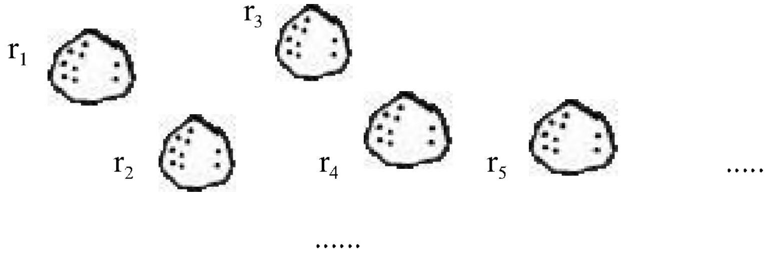
Counterfactuals Without Possible Worlds

Ever since the pioneering work of Stalnaker and Lewis[1], it has been customary to provide a semantics for counterfactuals statements in terms of possible worlds. Roughly speaking, the idea is that the counterfactual from A to C should be taken to be true just in case all of the closest worlds in which A is true are worlds in which C is true. Such a semantics is subject to some familiar difficulties - counterfactuals involving impossible antecedents, for example, or counterfactuals involving big changes consequential upon small changes. But it is not clear how seriously to take these difficulties - either because they might be met through some modification in the notion of closeness or because the intuitions on which the cases depend might be challenged or because the cases themselves might be dismissed as peripheral to the central use of the counterfactual construction; and nor has it been clear what a more satisfactory alternative to the possible world semantics might be put in its place.

I should like to suggest that the possible worlds semantics for counterfactuals faces a more serious difficulty, which cannot be so easily remedied. For the possible worlds semantics requires that the truth-value of a counterfactual statement should be preserved under the substitution of logically equivalent antecedents (since these will be true in the same possible worlds). But this substitution principle is incompatible with the combination of certain intuitively compelling counterfactual judgments and certain intuitively compelling principles of reasoning. Thus adoption of the semantics forces us to make an unpalatable choice between the particular counterfactual judgements, on the one hand, and the general principles of counterfactual reasoning, on the other. I should also like to propose an alternative semantics, using possible states in place of possible worlds, which avoids the difficulties and which is more satisfactory than the possible worlds semantics in a number of other respects.

Let us first consider the difficulties. Imagine an infinite landscape containing infinitely many rocks fairly close to one another (substitute stars in an infinite universe, if you like):



Use R_n for the sentence 'Rock n falls', R for the disjunction $R_1 \vee R_2 \vee R_3 \vee \ldots$, and '>' for the counterfactual operator (so that $A > C$ may be read 'if were the case that A then it would be the case that C'). Let us suppose that none of the rocks actually falls but that if a given rock r_n were to fall it would always fall in the direction of rock r_{n+1}. The following non-logical assumptions are then plausibly taken to be true (for n = 1, 2, 3, ...):

Positive Effect $R_n > R_{n+1}$
(if a given rock were to fall then the next rock would fall)

Negative Effect $R_{n+1} > \neg R_n$, n = 1, 2, 3, ...
(if a given rock were to fall then the previous rock would still stand)

Counterfactual Possibility $\neg(R > \neg R)$
(it is a counterfactual possibility that one of the rocks falls, i.e. a contradiction does not follow counterfactually from the supposition that a rock falls).

The following principles of reasoning are also very plausible:

Entailment $$\frac{}{C > C'}$$ given that C′ is a logical consequence of C

Substitution $$\frac{A > C}{A' > C}$$ given that A and A′ are logically equivalent

Transitivity $$\frac{A > B \quad A \wedge B > C}{A > C}$$

Disjunction $$\frac{A > C \quad B > C}{A \vee B > C}$$ given that A and B are logically exclusive

Conjunction $$\frac{A > C_1 \quad A > C_2 \quad A > C_3 \ldots}{A > C_1 \wedge C_2 \wedge C_3 \wedge ....}$$

From these assumptions we may, somewhat surprisingly, derive a contradiction. Details are left to the appendix but we may note here the broad outline of the proof. We show from the assumptions that if one of the rocks were to fall then it would not be the first and hence would be one that was second on. In the same way, we show that if one of the dominos from the second on were to fall, it would not be the second and hence would be one that was third on, from which it follows that if one of the dominos were to fall it would be one that was third on. Continuing in this way, we may show that if one of the dominos were to fall then it would not be any one of them, contrary to its being a counterfactual possibility that one of the dominos falls.

One of the assumptions should therefore be given up. But which?

There are three non-logical assumptions - Positive Effect, Negative Effect, and Counterfactual Possibility. Positive Effect and Counterfactual Possibility appear to be undeniable. But there are two qualms one might have about Negative Effect. The first is that it is not altogether clear that the counterfactual ‘if the second rock were to fall then the first would not fall' is true, for in considering the counterfactual possibility that the second rock falls, one should perhaps not keep fixed the fact that the first rock does not fall but allow that the second might fall by way of the first falling.

The second misgiving has to do with the lack of a connection between the antecedent and the consequent in Negative Effect. The second rock's falling, were it to happen, would not in any way be responsible for the first rock's not falling. But it might be thought that the truth of a counterfactual requires that the truth of the antecedent should somehow be responsible for - or connected to - the truth of the consequent.

I myself do not find either of these objections convincing. But rather than attempting to meet them head on, let me simply give a more complicated version of the case for which they do not arise (we might call it 'Goodman ad Infinitum' since it depends upon indefinitely duplicating Goodman's famous example of the match). Imagine an infinity of matches m_1, m_2, m_3 ..., each in different, causally isolated, space-time regions of the universe. We suppose that each match is dry, that there is plenty of oxygen in the atmosphere surrounding the match and that, in general, the conditions for a struck match to light are as propitious as they could be. Use S_n for 'match n is struck', W_n for 'match n is wet', and L_n for 'match n lights'. Let S be $S_1 \wedge S_2 \wedge S_3 \wedge \dots$ (each match is struck) and use:

$$M_1 \text{ for } S \wedge (W_1 \wedge \neg L_1) \wedge (W_2 \wedge \neg L_2) \wedge (W_3 \wedge \neg L_3) \wedge \dots$$
$$M_2 \text{ for } S \wedge (W_2 \wedge \neg L_2) \wedge (W_3 \wedge \neg L_3) \wedge (W_4 \wedge \neg L_4) \wedge \dots$$
$$\vdots$$
$$M_n \text{ for } S \wedge (W_n \wedge \neg L_n) \wedge (W_{n+1} \wedge \neg L_{n+1}) \wedge \dots \wedge (W_{n+2} \wedge \neg L_{n+2}) \wedge \dots$$
$$\vdots$$

Thus M_1 say that all of the matches are struck but are wet and do not light, M_2 says that all of the matches are struck and that all from the second on are wet and do not light, and M_n, in general, says that all of the matches are struck and that all from the n-th on are wet and do not light.

Let us also suppose that no match is in fact struck. It can then be argued that the sentences M_1, M_2, M_3, ... (in place of R_1, R_2, R_3, ...) will conform to the non-logical assumptions above:

Positive Effect M_{n+1} is a logical consequence of M_n (since it results from removing some of the conjuncts from M_n); and so $M_n > M_{n+1}$ by Entailment. Thus in this case there is no need for a special non-logical assumption; Positive Effect is guaranteed by the logic of counterfactuals alone.

Negative Effect It may surely be granted that $S_1 > L_1$ (if the first match were struck it would light) is true. But S_2, S_3, ..., W_2, $\neg L_2$,W_3 ,$\neg L_3$, ... are entirely irrelevant to L_1 being a counterfactual consequence of S_1, i.e. to whether the first match would light if struck, since they concern what happens in causally isolated regions of the universe; and so the counterfactual $S_1 \wedge S_2 \wedge S_3 \wedge \dots \wedge (W_2 \wedge \neg L_2) \wedge (W_3 \wedge \neg L_3) \wedge \dots > L_1$ (i.e., $M_2 > L_2$) should also be true. But $\neg M_1$ is a logical consequence of L_1 (since $\neg L_1$ is one of the conjuncts of M_1); and so $M_2 > \neg M_1$ should also be true. A similar argument establishes $M_{n+1} > \neg M_n$ for any n.

Counterfactual Possibility There appears to be nothing incoherent about the counterfactual supposition that all of the matches are struck but that all from some point on are wet and do not light.

One might worry that what went on in the other regions could be relevant in a non-causal way to what goes on in the given region. Thus, given that conditions for the match lighting were not propitious in all but finitely many of the specified regions, one might think that they would not then be propitious in the given region. But we may easily take care of this worry by making it part of the counterfactual supposition that there are already are infinitely many regions which are propitious for the lighting of a match. More generally, there is no reason for there to be any uniformity in what goes on in the different regions; it could involve the striking of a match in one region, the falling of a rock in another, the turning on of a light in yet another, and so on. Thus it would not even be clear how what went on in the other regions could be either causally or non-causally relevant to to what went on in the given region.

By modifying the original example in this way, we see that the prospects for disputing one of the non-logical assumptions do not look at all good. But what of the logical assumptions? It might be thought that here there is an obvious way out. For Lewis[2] has argued that the so-called 'Limit Assumption' might fail, with worlds getting closer and closer to the actual world without end. His example is of a line that is in fact under 1" in length. Consider now the worlds in which the line is longer than 1 inch. Then for any such world there will be a closer world in which the difference from one inch is reduced by a half.

Suppose now we adopt Lewis' version of the possible worlds semantics in which a non-vacuous counterfactual A > C is taken to be true if all of the sufficiently close worlds in which A are true are ones in which C is true. Then the failure of the Limit Assumption will lead to a failure in the infinitary version of the Conjunction Rule.[3] In the case at hand, for example, it will be true, under the counterfactual supposition that the line is longer than an inch, that it would be at most ½ inch in length, that it would be at most 1/4 inch in length, and so on ad infinitum, but it would not be true under the same counterfactual supposition that it would be at most ½ inch in length, at most 1/4 inch in length, and so on ad infinitum (since it would not then be longer than 1 inch).

But the counterfactual judgements upon which the purported counter-example to the Rule depend receive no intuitive support. We have no inclination to say, under the counterfactual supposition that the line is longer than an inch, that it would be at most ½ inch in length, that it would be at most 1/4 inch in length, and so on ad infinitum. Thus the correct conclusion to draw from the case is not 'so much the worse for the Conjunction Rule', but 'so much the worse for the Lewis semantics or his rejection of the Limit Assumption'.

Indeed, it is not clear how there could even be a convincing counter-example to the Rule. For the finitary version of the Rule is as obvious as any rule could be and yet, as Pollock has pointed out, the infinitary version of the rule appears to be 'just as obvious' as the finitary version and valid 'for exactly the same reason'.[4]

Moreover, taking this way out of the puzzle would render counterfactuals worthless for two of the central purposes for which they are required. For if we look at the derivation of a contradiction from our assumptions, we see that it does not require the full version of the infinitary rule but only the special case in which the counterfactual consequences are logically inconsistent with the counterfactual supposition[5]:

Infinitary Consistency

$$\frac{A > C_1 \quad A > C_2 \quad A > C_3 \ldots}{A > C_1 \wedge C_2 \wedge C_3 \wedge \ldots.} \quad \text{where } C_1, C_2, C_3, \ldots \text{ are jointly inconsistent}$$

This means that if the relevant application of the rule is to be challenged, then it must be allowed that a statement A might be a counterfactual possibility (since $A > C_1 \wedge C_2 \wedge C_3 \wedge \ldots.$ is not true) even though the counterfactual consequences of A are jointly inconsistent. Call a counterfactual supposition of this sort *paradoxical.*

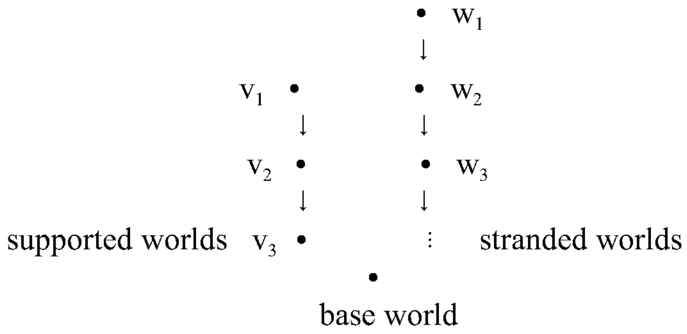
The trouble with paradoxical counterfactual suppositions is that they are of no use for decision making or theory testing. Suppose I am deciding between bringing it about that A or bringing it about that B. To this end, I consider the counterfactual consequences of bringing about the one or the other and then make my decision on the basis of a comparison of the consequences. But what if one or both of the counterfactual suppositions that I bring about A and that I bring about B are paradoxical? How then can I compare them, given that I can form no coherent conception of what the consequences of one or both of them are? Perhaps one is a matter of giving you some pain and the other a matter of giving you some pleasure. One feels that one should be able to decide in favor of giving you some pleasure (other things being equal). But how is this possible, on the present view, should one of the alternatives involve giving you some pleasure, though not any specific amount of pleasure - not 1utile, ½ a utile, 1/4 utile etc., while the other involves giving you some pain, though not any specific amount of pain - not 1 disutile, ½ a disutile, 1/4 a disutile etc?

Or again, suppose I wish to test a theory. To this end, I take a particular counterfactual possibility A and consider what consequences would follow, according to the theory, if A were to obtain. The theory is then disconfirmed if one or more of these consequences fail to hold and is otherwise confirmed. Perhaps the theory predicts what would happen if the temperature were to go up. One feels that one should be able to test the theory on the basis of what it would predict under this supposition. But what if the supposition is paradoxical? The temperature would go up, but not by at least 4C°, or by at least 2C°, or by at least 1C° etc. Then one of these counterfactual consequences must inevitably fail to hold (and, indeed, no actual testing of the theory would be required to see that this was so!). And yet we do not want to reject a theory simply because it tolerated paradoxical counterfactual suppositions, which might after all have been true.

Indeed, not only will these suppositions be of no use when they should be of use, they will also get in the way of making the counterfactual judgments that *can* properly be taken into account. For we do not want to find ourselves deciding between two alternatives, when one of them is paradoxical, or testing a theory on the basis of a paradoxical supposition. But this means that prior to making a decision or testing a theory, we will need to settle the question of whether the suppositions in question are paradoxical; and this is, in general, no easy task. Thus decision making and theory testing will become encumbered by the tricky preparatory exercise of assessing whether the counterfactual supposition in question can properly be made.

Lewis' version of the possible worlds semantics (in terms of sufficiently close worlds) is intended to accommodate failures of the Limit Assumption. But it turns out that there is another version of the possible worlds semantics which is also able to accommodate such failures and which in some other respects is to preferred.

To explain this alternative, let us introduce some terminology. Say that a member of a set of worlds is *supported* if there is a closest world of the set that is as least as close (to the given base world) as it is and otherwise say it is *stranded.* A supported world, when it looks 'down', can see a closest world to the base world but a stranded world can only see worlds that get closer and closer without end to the base world:



Say that a set of worlds is itself *supported* if each of its member worlds is supported and that otherwise it is *stranded.*

If the set of A-worlds is supported, then there would appear to be no reasonable alternative to the standard semantics: the counterfactual A > C will be true iff C is true in all of the closest A-worlds.[6] But it is not so clear what we should say when the set of A-worlds is stranded. One *might* adopt Lewis' proposal. But an alternative is to take A > C to be true iff C is true in all of the closest and all of the stranded A-worlds. If we cannot get as close as possible to the base world, so to speak, then we do not even try.

We can bring out the difference between the two proposals if we consider a chain of worlds w_1,w_2, w_3, ... that get closer and closer to the base world, as in the illustration above. Suppose that these worlds are exactly the A-worlds. Then under Lewis' proposal, the counterfactual A > C will be true iff C is true in all but finitely many of the worlds. But under the alternative proposal, the counterfactual A > C will be true iff C is true in all of the worlds, since all of them are stranded.

It turns out that the alternative semantics will validate all of the logical rules (including Conjunction) with the single exception of Disjunction. For suppose, in the above example, that A_1 is true in the worlds w_1,w_2, that A_2 is true in the worlds w_3,w_4, ..., and that C is true in all worlds but w_1. Then A_1 > C and A_2 > C are both true while $A_1 \vee A_2$ > C is not, since all of the worlds w_1,w_2, ... in which $A_1 \vee A_2$ are true are now stranded.[7]

But this alternative also comes at great cost. For we will have to give up Disjunction, which we have seen no independent reason to question. And we will have to give up many of our intuitive judgements concerning the truth of particular counterfactuals.

Consider the rock case, for example. Then it may be shown, given Positive and Negative Effect, that for every R_1-world there will be a closer R_2-world, for every R_2-world a closer R_3-

world, and so on.[8] It follows that each R-world is stranded. For any R world will be a R_k-world for some k and so, by Positive and Negative Effect, there will be a closer R_{k+1}-world and hence a closer R-world.

Given that each R-world is stranded, an arbitrary statement C will be a counterfactual consequence of R under the alternative semantics only if it is true in *all* of the R-worlds, i.e. only if it is entailed by R. More generally, let us say that the statement A is a *trivial* counterfactual supposition if A > C is true only in the 'trivial' case in which A entails C. Then under the alternative semantics, A will be a trivial counterfactual supposition if (and only if) each A-world is either a closest A-world or stranded.

But surely R is *not* a trivial counterfactual supposition. For a counterfactual consequence of $R = R_1 \vee R_2 \vee R_3 \vee ...$ is $R' = R_2 \vee R_3 \vee R_4 ...$; if one of the rocks from the first on were to fall then one of the rocks from the second on would fall. Yet R' is not entailed by R, since if only the first rock falls then R will be true and R' false.

We see that Lewis' proposal (in terms of sufficiently close worlds) and our own (in terms of closest or stranded worlds) are both unsatisfactory, though in somewhat different ways. Lewis' proposal fails to validate the intuitively acceptable principle of Conjunction while ours fails to validate the intuitively acceptable principle of Disjunction. And both yield a wide range of unacceptable counterfactual judgements. In the case of the rock, for example, it will be a counterfactual consequence of the supposition that a rock falls, under Lewis' proposal, that the first rock does not fall, that the second rock does not fall, and so on ad infinitum, while it will *not* be a counterfactual consequence of the supposition, under our own proposal, that a rock from second on will fall. In general, a possible supposition true only in stranded worlds will be paradoxical for Lewis - yielding more than it should, while it will be trivial for us - in general yielding less than it should.

Perhaps our own proposal is less objectionable than Lewis'. For: (i) Disjunction is perhaps less intuitively compelling than Conjunction; and (ii) unwanted triviality is perhaps less bothersome than unwanted paradoxicality and less of an impediment to the use of counterfactuals in decision making and theory testing.[9] But neither proposal can be considered acceptable.

Is there some other way out? I should like to suggest that it is the Rule of Substitution (permitting the substitution of logically equivalent antecedents) that should be given up. Indeed, this rule is the only one whose application in the reasoning of the puzzle is subject to serious doubt. For in order to derive the intermediate conclusion $[R_1 \vee (\neg R_1 \wedge R_2)] > \neg R_1$ (step (i) from the appendix), we have to make the inference from $R_2 > \neg R_1$ to $[(R_1 \wedge R_2) \vee (\neg R_1 \wedge R_2)] > \neg R_1$; and yet this inference does not appear to be valid. For granted that the first rock would not fall if the second rock were to fall, it hardly seems correct to say that the first rock would not fall if either the first and the second were to fall or the second but not the first were to fall. Thus the one and only obvious way to avoid the puzzle is to reject this step and the more general principle of Substitution upon which it rests.

If asked why the critical step from $R_2 > \neg R_1$ to $[(R_1 \wedge R_2) \vee (\neg R_1 \wedge R_2)] > \neg R_1$ appears to be invalid, it is natural to point to a consequence of the latter statement not had by the former. For the latter statement appears to imply $R_1 \wedge R_2 > \neg R_1$ (that the first rock would not fall if the first and second rocks were to fall), which certainly does not follow from $R_2 > \neg R_1$ (that the first rock would not fall if the second rock were to fall). And, in general, it might be thought that we should accept a kind of converse to Disjunction:

<u>Simplification</u>

$$\frac{A \vee B > C}{A\ (B)\ > C}$$

according to which the counterfactual from $A \vee B$ to C also licenses the counterfactual from A to C or from B to C.

Simplification is not not of course valid under the standard possible worlds semantics (even assuming a single closest world). For if the closest $A \vee B$-world is an $A \wedge C$-world, then $A \vee B$ will counterfactually imply C and yet B will not counterfactually imply C given that the closest B-world is not also C-world. Moreover, given Substitution, Simplification will license the inference from $A > C$ to $A \wedge B > C$[10], which is generally taken to be invalid. But however convincing these considerations may be for someone who already accepts the standard semantics or the Rule of Substitution, they have no force in the present context, where the semantics and the Rule are themselves in question.

Some philosophers have attempted to provide arguments against Simplification that are independent of the acceptance of Substitution or the standard semantics.[11] But most of these arguments are essentially defensive in character. Loewer[12], for example, thinks that there is no mismatch in logical form between the formal and the ordinary language version of Simplification and he accepts the standard possible worlds semantics under which the formal argument would not be valid. He therefore faces the problem of explaining why the argument appears to be valid; and to this end, he provides a pragmatic explanation of how one might gather the truth of its conclusion from the assertion of its premise. Thus such a line of reasoning serves merely to defend his view against potential counter-example. But there is nothing in it to suggest that someone else would be mistaken if he took the conclusion of the ordinary language version of Simplification to be a semantic, rather than a pragmatic, consequence of its premise.

There is, however, one case in which the evidence appears to point in the other direction. For certain 'excluder' counterfactuals appear to be straightforward counter-examples to the Rule. To use an example from McKay & van Inwagen[13], the counterfactual:

(1) If Spain had fought with the Axis or the Allies, she would have fought with the Axis does not appear to imply:

(2) If Spain had fought with the Allies she would have fought with the Axis,
as it should if Simplification were valid.

In the face of this apparent counter-example, I would like to suggest that Simplification *is* generally valid and that the counterfactual conclusion in the above argument does indeed follow from the counterfactual premise. What accounts for the appearance of invalidity is the operation of a principle of 'Suppositional Accommodation', according to which we always attempt to interpret a counterfactual in such a way that its antecedent A represents a genuine counterfactual possibility. It is on account of this principle that we are usually disinclined to accept any counterfactual of the form $A > \neg A$, since that would require treating A as something that was not a counterfactual possibility.

What now happens in the above argument is this. In asserting (or supposing) (1), we are presupposing that it is not a genuine possibility that Spain fought with the Allies. Indeed, the premise is a way of indicating that this is so; for the counterfactual premise 'if Spain had fought with the Allies or the Axis then she would have fought with the Axis' could not have been true if it had been taken to be a genuine possibility that Spain fought with the Allies. But when we move to (2), the principle of Suppositional Accommodation requires that we treat Spain fighting with the Allies as a genuine possibility; and given this accommodation in what is taken to be possible, (2) is no longer true. Thus what accounts for the appearance of invalidity is a shift in the relevant 'space of possibilities' as we move from premise to conclusion.

The principal objection to accepting Simplification and giving up Substitution is that it is then no longer clear how the logic or semantics for counterfactuals should go. Substitution does not generally fail and so when does it hold?[14] And if the truth-value of a counterfactual does not merely depend upon the sets of worlds in which its antecedent and consequent are true, then on what else does it depend?

I should now like to sketch an alternative logic and semantics for counterfactuals. The basic idea behind the semantics is to evaluate the antecedents of counterfactuals at possible states rather than at possible worlds. Although this might appear to be a minor departure, it yields a simple and natural solution to the puzzles and enjoys many other significant advantages over the possible worlds semantics.

It will be helpful to introduce this alternative approach by way of a comparison with the possible worlds approach. Under the possible worlds semantics, we are given a 'pluriverse' of possible worlds; and arbitrary statements are then taken to be true or false at each possible world. Truth-functional statements are subject to the familiar clauses. Thus $\neg$A will be true at a world if A is not true at the world; A & B will be true at a world iff A and B are true at the world; and A $\vee$ B will be true at a world iff A or B is true at the world.[15]

Under our alternative 'truth-maker' semantics, the pluriverse of possible worlds is replaced with a space of possible states - the monolithic blobs shatter into myriad fragments. Thus not only will will there be a possible world in which I am sitting, you are standing, we are talking, etc. etc., there will also be a possible state in which I am sitting, a possible state in which you are standing, a possible state in which I am sitting and you are standing, and so on.

We take the space of states to be endowed with mereological structure. Thus the state of my sitting and being asleep will be composed of the state of my sitting and the state of my being asleep, and the state of a given patch being red all over will have no part in common with the state of the patch being green all over (unless it concern the existence or constitution of the patch, without regard to its color).

We take the view that the fusion of the some states will exist just in case the fused states are compatible, i.e. just in case it is possible for all of them to obtain. Thus the fusion of the state of my sitting and the state of my being asleep will exist since the two component states are (all too) compatible. On the other hand, the fusion of the state of the patch being red all over and the state of the patch being green all over will not exist, since no patch can be both red all over and green all over.

The fusion of states will be what mathematicians' call a *least upper bound*: it is an *upper* *bound* in that each of the states is a part of the fusion; and it is the *least* upper bound in that it is a part of every other upper bound. For some states to be compatible is for them to have an upper bound. We are therefore led to the following important principle:

CLUB (Conditional Least Upper Bound) any states will have a least upper bound as long as they have an upper bound.

From a formal point of view, we might take a state space to be a set S endowed with a part-whole structure $\sqsubseteq$, where $\sqsubseteq$ is a relation on S subject to reflexivity, antisymmetry, transitivity and CLUB. It is a state space, in this sense, that plays the same role within the truthmaker semantics as is played by the pluriverse within the possible worlds semantics.

Statements are true or false *at* possible worlds but are *made* true or false - i.e., are *verified* or *falsified* - by possible states. Whereas a statement is either true or false at any given possible world, it may not be made true or false by any given possible state. The possible state of my sitting for example does not settle the question of whether I am asleep. For this reason, we cannot give the usual classical clauses for when a statement is verified or falsified by a possible state.

There are a number of different proposals as to how the classical clauses should be modified in the presence of partiality, but the guiding principle behind our own account is that the state should be wholly relevant to the truth (or falsity) of the statement that it verifies (or falsifies). We are thereby led to the following critical clauses (where $s \sqcup t$ is used for the fusion of the states s and t):

(i)$^+$ state s verifies $\neg A$ iff s falsifies A;

(i)$^-$ s falsifies $\neg A$ iff s verifies A;

(ii)$^+$ s verifies $A \wedge B$ iff s is the fusion $s_1 \sqcup s_2$ of a state s_1 that verifies A and a state s_2 that verifies B;

(ii)$^-$ s falsifies $A \wedge B$ iff s falsifies A or s falsifies B or s falsifies $A \vee B$;

(iii)$^+$ s verifies $A \vee B$ iff s verifies A or s verifies B or s verifies $A \wedge B$;

(iii)$^-$ s falsifies $A \vee B$ iff s is the fusion $s_1 \sqcup s_2$ of a state s_1 that falsifies A and a state s_2 that falsifies B.[16]

According to clause (ii)$^+$, a state that verifies a conjunction should be composed of states that verify the respective conjuncts and likewise, according to (iii)$^-$, a state that falsifies a disjunction should be composed of states that falsify the respective disjuncts. We should note that a verifier of $A \wedge B$ is also allowed to be a verifier of $A \vee B$ and that a falsifier of $A \vee B$ is also allowed to be a falsifier of $A \wedge B$, although we might adopt a more exclusive version of the clauses under which these cases are not allowed).

Whereas the possible worlds semantics merely tells us whether a statement is true or false at a possible world, the present semantics tells us what it is *in* the world that makes the statement true if it is true or what it is in the world that makes it false if it is false. The difference between the two approaches comes out very clearly with instances of the Law of Excluded Middle - such as ‘it is raining or not raining’ versus ‘it is windy or not windy’. These two statements are true in the same possible worlds, viz. all of them, but what makes them true in those worlds is very different, the state of the rain in the one case and the state of the wind in the other.

Verification and falsification as defined above are a form of *exact* verification or falsification; and there is no guarantee, if a state s verifies a statement, that a more comprehensive state s ⊔ t will also verify the statement. In terms of exact verification and falsification - which we designate by ‘s ||- A’ and ‘s -|| A’ - we may define two looser notions of verification and falsification. A state *inexactly* verifies a statement (which we write as ‘s ||> A’) if it contains a state that exactly verifiers the statement. Thus the state of its being windy and wet will not exactly verify that it is windy since it contains the irrelevant state of its being wet, but it will inexactly verify that it is windy since it contains the relevant state of its being windy, one which exactly verifies the statement that it is windy; and similarly for falsification.

A weaker notion still is that of *loose* or *classical* verification or falsification. A state loosely verifies a statement (which we write as ‘s |= A’) if it incompatible with any falsifier of the statement; and similarly, a state loosely falsifies a statement if it is incompatible with any verifier of the statement. A loose verifier will require that the statement be true, since it excludes the possibility of a falsifier, but it may not itself either be, or contain, a verifier for the statement. The state of its being wet, for example, will loosely verify the statement that is windy or not windy, since no state whatever will falsify the statement, but the state does not verify or contain anything that verifies the statement.

Corresponding to the three notions of verification are three notions of consequence. Thus we may say that C is an *exact* consequence of the statement A - or that A *exactly* entails C - if any state that exactly verifies A exactly verifies C; and similarly for the notions of *inexact* and *classical* consequence but with inexact or classical verification in place of exact verification. Note that A ∨ B will be an exact consequence of A, since any state that verifies A will also verify A ∨ B. But A will not in general be an exact consequence of A ∧ B, since the B-part of the state that verifies A ∧ B may not be relevant to the verification of A. Thus even though there is no logical distinction, from the classical perspective, between inferring a disjunction from a disjunct and inferring a conjunct from a conjunction, the two cases are fundamentally different from the present perspective. Or again, B ∨ ¬B will be a classical consequence of A (since nothing falsifies B ∨ ¬B) but it will not in general be an inexact consequence (since an exact verifier for A need not contain an exact verifier or falsifier for B).

Let us now show how to extend the above semantic framework to counterfactuals; and let us begin with the relatively modest task of saying when a counterfactual statement is true or false at a given world (or world-state).[17] Our account of the truth-conditions for counterfactuals will be based upon two main ideas. The first is the *Universal Realizability of the Antecedent*. What this means is that a counterfactual A > C will only be taken to be true when it is true for *any way* in which its antecedent A might be true. The second idea is the *Universal Verifiability of the Consequent*. What this means is that a counterfactual will be taken to be true, given some way in which its antecedent might be true, only when its consequent is made true under *any outcome* of the way in which its antecedent is true.

We capture the idea of the different ways in which the antecedent A might be true in terms of the different states that exactly verify it; and we capture the idea of the consequent being made true under any outcome of the way in which the antecedent is true in terms of its being inexactly verified by any outcome of an exactly verifying state for the antecedent. We therefore arrive at the following truth-conditions:

The counterfactual A > C is true if any possible outcome of an A-state contains a C-state.

It may help to be a little more precise, making explicit the reference to the world at which the counterfactual is to be evaluated. Let us use '$t \rightarrow_w u$' to indicate that u is a possible outcome of imposing the change t on the world w:

$$\begin{tikzcd} & & u \\ t \arrow[r] & \boxed{w} \arrow[ur] \arrow[r] \arrow[dr] & u' \\ & & u'' \end{tikzcd}$$

We then declare:

A > C is true at w iff u inexactly verifies C whenever t exactly verifies A and u is a possible outcome of u relative to w.

Or in symbolic terms:

w |= A > C if u ||> C whenever t ||- A and $t \rightarrow_w u$.

Note that two notions of verification are employed on the right. The antecedent state t must *exactly* verify the antecedent A, since it should correspond to a way - i.e. to an exact way - in which the antecedent is true. But the outcome u need only *inexactly* verify the consequent C, since we do not wish to insist that the outcome should itself be a way in which the consequent is true but merely that it should embody a way in which the consequent is true.

The notion of classical verification (|=) is employed on the left though, since classical and inexact verification coincide for world-states, we could equally well have employed the notion of inexact verification. Given that |= is used on the left and ||- (or ||>) on the right, we cannot use the above clause 'recursively' to evaluate counterfactuals A > C whose antecedent or consequent also contain counterfactuals. The semantics can in fact be extended to embedded counterfactuals, but this is not something we shall consider in the present paper.

The informal terminology of 'outcomes' should not mislead. An outcome is naturally taken to be a future causal outcome. But such a narrow interpretation is not required. In the case of backtracking counterfactuals, for example, the relation → could be taken to be a backwards arrow, relating the given state to the states that would have had to obtain for it to obtain; and in other cases (such as 'if this peg had been round, then it would not have fit the hole'), the relation could be taken to be more logical or conceptual in character.

Let me make some remarks about the logic of counterfactuals under the present semantics. There are different notions of validity in play and we shall take an inference:

$$\frac{A_1 \quad A_2 \quad A_3 \ \ldots}{C}$$

to be valid if it is classically valid, i.e. if the conclusion is classically entailed by the premises.

We first state some rules that are valid simply from the form of the truth-conditions without regard to any conditions that might be imposed on the transition relation →; and to this end, we use $\top$ as a truth constant that is verified by the 'null' state 0 alone yet never falsified and we use $\bot$ for its negation, $\neg\top$. The following rules are then valid[18]:

Inclusive Disjunction
$$\frac{A > C \quad B > C \; (A \wedge B) > C}{(A \vee B) > C}$$

Conjunction
$$\frac{A > B \quad A > C}{A > B \wedge C}$$

Exact Strengthening
$$\frac{A > C}{A' > C} \qquad \text{given that } A' \text{ exactly entails } A$$

Inexact Weakening
$$\frac{A > C}{A > C'} \qquad \text{given that } C \text{ inexactly entails } C'$$

Triviality
$$\frac{}{A > \top} \qquad\qquad \frac{}{\bot > C}$$

The proofs of validity are straightforward but, given the unfamiliar nature of the semantics, we spell out the details:

Inclusive Disjunction In order to show that this rule is valid, we need to show that the conclusion $A \vee B > C$ of the rule will be true in an arbitrary world w given that its premises $A > C$, $B > C$, $(A \wedge B) > C$ are true at w.[19] So take any state t for which t ||-$A \vee B$ and suppose $t \rightarrow_w u$. We need to show u ||> C. By the verification clauses for $\vee$, either (i) t ||- A or (ii) t ||- B or (iii) t ||- $A \wedge B$. In case (i), it follows, given that $A > C$ is true at w, that u ||> C, as required; and similarly for the other cases.

Conjunction To establish validity, suppose t ||- A and $t \rightarrow_w u$. Then u ||> B and u ||- C, given the truth of $A > B$ and $A > C$ at w; and so for some sub-states u_1 and u_2 of u, u_1 ||- B and u_2 ||- C. By Club, $u_1 \sqcup u_2$ exists and is a sub-state of u. But then $u_1 \sqcup u_2$ ||- $B \wedge C$; and so u ||> $B \wedge C$, as required.

Exact Strengthening Suppose w |= $A > C$, $t \rightarrow_w u$, and t ||- A′. Then t ||- A given that A′ exactly implies A; and so u ||- C, as required.

Inexact Weakening Suppose that w |= $A > C$, t ||- A and $t \rightarrow_w u$. Then u ||> C. So u ||> C′ given that C′ is an inexact consequence of C.

Triviality Suppose t ||- A and $t \rightarrow_w u$. Then u contains the null state 0, which verifies $\top$. On the other hand, no state t exactly verifies $\bot$ and so, vacuously, u ||> C whenever t ||- $\bot$ and $t \rightarrow_w u$.

Some further comments on these rules may be helpful:

(1) Note the extra premise $(A \wedge B) > C$ in the rule of Inclusive Disjunction. This is

necessary because we have allowed a verifier for $A \wedge B$ also to be a verifier for $A \vee B$. If this option had been excluded, then the extra premise would not have been required.

(2) If infinitary conjunctions were allowed, then the infinitary version of Conjunction could be established in the same way as the finitary version.

(3) The rule of Simplification above is a special case of Exact Strengthening since $A \vee B$ is an exact consequence of A. Thus we see that Simplification is not an isolated phenomenon but merely a manifestation of a more general aspect of the logic. We also see that the counterfactual behaves like a standard conditional in allowing both strengthening of the antecedent and a corresponding weakening of the consequent. But the relevant forms of strengthening and weakening should be appropriately understood in terms of the 'exact' content of the statements upon which they operate.

There are some plausible conditions that might be imposed on the transition relation and that lead to the validity of some further rules. Let us discuss each of them in turn. We might assume, first of all, that the outcome of any state should include that state:

Inclusion If $t \rightarrow_w u$ then $t \sqsubseteq u$.

From Inclusion, we obtain the validity of:

Identity
$$\frac{}{A > A}$$

For take any world-state w. Suppose $t \Vdash A$ and $t \rightarrow_w u$. By Inclusion, $t \sqsubseteq u$; and so $u \,||\!> A$, as required.

We might also assume that one of the outcomes of an actual state should itself be an actual state:

Actuality If $t \sqsubseteq w$ then $t \rightarrow_w u$ for some $u \sqsubseteq w$.

From Actuality, we obtain the validity of:

Modus Ponens
$$\frac{A > B \quad A}{B}$$

For suppose $w \models A$. Then for some $t \sqsubseteq w$, $t \Vdash A$. By Preservation, $t \rightarrow_w u$ for some $u \sqsubseteq w$. Given $w \models A > B$, $u \,||\!> B$; and since $u \sqsubseteq w$, $w \models B$.

Another plausible assumption is that an outcome of a given state should remain an outcome of the state in combination with any part of the outcome. In other words, it should be possible to incorporate any part of the outcome into the given state without affecting its status as an outcome:

Incorporation If $t \rightarrow_w u$ and $u' \sqsubseteq u$ then $t \sqcup u' \rightarrow_w u$;

From Incorporation, we obtain the validity of:

Transitivity
$$\frac{A > B \quad A \wedge B > C}{A > C}$$

For suppose t ||- A and $t \to_w u$. We need to show u |= C. Given w ||- A > B, u ||> B. So for some $u' \sqsubseteq u$, u' ||- B; and consequently $t \sqcup u'$ ||- $A \wedge B$. By Incorporation, $t \sqcup u' \to_w u$; and so given w ||- $A \wedge B > C$, u ||> C, as required.

A final assumption, though somewhat less plausible than the others, is that each outcome of a given state should itself be a world-state:

Completeness $t \to_w u$ only if u is a world-state.[20]

Adopting this condition enables us to make a modest extension of the semantics to counterfactuals A > C whose consequent may also contain counterfactuals, since this only requires that we ascertain when thsese counterfactuals are *in*exactly verified by a given world-state.

From Completeness, we obtain the validity of:

Classical Weakening
$$\frac{A > C}{A > C'} \quad \text{given that } C' \text{ is a classical consequence of } C$$

For suppose that w |= A > C, t ||- A and $t \to_w u$. Then u ||> C; and so u |= C. Since C' is a classical consequence of C, u |= C. By Completeness, u is a world-state; and so u ||> C, as required.

We are now in a position to resolve the puzzle. I will not go into the formal details but, by adopting the present semantic framework, we can construct a model in which all of the non-logical assumptions of the puzzle are true and all of the logical rules, with the single exception of the offending instances of Substitution, are valid. Indeed, the model can be made to correspond to our intuitive conception of the scenario, in which the state of the n-th rock falling will have as its single outcome the state of all rocks from the n-th on falling. We thereby avoid the difficulties that beset the possible worlds approach and provide a completely satisfying semantical explanation of why and when the rule of Substitution will hold.

There are two objections that I have sometimes heard made against the present semantics in relation to the possible worlds semantics. It has been argued, in the first place, that it is relatively problematic in its ontological commitments. For not only must we presuppose a pluriverse of possible worlds, but also a space of possible states of which the worlds are composed. But it is not in general clear how a world is to be divided into different possible states. Do we have a state of a patch being red, say, or only of its being a particular shade of red or a state of Fido's being a dog, say, or only of his being a cocker spaniel? It has been argued, in the second place, that the present semantics is relatively problematic in its conceptual commitments. For the transition relation must itself be understood in terms of counterfactuals. Thus to say that u is a possible outcome of t in w is just to say that, in w, u might obtain if t were to obtain (and also that u is maximal in this respect). But under the possible worlds semantics, we have an *analysis* of the truth-conditions of counterfactuals in terms of the closeness or similarity of possible worlds.

It is not altogether clear that the possible worlds semantics has the advantages claimed for it under these two objections. For similarity of worlds is partly a matter of agreement in particular fact. But what is a particular fact? Any answer in effect presupposes a space of possible states. Again, it is well known that the truth-conditions for counterfactual cannot be understood in terms of the ordinary notion of similarity but only in terms of a suitably doctored notion. But it is then far from clear that the requisite 'doctoring' can itself be understood independently of counterfactual considerations. I doubt, in any case, that we should require a semantics to provide us with an analysis of the locutions with which it deals rather than with a perspicuous account of how the truth-conditions of the sentences containing the locutions are to be determined.

But it does not really matter whether the possible worlds semantics has these supposed advantages. For we can always piggy back a version of the present semantics on the ontological and conceptual resources of the possible worlds semantics. For we may take the atomic states of the state space to be the truth-sets of the atomic sentences of the language and their negations (i.e., those sets of worlds in which the atomic sentences are respectively true and false); and we can take u to be an outcome of t in w if u is one of the closest worlds to w in which t obtains. In this way, there is no need for the truthmaker semantics to go beyond the ontological or conceptual resources of the possible worlds semantics. I do not recommend this as an option, since I believe that there is something to be gained from having a richer and more varied conception of the state space and the transition relation; but it is always available to someone with these qualms.

I also believe that the present semantics has enormous advantages over the possible worlds semantics, quite apart from the puzzle. These are largely a matter of implementation. We are presented with a concrete scenario in which we wish to make certain kinds of counterfactual judgment (rocks falling as a result of other rocks falling, gases expanding under pressure etc.) and we wish to form a concrete picture of how the underlying semantics should proceed. Let me briefly discuss two cases of this sort.

The first involves causal modeling[21]. The values of some variables depend upon the values of others. Suppose, for example, that my car is stuck in the mud and two friendly neighbors will push on my command. We may then take there to be four variables ME, YOU, HIM, IT. ME takes the value 1 if I shout 'push' and otherwise takes the value 0; YOU takes the value 1 if you push and otherwise takes the value 0, and similarly for HIM; and IT takes the value 1 if the car moves and otherwise takes the value 0. We may suppose, moreover, that the variables YOU and HIM depend upon the variable ME and take the value 1 just in case ME takes the value 1 and that the variable IT depends upon the variables YOU and HIM and takes the value 1 just in case YOU and HIM both take the value 1.

Suppose that I did in fact shout 'push' and that, as a consequence, you and he both push and the car moves (this corresponds to the variables ME, YOU, HIM, and IT all taking the value 1). Suppose now, counterfactually, that you did not push (so that YOU takes the value 0). We then want to say that the car would not move (IT takes the value 0) and we get this result, under the so-called 'interventionist' semantics, by first breaking the dependence of YOU on ME and then recomputing the values of the dependent variables. Thus the value of HIM remains at 1 while IT takes the value 0, given that YOU takes the value 0 and HIM the value 1.

We would like to be able to extend this semantics to counterfactuals whose antecedents are arbitrary truth-functional compounds of assignment statements (as with 'YOU takes the value 0 or HIM takes the value 0'). But it hard to see how to do this under the possible world semantics (whether with a similarity relation of without), since we somehow need to associate the truth-set of the antecedent with an assignment of values to the variables or to a 'disjunctive' set of such assignments; and there appears, in general, to be no reasonable way in which this can be done.[22]

These difficulties disappear under the truth-maker approach. For the assignments of values to a given variable can be taken to be an atomic states (within an appropriate state space). Fusions of such states will correspond to assignments of values to the variables; and truth-functionally complex antecedents will be verified by such fusions - that is, in effect, by assignments of values to the variables. But the transition relation can then tell us what the outcome of imposing such an assignment on a given world will be. Thus by taking the verifiers of an antecedent to be states rather than worlds, the difficulties in extending the interventionist semantics to truth-functional antecedents are avoided.[23]

The second case concerns the concept of 'inertial' change. Consider a 'block' universe:

```
       __
c    |__|
b    |__|            __
a  _|__|_________ |__|_ d
    ||                 ||  T
```

in which block c is on b, block b on a, and blocks a and d are on the table, T. Let the atomic states of our state space consist of one block being on top of another or being on the table. Consider now the non-actual state in which a is on c. This is singly compatible with the actual state of b being on a and also with the actual state of c being on b. But it is not jointly compatible with these two states. Consider, by contrast, a non-actual state in which block c is on d. This is compatible with the following actual atomic states: b on a, a on T, and d on T. But in this case, it is jointly compatible with each of these states and, more generally, is jointly compatible with all of the actual states with which it is singly compatible.

Call a state s of this sort *inertial* (or an *inertial change* if s is non-actual); and call a supposition A *inertial* if each of its verifiers is inertial. Then inertial changes and suppositions are, I believe, of very special significance. For it is relatively easy to determine the counterfactual consequences of an inertial supposition A. For as long as the verifier (or verifiers) of a given truth are compatible with the verifiers of A, we can safely assume that the truth will still obtain under the supposition of A. Thus given the truth of B and its compatibility with A, B will still be true under the supposition of A.

We might call this rule 'the Principle of Inertia'; and it may be argued that the inferential efficiencies afforded by this principle are of great help in engineering a solution to the frame problem from AI. I do not wish to explore this idea here, but let us note that any reasonable way of defining the concept of an inertial change will require reference to a state space and that any reasonable way of formulating the Principle of Inertia will require appeal to something like the truthmaker semantics, since the required compatibility of B with A is a compatibility in their verifiers and not merely a modal matter of there being a possible world in which each is true. Thus the state-based approach is again able to prove its worth against the traditional world-based approach.[24]

Appendix

We give the derivation of the contradiction from the stated assumptions. To this end, it will be helpful to make use of the following derived rules:

<u>Transitivity′</u> $$\frac{A > B \quad B > C}{A > C}$$ (as long as A is a logical consequence of B)

<u>Weakening</u> $$\frac{A > C}{A > C'}$$ (as long as C′ is a logical consequence of C)

<u>Proof of Transitivity′</u> Suppose that A is a logical consequence of B. Then B and $A \wedge B$ are logically equivalent. Given $A > B \quad B > C$, we obtain $A \wedge B > C$ by Substitution and so, from $A > B$ and $A \wedge B > C$, we obtain $A > C$ by Transitivity.
<u>Proof of Weakening</u> Suppose C′ is a logical consequence of C. Then C′ is also a logical consequence of $A \wedge C$. By Entailment, $A \wedge C > C'$ is a theorem. So given $A > C$, we obtain $A > C'$ by Transitivity.

(i) In order to derive a contradiction, we first show that $[R_1 \vee (\neg R_1 \wedge R_2)] > \neg R_1$ is derivable:

(1) $R_2 > \neg R_1$ Negative Effect
(2) $[(R_1 \wedge R_2) \vee (\neg R_1 \wedge R_2)] > \neg R_1$ from (1) by Substitution
(3) $R_1 > R_2$ Positive Effect
(4) $R_1 > (R_1 \wedge R_2)$ from (3) by Entailment $\wedge$ Finite Conj.
(5) $R_1 > [(R_1 \wedge R_2) \vee (\neg R_1 \wedge R_2)]$ from (4) by Weakening
(6) $\neg R_1 \wedge R_2 > [(R_1 \wedge R_2) \vee (\neg R_1 \wedge R_2)]$ by Entailment
(7) $[R_1 \vee (\neg R_1 \wedge R_2)] > [(R_1 \wedge R_2) \vee (\neg R_1 \wedge R_2)]$ from (5) and (6) by Disjunction
(8) $[R_1 \vee (\neg R_1 \wedge R_2)] > \neg R_1$ from (7) and (2) by Transitivity′

(ii) Next we show that:

(1:¬1) $R_1 \vee R_2 \vee R_3 \vee ... > \neg R_1$

is derivable. $[R_1 \vee (\neg R_1 \wedge R_2)] > \neg R_1$ is derivable by (i) above and $[\neg R_1 \wedge \neg R_2 \wedge (R_3 \vee R_4 \vee ...)] > \neg R_1$ is derivable by Entailment. So $R_1 \vee (\neg R_1 \wedge R_2) \vee [\neg R_1 \wedge \neg R_2 \wedge (R_3 \vee R_4 \vee ...)] > \neg R_1$ is derivable by Disjunction. But $R_1 \vee (\neg R_1 \wedge R_2) \vee [\neg R_1 \wedge \neg R_2 \wedge (R_3 \vee R_4 \vee ...)]$ is logically equivalent to $R_1 \vee R_2 \vee R_3 \vee ....$ So $R_1 \vee R_2 \vee R_3 \vee ... > \neg R_1$ is derivable by Substitution.

(iii) We are now in a position to derive a contradiction. From the derivability of (1:¬1) above, it follows by Entailment and Conjunction that $R_1 \vee R_2 \vee R_3 \vee .... > \neg R_1 \wedge (R_1 \vee R_2 \vee R_3 \vee ...)$ is derivable. But $R_2 \vee R_3 \vee ...$ is a logical consequence of $\neg R_1 \wedge (R_1 \vee R_2 \vee R_3 \vee ...)$; and so by Weakening:

(1:2) $R_1 \vee R_2 \vee R_3 \vee ... > R_2 \vee R_3 \vee ...$

is derivable.
In exactly the same manner in which we derived (1:¬1), we can also derive:

(2:¬2) $R_2 \vee R_3 \vee R_4 \vee ... > \neg R_2$.

But $R_1 \vee R_2 \vee R_3 \vee ...$ is a logical consequence of $R_2 \vee R_3 \vee R_4 \vee ...$; and so by Transitivity′ applied to (1:2) and (2:¬2):

(1: ¬2) $R_1 \vee R_2 \vee R_3 \vee ... > \neg R_2$

is derivable.

Proceeding in this manner, we establish the derivability of:

(1: ¬n) $R_1 \vee R_2 \vee R_3 \vee ... > \neg R_n$, for n = 1, 2, ...

So by Conjunction, $R_1 \vee R_2 \vee R_3 \vee ... > \neg R_1 \wedge \neg R_2 \wedge \neg R_3 \wedge ...$ is derivable; and hence $R_1 \vee R_2 \vee R_3 \vee ... > \neg(R_1 \vee R_2 \vee R_3 \vee ...)$ is derivable by Weakening, contrary to Possibility.

[1] Stalnaker, 'A Theory of Conditionals' in N. Rescher (ed.) *Studies in Logical Theory*, American Philosophical Quarterly Monograph Series , No. 2' (Oxford: Blackwell, 1968), 98-112 and Lewis, *Counterfactuals* (Oxford: Blackwell, 1973).

[2] Lewis (1973) ibid., p. 20. A related argument is to be found in Lewis, 'Ordering Semantics and Premise Semantics for Counterfactuals', Journal of Philosophical Logic 10 (2) (1981, 217-34), 229-30.

[3] As pointed out in: Pollock, 'The 'Possible Worlds' Analysis of Counterfactuals', (Philosophical Studies 29: 6, 1976a), p. 471; Pollock, *Subjunctive Reasoning*' (Dordrecht: Holland, 1976b), p. 20, and Hertzberger, 'Counterfactuals and Consistency', Journal of Philosophy v. 76.2 (1979), pp. 83-88.

[4] Pollock (1976a) ibid, p. 471 and Pollock (1976b) ibid, p.20.

[5] Let R be the disjunction $R_1 \vee R_2 \vee ....$ Then the argument only requires that we infer $R > R \wedge \neg R_1 \wedge \neg R_2 \wedge ...$ from $R > R$, $R > \neg R_1$, $R > \neg R_2$, ...

[6] I ignore the complications which arise from there being 'non-entertainable' worlds.

[7] And when it comes to the argument itself, we might question the step in which we go from $[R_1 \vee (\neg R_1 \wedge R_2)] > \neg R_1$ and $[\neg R_1 \wedge \neg R_2 \wedge (R_3 \vee R_4 \vee ...)] > \neg R_1$ to $R_1 \vee (\neg R_1 \wedge R_2) \vee [\neg R_1 \wedge \neg R_2 \wedge (R_3 \vee R_4 \vee ...)] > \neg R_1$.

[8] For take any R_k-world w. Then w will be at least as far as a stranded or closest R_k-world w′. By Positive Effect, w′ will be a R_{k+1}-world; and so it will be at least as far as a stranded or closest R_{k+1}-world v. But w′ cannot be just as far as v since otherwise Negative Effect would not hold.

[9] Though, in all fairness, I should point out that paradoxicality under Lewis' proposal will be less widespread than triviality under our own; for the supposition A will the paradoxical for Lewis just in case every A-world is stranded while it will trivial for us just in case every A-world is a closest A-world or stranded.

[10] Since we can go from $A > C$ to $[(A \vee (A \wedge B)] > C$ by Substitution and then to $(A \wedge B) > C$ by Simplification.

11. A review of some of this work is to be found in §1.8 of Nute & Cross ‘Conditional Logic’ in *Handbook of Philosophical Logic* 2nd edition (eds. D. M. Gabbay & F. Guenthner), (2002, Dordrecht: Kluwer), 1-98.

12. Loewer ‘Counterfactuals with Disjunctive Antecedents’, Journal of Philosophy 73 (1976), 531-6.

13. McKay & van Inwagen ‘Counterfactuals with Disjunctive Antecedents’, Philosophical Studies 31 (1977), 353-56.

14. In a remarkable reversal, Nute in‘Conversational Scorekeeping and Conditionals’, Journal of Philosophical Logic 9 (1980, 153-66), p.161, went from accepting Simplification to rejecting it for just this reason

15. For simplicity, I will just deal with truth-functional complexity in what follows and ignore the quantifiers.

16. Versions of the semantics are to be found in van Fraassen ‘Facts and Tautological Entailments’, Journal of Philosophy (1969), 66:15, 477-87 and Schubert ‘The Situations We Talk About’, in Logic-based Artificial Intelligence’ (ed. J. Minker), 407-39, (2000, Dordrecht: Kluwer); and I had it in mind in my review of David Lewis’ ‘Counterfactuals’, Mind 84 (1975), 451-8; reprinted in ‘Modality and Tense: Philosophical Papers’, (2000, Oxford: Clarendon Press) 366-70. There is a significant connection with the metaphysical notion of *ground*, discussed in §5 of Fine ‘Guide to Ground’, to be published in a collection of papers edited by F. Correia & ..., (2011, Cambridge: Cambridge University Press).

17. A world, within the present framework, may be identified with a state which either contains or is incompatible with any other state. We assume that our state spaces are ones in which each state is contained in a world-state.

18. Supplemented with appropriate rules for truth-functional logic, the rules will in fact be sound and complete for the proposed ‘minimalist’ semantics.

19. Strictly speaking, within a formal semantics, this is all relative to a ‘model’ within which the formulas are evaluated.

20. An alternative to Completeness, in some ways more satisfactory, is to modify the truth-conditions for counterfactuals so that $A > C$ is true at a world w if C is *loosely* verified by u whenever $t \,||\!- A$ and $t \rightarrow_w u$.

21. An account along these lines is presented in Pearl’s *Causality* (2000, Cambridge: Cambridge University Press).

22. Some of these difficulties are discussed by Rachael Briggs in her unpublished 'Interventionist Counterfactuals'. We can actually use the previous puzzle to *prove* that no reasonable solution to this problem exists.

23. Though certain details in how exactly the semantics is to be applied will still need to be decided.

24. The present paper is largely based upon the Nagel Lecture I gave at Columbia in 2010. Some of the material from the paper was also presented at the Whitehead Lectures at Harvard, 2009, the Townsend Lectures at Berkeley, 2010, and a conference on Propositions and Same-Saying at Sydney University, 2010, and at talks to the philosophy departments of University of Miami and Virginia Commonwealth University. I should like to thank the audiences at those meetings for many helpful comments.