n]\!]$, which amounts to showing that $f(\{1, ..., n+1\}, @) = \{n+1\}$ is a subset of $W \setminus \{1, ..., n\}$, which of course, it is. By contrast, $[\![\bigvee_n P_{\geq n}]\!] = \bigcup_n \{1, ..., n\} = \mathbb{N}$, and so $f(\mathbb{N}, @)$ is some arbitrary singleton contained in $\mathbb{N}$ — it is not the emptyset. Thus we validate Anti-Zenoism.

This is, of course, the version of the view that accepts Conditional Excluded Middle. A variant semantics drops the constraint CEM, and an alternative model of Benardete's paradox is available where we instead set $f(A,@) = A$ in the case that A has no maximum. This comports with our stipulation that the counterfactuals with the antecedent that the man made it some distance past A are generally false (unless the consequent is entailed by the antecedent).

By a simple reinterpretion of what our worlds represent, we may use the above to model Fine's example of the slope in such a way that Consistency, Positive Effect and Negative Effect are all validated. In the alternative interpretation, the world $n \in \mathbb{N}$ represents a world in which the nth ball is the first to fall (and the remaining balls fall), and @ a world where no balls fall. This demonstrates that the Disjunction denier also has the resources to accommodate Fine's puzzle, if they wish to.

The case of Yablo's button is different. If we are only interested in modeling Chocolate Preference and Consistency, the above model will do. But we also had a further principle, Zap Avoidance, that has no analogue in the other two puzzles. Instead we might model the worlds with functions from numbers to 1s and 0s, along with the actual world @ in which you don't play Yablo's button. A function which maps n to 1 represents a world in which you press the button on the $-n$th day, and a function which maps n to 0 a world where you decline to. Thus:

- $W = 2^{\mathbb{N}} \cup \{@\}$
- $[\![D_n]\!] = \{t \in 2^{\mathbb{N}} \mid t(n) = 0\}$
- $A = 2^{\mathbb{N}}$

Thus $[\![D_{\geq n}]\!] = \{t \in 2^{\mathbb{N}} \mid t(m) = 0 \text{ for all } m \geq n\}$. As before we treat $f(A, t)$ arbitrarily, subject to the above constraints, when $t \neq @$. So we just need to treat the case where $t = @$. Define a function, $best_n(t)$ that takes a sequence, and tells you what the best thing to do would have been on day n given the sequence of plays so far:

> $best_n(t) := 1$ if $t(m) = 0$ for every $m > n$ (the display reads 'Chocolate' on day n).
>
> $best_n(t) := 0$ otherwise. I.e., if $t(m) = 1$ for some $m > n$ (the display reads 'Zap' on day n).

We may then define a selection function, letting earlier bullet points take precedence over the later:

- $f(X, t) = \emptyset$ when $X = \emptyset$
- $f(X, t) = \{@\}$ if $@ \in X$
- $f(X, t) = \{t'\}$ where $t' \in X$ is such that, whenever $best_n(s) = b$ for every $s \in X$, $t'(n) = b$.
- $f(X, t)$ some arbitrary singleton contained in X if there is no t' like the above.

Informally, if every world in X agrees about what the best action would be on day n, then the world that would have been the case if X must be a world where that action is made on day n, assuming there is an X world like this. Note that when X is $[\![\neg D_{\geq n+1}]\!] = \{t \in 2^{\mathbb{N}} \mid t(m) = 1 \text{ for some } m \geq n+1\}$ then $best_1(t) = ... = best_n(t) = 0$ for every $t \in X$ (and for $m > n$, $best_m(t)$ can be both 1 and 0 across X). Thus any function mapping $1, ..., n$ to 0, and mapping some $m > n$ to 1 will both belong to X and satisfy the constraint of the second clause. Thus we see that $[\![A \wedge \neg D_{n+1} \mathrel{\Box\!\!\rightarrow} D_n]\!]$ contains @, since $f(X, @)$ must be the singleton of a function mapping n to 0. Similarly, when A is $[\![D_{\geq n+1}]\!]$, then the sequence mapping n to 1, everything else to 0 will satisfy the constraint. So $[\![A \wedge D_{n+1} \mathrel{\Box\!\!\rightarrow} \neg D_n]\!]$ contains @ since in that case $f(X, @)$ is the singleton of a sequence mapping n to 1. In order to get a version of this view without Conditional Excluded Middle, one can replace the fourth bullet point with $f(X, t) = X$, and the third bullet point so that $f(X, t)$ is the set of all $t' \in X$ satisfying the condition that whenever $best_n(s) = b$ for every $s \in X$, $t'(n) = b$.

What should we make of these models? Note, first, that as predicted, there are counterexamples to Disjunction, and even Weak Disjunction, in these models. It follows that the selection functions cannot be given the standard interpretation in which $f(A, x)$ denotes the singleton of the closest A-world to x (or the set of closest A-worlds). This much should be no surprise.

4.2 Interpreting the Semantics: Random Selection

One thing the models demonstrate is the consistency of Divine Dispositions, Anti-Zenoism and other problematic principles, with the logic LC which we have shown to be independently desirable. This is a non-trivial point that shouldn't be neglected. Yet one might still ask for more: after all, the logics of Lewis, Stalnaker and others are supported by a clear picture of what it takes for a counterfactual to be true or false based on the notion of similarity. What comparable underpinning is available for LC?

Moritz Schulz (Schulz, 2014), (Schulz, 2017) defends a semantics for counterfactuals which invalidates the principles we have been discussing. Schulz, like Lewis, rejects the uniqueness assumption for similarity, but like Stalnaker accepts Conditional Excluded Middle. For Schulz, the counterfactual selection function, $f(A, x)$ picks an element arbitrarily from among the most similar A-worlds.[58] For Schulz, arbitrary selection is a primitive notion to which counterfactuality is being reduced.

Unfortunately, when there are infinite descending chains of ever similar worlds, there is no such thing as the most similar worlds to select from. In this case Schulz says we should select from a set of sufficiently similar worlds (see §7.4 (Schulz, 2017)). The resulting view will substantiate a view of our

[58] Schulz replaces talk of 'similarity' with 'relevance', but he attributes the view so stated to Lewis, so I will assume that being 'more relevant than' plays roughly the same role in Schulz's discussion as being 'more similar than' does for Lewis. (Although, according to Schulz, they come apart when the limit assumption fails.)

paradoxes that has roughly the same shape as the view described earlier. For instance, on the counterfactual supposition that the man makes it some distance past A, we pick some suitable cutoff — perhaps, that the man makes it a meter past A — and select arbitrarily from the worlds in which the man makes it at most a meter past A. But the view, since it involves similarity, still has the trappings of that theory. For instance, suppose that I am in fact 5 foot tall, and that, other things being equal, worlds in which I am closer to my actual height are closer to actuality. Accordingly, the closest worlds where I am at least 6 foot tall, I am exactly 6 foot tall. So for Lewis, Stalnaker and Schulz alike, the seemingly false counterfactual 'if I had been at least 6 foot tall, then I would have been exactly 6 foot tall' must be true.[59]

An alternative version of this idea is developed in (Bacon, 2015) for indicative conditionals. In this version, similarity plays no role in the theory. Since this is the version I prefer, let me spell it out in a little more detail. The primitive of the theory is a function I will dub the *ur-selection function*: a function $f : P(W) \times W \to P(W)$ subject to the constraints MP, ID, CEM, and instead of AB, the stronger condition that $f(A, x) = \emptyset$ iff $A = \emptyset$.[60] The ur-selection function does not correspond to a conditional uttered in any ordinary context. An utterance of an indicative conditional may be interpreted as follows. In a given context there is usually some salient evidence — usually the speaker's — determining an accessibility relation, R. When someone utters an indicative conditional in such a context they express a conditional determined by the following selection function, writing $R(x)$ for $\{y \mid Rxy\}$:

$$f_R(A, x) = f(A \cap R(x), x)$$

As in Schulz's proposal, f is to be interpreted in terms of random selection. One understands $f(A, x)$ as selecting an A-world at random, and consequently f_R as selecting an accessible A-world at random. But unlike Schulz, we may select at random from any accessible A-world, not merely the closest. It is clear why, on this semantics, Disjunction is not valid: if I select (accessible) worlds at random from A, B and $A \cup B$ respectively, there is no guarantee I will pick any two worlds the same.[61] Thus if we pick a C that is true at the selected A world and the selected B world, but not the selected $A \cup B$ world, $A \to C$ and $B \to C$ are true but $A \vee B \to C$ false.

This formalism may be extended to counterfactuals. The speaker's evidence is not relevant for the evaluation of a counterfactual. Instead we find that more

[59] Schulz might lean more heavily on the distinction between his notion of 'relevance' and Lewis and Stalnaker's similarity. But severing the link between similarity and relevance makes the complaint that relevance is an unexplained primitive of the theory more acute.

[60] This last constraint secures a logic that extends LC, including Aburdity$^+$, B and S4. LC, by contrast, can be thought of the logic of selection functions generated by taking a selection function satisfying the stronger conditions, and restricting by an accessibility relation, as described below.

[61] In the special circumstance that $A = B$, we do have this guarantee since we only select once for each set. Similarly, when A and B are singletons we are certain to make the either the same choice as A or as B when we select from $A \cup B$.

'objective' accessibility relations secure counterfactual readings of the conditional in a given context. Here is one candidate: given a salient time provided by the context, often indicated by the antecedent, the relation E_t holds between world x and y if they agree, or substantially agree, about matters of particular fact up until t, and continue after t in the most likely ways given the laws at x and y. E_t is rarely the evidence of a person, for it would require them to know the entire history of the universe up until time t, and this explains why counterfactual and indicative conditionals with the same antecedents and consequents often diverge in truth value.

What does it mean to randomly select an antecedent world? There are many different processes for selecting something at random, like rolling a die, or spinning a wheel, and they do not all amount to the same thing. Unlike Schulz, who grounds this idea in a theory of arbitrary reference inspired by (Breckenridge and Magidor, 2012), I prefer to understand this as a primitively conditional notion. If I wished to randomly select between 'heads' and 'tails', I could simply take a coin out of my pocket and flip it. Alternatively, I could leave the coin in my pocket, and consider the way the coin would have landed had I flipped. Evidently, in the latter case I would not be able to observe the result, but given Conditional Excluded Middle, this is also a way of selecting heads or tails, albeit an inherently conditional one. There are benefits to this interpretation too: when a proposition A is true, then the selection of an A world is not random — MP constrains us to select the actual world. This constraint has to be baked in by hand in Schulz's framework. But it is readily explained once we concede the conditional nature of the selection process, since the world that would have obtained if A, when A is a truth, has to be the world that *in fact* obtains, by Modus Ponens.

The notion of random selection, then, is not much more than a helpful heuristic: it is not supposed to be a notion to which conditionality can be reductively defined. In this respect, I am in good company: Stalnaker, for example, is quick to concede that similarity is not a notion we had antecedently, to which conditionality can be reduced, but rather a notion that is as much to be understood in terms of conditionals as conditionals in terms of it ((Stalnaker, 1987)). Lewis is similarly explicit about this in *Counterfactuals* (Lewis, 1973). But the notion of similarity is not useless: it is heuristically valuable, and proves its worth by imposing structural constraints on selection functions, predicting a rich and powerful logic.

Yet I believe that I am in a better position that Stalnaker or Lewis, as I have been emphasizing connections between the ur-selection function and probabilistic notions that narrow its role down considerably. These connections substantively rule out many interpretations of the ur-selection function, including Lewis's, Stalnaker's, the material conditional ($f(A, x) = \emptyset$ whenever $x \notin A$, and $\{x\}$ otherwise), the strict conditional ($f(A, x) = R(x) \cap A$, for some accessibility relation R) and many others. Now we have set up the formal framework, we are finally in a position to state that connection explicitly.

If I am about to randomly select a ball from a bag, I might represent that formally using a random variable. Informally, a random variable is a non-rigid name for the ball that I pick, or more formally, a function mapping possible worlds to the ball that I pick in that world, $f : W \to B$, where B is the set of balls. The probability of the selected ball being in a given subset $B' \subseteq B$ is given by the the probability $Pr(\{w : f(w) \in B'\})$. For each antecedent proposition, A, a random variable for a randomly selected A world would be a function $f : W \to A$. We may thus reconceive of the ur-selection function as a collection of random variables, $f_A : W \to A$, one for each consistent antecedent, defined by setting $f_A(x)$ to be the unique member of $f(A, x)$. Our constraint then amounts to the idea that the probability of selecting a given A world is directly proportional to the probability of that world. When a proposition may be made up entirely of worlds with 0 probability, a slight strengthening is needed, and our thesis then becomes:

Proportionality For any rational ur-prior: the probability that the selected A-world, f_A, belongs to $B \subseteq A$ is proportional to the probability of B (provided the probability of A is non-zero).

Letting $\to$ denote the ur-conditional — the conditional defined by the ur-selection function — Proportionality ensures that for any ur-prior, Pr, $Pr(A \to B) = Pr(f_A \in B) = Pr(B \mid A)$. This constraint is important for delivering judgments about the probabilities of indicatives. On the assumption that the accessibility relation for evaluating counterfactuals, E_t, is the complete history up until t, and that the initial chances form an acceptable ur-prior, we can derive chance-theoretic versions of the connection from the Principal Principle.

This ends my brief outline of a theory of conditionals that validates the logic LC. This, of course, falls far short of a proper defense of the theory, which I will have to defer to future work (the beginnings of which can be found in (Bacon, 2015)). But such a defense would go well beyond my more modest aims here: to simply convey the sense that there are interpretations of conditionals that are not at all *ad hoc* which validate the logic of LC, invalidate Disjunction and other such principles, and in which our paradoxes of infinity can be resolved.

References

Albert J. J. Anglberger, Johannes Korbmacher, and Federico L. G. Faroldi. An exact truthmaker semantics for permission and obligation. In Olivier Roy, Allard Tamminga, and Malte Willer, editors, *Deontic Logic and Normative Systems*, pages 16–31. College Publications, 2016.

Andrew Bacon. Conditional logics supporting stalnaker's thesis. Unpublished.

Andrew Bacon. A paradox for supertask decision makers. *Philosophical Studies*, 153(2):307, 2011.

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

José A. Benardete. *Infinity: An Essay in Metaphysics*. Clarendon Press, 1964.

Wylie Breckenridge and Ofra Magidor. Arbitrary reference. *Philosophical Studies*, 158(3):377–400, 2012.

Michael Caie. Benardete's paradox and the logic of counterfactuals. *Analysis*, 78(1):22–34, 2018.

Brian F. Chellas. Basic conditional logic. *Journal of Philosophical Logic*, 4(2): 133–153, 1975.

Charles B. Cross and Donald Nute. Conditionals: From philosophy to computer science, edited by g. crocco, l. fariñas del cerro, and a. herzig, studies in logic and computation, no. 5, clarendon press, oxford university press, oxford and new york1995, viii + 368 pp. *Journal of Symbolic Logic*, 62(4):1487–1490, 1997. doi: 10.2307/2275657.

Kit Fine. Compliance and command iii, imperatives and deontic conditionals. Unpublished.

Kit Fine. A difficulty for the possible worlds analysis of counterfactuals. *Synthese*, 189(1):29–57, 2012a.

Kit Fine. Counterfactuals without possible worlds. *Journal of Philosophy*, 109 (3):221–246, 2012b.

Kit Fine. Compliance and command i—categorical imperatives. *Review of Symbolic Logic*, 11(4):609–633, 2018a. doi: 10.1017/S175502031700020X.

Kit Fine. Compliance and command ii, imperatives and deontics. *Review of Symbolic Logic*, 11(4):634–664, 2018b. doi: 10.1017/S1755020318000059.

Danny Fox. Free choice and the theory of scalar implicatures* mit,.

Allan Gibbard and William Harper. Counterfactuals and two kinds of expected utility. In A. Hooker, J. J. Leach, and E. F. McClennen, editors, *Foundations and Applications of Decision Theory*, pages 125–162. D. Reidel, 1978.

Anthony S. Gillies. Counterfactual scorekeeping. *Linguistics and Philosophy*, 30(3):329–360, 2007.

Simon Goldstein. Free choice impossibility results. *Journal of Philosophical Logic*, forthcoming.

Hans G. Herzberger. Counterfactuals and consistency. *Journal of Philosophy*, 76(2):83–88, 1979.

Mark Johnston. How to speak of the colors. *Philosophical Studies*, 68(3):221–263, 1992.

Angelika Kratzer. What 'must' and 'can' must and can mean. *Linguistics and Philosophy*, 1(3):337–355, 1977. doi: 10.1007/BF00353453.

David K. Lewis. *Counterfactuals*. Blackwell, 1973.

David K. Lewis. Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic*, 10(2):217–234, 1981. doi: 10.1007/BF00248850.

Matthew Mandelkern. Talking about worlds. *Philosophical Perspectives*, 33, forthcoming.

Sarah Moss. On the pragmatics of counterfactuals. *Noûs*, 46(3):561–586, 2012.

Sarah Moss. Subjunctive credences and semantic humility. *Philosophy and Phenomenological Research*, 87(2):251–278, 2013.

John Pollock. *Subjunctive Reasoning*. Reidel, 1976.

Moritz Schulz. Counterfactuals and arbitrariness. *Mind*, 123(492):1021–1055, 2014.

Moritz Schulz. *Counterfactuals and Probability*. Oxford University Press, 2017.

Brian Skyrms. The prior propensity account of subjunctive conditionals. In *The University of Western Ontario Series in Philosophy of Science*, volume 15, pages 259–265. 1980.

R. Stalnaker. Letter to van fraassen. *WL Harper and CA Hooker (1976)*, pages 302–306, 1976.

Robert C. Stalnaker. A theory of conditionals. In Nicholas Rescher, editor, *Studies in Logical Theory (American Philosophical Quarterly Monographs 2)*, pages 98–112. Oxford: Blackwell, 1968.

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

Robert C. Stalnaker. A defense of conditional excluded middle. In William Harper, Robert C. Stalnaker, and Glenn Pearce, editors, *Ifs*, pages 87–104. Reidel, 1981.

Robert C Stalnaker. *Inquiry*. MIT Press Cambridge, 1987.

Eric Swanson. On the treatment of incomparability in ordering semantics and premise semantics. *Journal of Philosophical Logic*, 40(6):693–713, 2011. doi: 10.1007/s10992-010-9157-z.

Eric Swanson. Conditional excluded middle without the limit assumption. *Philosophy and Phenomenological Research*, 85(2):301–321, 2012.

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

F. J. M. M. Veltman. Prejudices, presuppositions, and the theory of counterfactuals. 1976.

Kai von Fintel. Counterfactuals in a dynamic context. In Michael Kenstowicz, editor, *Ken Hale: A Life in Language*. MIT Press, Cambridge, 2001.

J. Robert G. Williams. Defending conditional excluded middle. *Noûs*, 44(4): 650–668, 2010.

S. Yablo. A reply to new zeno. *Analysis*, 60(2):148–151, 2000.