Counterfactuals and Causation: Transitivity

By

Miloš Radovanović

Submitted to

Central European University

Department of Philosophy

In partial fulfillment of the requirements for the degree of Master of Arts

Supervisor: Ferenc Huoranszki

Budapest, Hungary

2012

Abstract

Lewis claimed that counterfactuals are not transitive and that causation is. I present views according to which counterfactuals are transitive, dismissing Lewis' reasons to think otherwise. I move on to Lewis' theory/theories of causation, in which the transitivity of causation is left somewhat unexplained. I explore what transitivity of counterfactuals can mean for the transitivity of causation, and whether it can solve the problems raised by counterexamples to transitivity of causation. I find that even if counterfactuals are transitive, that helps us little with problems of intransitive sequences of causal statements. Moreover, it seems that transitive counterfactuals pose an equally bad, or even worse, basis for causation then Lewis' intransitive counterfactuals do.

Table of Contents:

Introduction	1
1. Stalnaker-Lewis' Counterfactuals and Transitivity	3
2. Strict Implication Counterfactuals and Transitivity	6
3. Lewis' Causation and transitivity	17
4. Transitivity of Counterfactuals and Causation	22
Conclusion	30
Bibliography	33

Introduction

In 1973 David Lewis presented both his theory of counterfactuals and his counterfactual theory of causation. Since then, a great deal of papers regarding those theories was written presenting problems and counterexamples to the original theories. It is interesting, however, that these critiques are typically not related. This fact strikes one as even more surprising having in mind that it is very reasonable to expect that tempering with counterfactuals is going to have some consequences for theories that use counterfactuals as their foundation, as is case with Lewis' theories of causation. This paper presents an effort to connect some problems of Lewis' theory of counterfactuals with his theories of causation.

I will mainly be dealing with problems regarding the transitivity of counterfactuals and the transitivity of causation. Lewis himself claimed that counterfactuals are not transitive, while maintaining that causation is a transitive relation. However, in the relevant literature one can find numerous counterexamples designed to show that transitivity fails in both cases. In this paper, I explore what consequences transitivity of counterfactuals may have to transitivity of causation, as Lewis defines it.

I will first (briefly) present Stalnaker-Lewis' theory of counterfactuals and go on to show how transitivity is dealt with within this theory. I will move on to theories of counterfactuals that take counterfactuals to be transitive after all. I will call such theories *strict implication theories*. Hopefully, the reader will be convinced that these are superior to Lewis' theory, but my argumentation does not depend on that, so I will not go into a detailed debate as to which type of theory, if any, one should accept.

Then, I will move on to Lewis' theories of causation and counterexamples to the transitivity of causation. Analyzing causation in terms of counterfactuals, I will apply the same argumentation used within the strict implication theories to show that counterfactuals

are transitive, in order to explore whether that can solve the problem of transitivity of causation. My point will be that the very same argumentation can be applied, and this may make one think that those who accept the strict implication theories can use their already established arsenal to defend the transitivity of causation. If that were true, strict implication theorists would, accepting Lewis' theory/theories of causation, have a clear path to the transitivity of causation; the path Lewis himself would be unable to take without dismissing his theory of counterfactuals. However, I will show that once one reaches the end of this road he will not find what he might have hoped for. It will turn out that Lewis' theory/theories of causation are not at all successful when combined with counterfactuals understood as within strict implication theories. Accepting these theories will then have unacceptable consequences to causal statements interpreted by Lewis' theory/theories of causation.

1. Stalnaker-Lewis' Counterfactuals and Transitivity

According to what has become known as the Stalnaker-Lewis theory or standard theory of counterfactuals, conditional $A > B^1$ is true in a world *a* if and only if:

a) There are no possible A-worlds (worlds in which A is true), i.e. A is impossible relative to *a*, or if

b) B is true in all A-worlds closest to a^{2} .

Closeness of possible worlds to the actual world a is determined by the 'overall similarity relation', which presents a total ordering of the possible worlds accessible from a.

However, theory presented is more in line with Lewis' theory than Stalnaker's. Stalnaker's theory supports the "uniqueness assumption", according to which there is always exactly one possible antecedent-world that is the closest (most similar) to the actual world. Thus, for Stalnaker a counterfactual A > B is true in a world *a* if and only if B is true in the A-world most similar to the actual world *a*. Also, according to Stalnaker the similarity relation provides one with partial and not total ordering of the possible worlds accessible from *a*. Lewis, on the other hand, presupposes total ordering, but dismisses the "uniqueness assumption" together with the so called "limit-assumption", i.e. he allows that for some antecedents there are no closest antecedent-worlds (in the sense in which there is no smallest real number bigger than 2).³ Although the theory outlined above is more in line with Lewis' than Stalnaker's theory of counterfactuals, the differences between these two accounts make no particular difference when it comes to transitivity. Transitivity being the focus of this

¹ I will use ">" as a symbol of counterfactual implication. ">" should be read as "counterfactually implies" or "Had it been the case that ..., it would be the case that ... "

² See: Stalnaker 1968: 104 and Lewis 1973: 16

³ Stanlaker 1968; Lewis 1973

paper, we will no longer be dealing with these subtle differences and will proceed using the 'standard theory'.

According to this, standard theory, the truth value of a certain counterfactual is established by determining whether there are possible antecedent-worlds accessible from the actual world and, if there are such worlds, whether in those of them that are most similar (closest) to the actual one, consequent holds. For example, "Had it rained, the streets would have been wet" is true because in all the closest (most similar) possible worlds in which it rains, the streets are wet. In other words, it would take a bigger departure from the actual world to get to the world in which antecedent is true and consequent false, then to the world in which both are true.

Now, worlds relevant for determining the truth value of one conditional can be completely different from those relevant for some other conditional. Context provides us with the ordering of the worlds, but the antecedent is the one that picks up those that matter for a certain conditional. Thus, there is a possibility that all the closest A-worlds are B-worlds (A > B), while all the closest B-worlds are C-worlds (B > C), but that not all the closest A-worlds are not transitive. Stalnaker and Lewis use the very same example (Hoover) to show that their theories match up with, support and explain our ordinary language intuitions when it comes to the transitivity of counterfactuals:⁴

- (1) Had John Edgar Hoover been a Communist, he would have been a traitor.
- (2) Had he been born in Russia, he would have been a Communist.

(3) Had Hoover been born in Russia, he would have been a traitor.

While (1) and (2) seem intuitively true, (3) seems false, and Stalnaker-Lewis theory offers us an explanation of why this is so.

⁴ Stalnaker 1968: 106; Lewis 1973: 33

Since Hoover was an FBI director famous for his fight against the left-wingers, the closest possible worlds in which he himself is a Communist are those in which he is also a traitor; which makes (1) true.

Since Russia of that time, on the other hand, was a Communist country with a very powerful left-oriented propaganda, the closest worlds in which Hoover is born in Russia are also the worlds in which he himself is a Communist; which makes (2) true.

Still, there is no reason to think that in all the closest worlds in which Hoover was born in Russia he is also a traitor. Russians are not typically treacherous just because they were born in Russia, so in the antecedent-worlds most similar to the actual world there is no reason to think that the Russian-Hoover is necessarily a traitor; which makes (3) false. Thus, in (Hoover) case, we have that B > C, A > B and $\sim(A > C)$, i.e. we have that transitivity fails.

Therefore, according to Stalnaker-Lewis' theory, counterfactuals are not transitive, and (Hoover) is a typical example as to why this is so. This means that if one wants to claim that counterfactuals are, after all, transitive, he must explain our intuitions about the conditionals (1), (2) and (3) and show where exactly does Stalnaker-Lewis' theory go astray.

2. Strict Implication Counterfactuals and Transitivity

Examples such as (Hoover) were practically taken to serve as a proof that transitivity fails when it comes to counterfactuals. However, not all were convinced by such examples, and different interpretation of these (counter)examples emerged. Among the first of the unconvinced ones was Warmbrod (1981) who thought that counterfactuals should be treated as strict implications.⁵

On Warmbrōd's theory, a counterfactual A > B is true in a model of ordered quadruples (a, K, R, | |) if and only if, for each $w \in K$ such that aRw, material implication $A \rightarrow B$ is true in w. Here a stands for the actual world, w for some possible worlds and K for a set of all possible worlds. R is a binary relation defined on K which determines which of the possible worlds are accessible from a, i.e. accessibility relation, and | | is an evaluation function translating propositions to the appropriate sets of worlds (proposition p to the set | p |, a subset of K).⁶ In other words, a counterfactual A > B is true in a world a if and only if the material implication $A \rightarrow B$ holds throughout all the worlds accessible from a.

Crucial difference between the Stalnaker-Lewis type and strict implication theories such as Warmbrod's is found in how they treat context-sensitivity of counterfactuals. Lewis' relation of overall world similarity that provides one with total ordering of the worlds is context dependant, and in a given context, some differences between the actual and other possible worlds will be more, and some will be less important. So, different contexts require different similarity relations and different worlds orderings.⁷ Lewis (1986c) offers a detailed analysis as to what would be the default similarity relation and what usually matters the most

⁵ Wright (1983), Lowe (1990) and Broggard and Salerno (2008) vastly share Warmbröd's intuitions and develop similar arguments.

⁶ Warmbrod 1981: 278

⁷ Lewis 1986a, "Counterfactual Dependence and Time's Arrow": 52

when talking about similarity. Where strict implication theories differ from standard theory is that within the Stalnaker-Lewis approach once the overall world similarity relation is adjusted to a given context, all conditionals are evaluated according to the same measure of similarity. Strict implication theories, on the other hand, do not pretend to evaluate all counterfactuals 'simultaneously'. According to these theories, some counterfactual conditionals cannot be evaluated together, as they require different models. Why is that so?

Early in a conversation, some presuppositions or assumptions are made, explicitly or implicitly. Counterfactuals typically impose some implicit presuppositions that are assumed further in the discussion. Take (Hoover) example. Once (1) is uttered it is from that moment assumed that Hoover one has in mind is an American-born FBI director famous for his fight against Communism etc. Without such assumptions there is no reason to believe that (1) is true. But, for as long as we are in the same discourse, we must not contradict such assumptions. If we ifwere to do so, we would be saying something that is either (I) obviously false or (II) is the beginning of another discourse. Thus, if we utter say p, where p contradicts some of the previously made assumptions, then p is either (I) false, if it is to be evaluated in the same model with the rest of the discourse, or (II) p is not a part of the same discourse with was previously said because it cannot be evaluated in the same model with previous

The model here is a typical model of modal logic, such as a model presented in explaining Warmbrōd's theory above – (a, K, R, | | |). The difference between ordinary modal logic (for example S5) and Warmbrōd's semantics is how we understand the accessibility relation. In Warmbrōd's model not all possible worlds are accessible, as is the case with S5, but only possible worlds relevantly and sufficiently similar to the actual world. For understanding the intuitive significance of the accessibility relation the following restriction is crucial: every assumed or presupposed proposition of a certain conversation must be true in

all the accessible worlds. For example, once (1) is uttered, propositions claiming that Hoover is an American-born, or that he is an FBI director etc., must be true in all the accessible worlds. Therefore, each assumption has the status of truth within the scope of accessible worlds i.e. is treated as a necessity further on in the conversation. This means that if an antecedent of some counterfactual contradicts some of the previously made assumptions, then that antecedent is impossible, and the conditional is vacuously true, because material implications with false antecedents are always true (case (a) in the terms of standard theories).⁸ For such a model, Warmbrōd would say that its accessibility relation is not *normal*, where an accessibility relation is normal for a certain discourse if and only if in the set of accessible worlds for every antecedent there can be found at least one possible world in which that antecedent is true.⁹ The accessibility relation is, however, determined by the first conditional of the discourse.

Therefore, for as long as the first conditional of the discourse is concerned, there will be no differences in truth value no matter whether we use Stalnaker-Lewis' theory or Warmbrōd's – accessible antecedent worlds according to Warmbrōd will be exactly those worlds that Stalnaker-Lewis' theory would pick out as the closest possible worlds.

The difference appears when evaluating the conditionals that follow, because the accessibility relation does not adept to them, but they find it, so to speak, finished and ready. While Stalnaker and Lewis evaluate counterfactual conditionals as isolated, Warmbrōd insists on evaluating them relative to a piece of discourse they are a part of. What was assumed prior to some counterfactual, then, determines the context and accessibility relation according to which that counterfactual will be evaluated.

⁸ Warmbrōd himself is hasitant to say that counterfactual conditionals with impossible antecedents have any truth value because he believes that our intuitions about such cases are not sufficiently clear. However, when analyzing a counterexample to the transitivity of counterfactuals, he also leans in the direction our analysis is leaning.

⁹ Warmbrōd's 1981: 282

This means that according to the strict implication theories (Hoover) is not a counterexample to the transitivity of counterfactual conditionals. Namely, once (1) is uttered, it is from that moment on assumed that Hoover is an American-born, that he is an FBI director etc. But then, both (2) and (3) have impossible antecedents (false in all the accessible worlds) because among accessible worlds there are no worlds in which Hoover was born in Russia. We have excluded such worlds from the accessible ones by the assumptions we had to make in order to have (1) true. Thus, conditionals (2) and (3) have impossible antecedents when uttered after (1), which means that they are either (I) trivially true, or (II) (2) and (3) are not parts of the same discourse with (1), because they require a different model in which they would be non-trivially true or false i.e. they require a model with a 'normal' accessibility relation. In one case (I) we have that all three conditionals are true, and in the other case (II) they are not to be considered as a part of the same discourse with one. One way or another, (Hoover) fails to provide a counterexample to the transitivity of counterfactuals.

An important point to make here is that just replacing the order of conditionals or premises affects the plausibility of (Hoover) as a counterexample. However, such a reordering would provide us with a normal accessibility relation.

(2) Had Hoover been born in Russia, he would have been a Communist.

(1) Had he been a Communist, he would have been a traitor.

(3) Had Hoover been born in Russia, he would have been a traitor.¹⁰

It seems that we are much less inclined to accept (1) as true in this order and this is something strict implication theories can offer an explanation of. Indeed, it seems to be impossible to explain away this phenomenon if one insists that there is no context-shifting somewhere along the argument.¹¹ According to strict implication theories, what is going on

¹⁰ Warmbrod 1981: 283

¹¹ Both Warmbrod (1981) and von Fintel (2001) stress this point.

here is that now that (2) is uttered prior to (1), it is (2) that imposes assumptions and determines the accessibility relation.

For (2) to be true, it is necessary that in all the accessible worlds in which Hoover is born in Russia, he is also a Communist. We must therefore assume that historical circumstances in Russia were such that Russians were typically Communists; typically enough for Hoover to be a Communist just because being born there. But then, what happens to (1)? Obviously, we have no reason to believe that (1) is true. We have no reason to say that all Communist Hoovers from all the accessible worlds are traitors. Among the accessible worlds there are worlds in which Hoover is a Russian, and it is in no way treacherous for a Russian to be a Communist. Russian Communists are not necessarily traitors because, as we said, among the accessible worlds all the assumptions (2) imposes are true, and one of those assumptions is that the social setting in Russia of the time - Communist traitors were a minority (even according to Stalin). Therefore, (1) is false. It is obvious that since according to this accessibility relation (1) is false, we need not worry much about (3) coming out false as well. Conditional (3) is false, for the very same reasons (1) is, but now that we have one true premise (2) and one false premise (1), (3) being false makes no problems for the transitivity of counterfactuals. What we have is one true premise, one false premise, and a false conclusion, which means that, once again, (Hoover) does not constitute a counterexample to the transitivity of counterfactuals.

Other counterexamples to the transitivity, as well as counterexamples to other rules of inference valid for Warmbrōd's strict implication and not valid for Stalnaker-Lewis conditionals (strengthening the antecedent and counterposition) can be analyzed and explained away in a similar fashion.¹²

¹² See: Warmbröd 1981, von Fintel 2001, and Brogaard and Salerno 2008

From the point of view of such strict implication theories, all of these counterexamples are founded on some, more or less subtle, context-shifts which Stalnaker-Lewis type theories typically do not take into consideration. All these counterexamples can then, with a relevantly similar argumentation, be characterized as cases in which we have propositions that contradict previously established assumptions, which results either in a kind of triviality, or in a change of model. This is what makes this approach 'static', as it tends to evaluate all counterfactuals of a discourse according to one fixed accessibility relation.

More recently, von Fintel (2001) has proposed an account that treats counterfactuals as strict implications, but uses dynamic semantics. The basic idea of dynamic semantics is a very simple one. What we had so far is that some expression 'a', in a certain context c is mapped to a certain proposition p. The novelty of dynamic semantics is that there are cases in which the expression 'a' actually influences the original context c and changes it to c'. In such cases 'a' should not be evaluated in the initial context c, but in the context c'. von Fintel believes that this dynamic approach is much more in line with what goes on when we use counterfactual conditionals.¹³

This position is very much in line with the work of Warmbrōd, and von Fintel discusses Warmbrōd's proposal as a precedent of his own account. What is a definite similarity between the two is that both authors take it that the real problem with examples of nonmonotonicity of counterfactuals is that in such cases we are dealing with conditionals that make claims about completely different sets of worlds. Also, von Fintel too takes it for granted that Stalnaker-Lewis' semantics correctly describes the truth-conditions of a counterfactual uttered in the initial context.¹⁴ Therefore, isolated counterfactuals will be assigned the very same truth values within von Fintel's theory as withing Stalnaker-Lewis

¹³ Von Fintel 2001: 123

¹⁴ *Ibid.* 127

theory. The crucial difference surfaces when dealing with sequences of counterfactuals. The counterfactuals that follow some other such conditionals will rarely find the same initial context, as that context is typically updated by the utterance of prior conditionals. If a counterfactual A > B is uttered in the initial context *c*, and this conditional changes that context to *c*', it will be interpreted according to *c*'. But if a conditional B > C follows, it will not have *c* as it initial context, but *c*'. Then B > C updates the context *c*' to *c*'' and is evaluated according to *c*''. As von Fintel puts it:

A crucial feature of the account is that the modal horizon is passed on from one counterfactual to the next and that it continually evolves to include more and more possibilities. Where the analyses will crucially differ is in cases where a counterfactual early in a sequence has brought into play some remote possibility... a later counterfactual cannot ignore any possibilities as far out as the possibilities concerned earlier.¹⁵

What von Fintel says here is that modal horizon can be broadened, but cannot shrink¹⁶, as the counterfactuals that follow cannot ignore the possible worlds introduced into the set of accessible worlds by some earlier counterfactuals. This feature of von Fintel's approach is exactly what guarantees that in our 'modal horizon' we will always have a possible world for every antecedent of the discourse, i.e. it ensures that the accessibility relation will be what Warmbrōd calls a 'normal' accessibility relation.

If the initial context *c* does not include any A-worlds, uttering A > B in that context *c* updates the context to *c*' and the modal horizon is broadened to allow for the most similar or closest A-worlds. However, once the 'horizon' is broadened in such a way, all the possible worlds introduced into the 'horizon' (or all the possible worlds that can be reached by the accessibility relation in Warmbröd's terminology) are relevant for evaluating the following counterfactuals. Since this theory treats counterfactuals as strict implications, it will not then be sufficient for the following conditional, say B > C, to be true, that in all the closest B-

¹⁵ *Ibid.* 129

¹⁶ *Ibid.* 131

worlds in the modal horizon C is true. What is required is that C is true throughout all the Bworlds in the horizon, i.e. that material implication $B \rightarrow C$ is true in all the accessible worlds. The question around the corner here is what this means for monotonocity of counterfactuals or, more precisely, what this means for the transitivity of counterfactuals and our (Hoover) example?

Von Fintel actually uses (Hoover) as evidence for his theory and against Stalnaker-Lewis' approach.¹⁷ Let us have another look at the example. We have presented it in two different orders:

(1) Had John Edgar Hoover been a Communist, he would have been a traitor.

(2) Had he been born in Russia, he would have been a Communist.

(3) Had Hoover been born in Russia, he would have been a traitor.

and:

(2) Had Hoover been born in Russia, he would have been a Communist.

(1) Had he been a Communist, he would have been a traitor.

(3) Had Hoover been born in Russia, he would have been a traitor.

We said that while in the first sequence we are somewhat inclined to accept (1) and (2) as true and dismiss (3) as false, once (2) is uttered prior to (1) we do not feel tempted to accept (1) as true. von Fintel notices that the 'natural way of reading the second premise [in the second sequence] is as taking into account a set of Communist scenarios *including* those introduced by the first premise'¹⁸. This is why we take (1) to be false when uttered after (2). More importantly, von Fintel raises the question as to where this difference in reading the two sequences comes from. If there is no context-shifting somewhere along the line of the argument, then how can one explain this asymmetry? Indeed, this seems to be impossible task

¹⁷ *Ibid*.

¹⁸ *Ibid.* 132

for Stalnaker-Lewis type theories as they systematically overlook the difference between the two cases, so it seems that such examples speak in favor of strict implication approach.

Now, it is clear that in the second (2)-(1)-(3) case the first conditional uttered provides us with a broad modal horizon in which (1) and (3) are false. Conditional (2) introduces the possible worlds in which we have a Russian (and Communist) Hoover into consideration. Once those worlds are in our modal horizon or once they fall under the reach of the accessibility relation (in Warmbrōd's terminology), the following conditionals (1) and (3) are false, as material implications they instantiate are not true throughout the 'modal horizon' or throughout all the accessible worlds. In this case, the first conditional provides us with a sufficiently broad horizon to evaluate all the conditionals of the discourse, and the evaluation provides us with one true premise (2), one false premise (1) and a false conclusion (3). Thus, in this case there is no need to fear that this sequence of counterfactuals is intransitive. A more interesting question is what happens with the first, (1)-(2)-(3) case?

When (1) is the first conditional of the sequence, we need only introduce those possible worlds in which Hoover is a Communist, but still an FBI director, American =-born etc. So, if we had the initial context c which consisted only out of our actual world a, we broaden our modal horizon as much as necessary to reach some antecedent worlds – those in which Hoover is a Communist. This leaves us with an updated context c', according to which among the relevant worlds we now have Communist-Hoover worlds. The following conditional, (2), finds no adequate worlds in such a setting, and broadens the modal horizon further on in order to allow for worlds in which Hoover is a Russian-born. This updates the context c' to c'', as we now have more worlds in our consideration. However, conditional (3) does not require us to introduce some new worlds, as (2) has already equipped us with the worlds in which Hoover was born in Russia, so (3) is evaluated in the context it finds – c''.

(1) as true in the context c', we evaluate (2) as true in the context c'', and we evaluate (3) as false in the context c''. In fact, von Fintel calls such an inference 'dynamically valid' but he adds that "[w]hile this seems nice, there is reason to think that dynamic entailment is in fact not the notion that we use to assess logical arguments".¹⁹ Here we can see both how von Fintel's theory resembles and how it differs from Warmbröd's.

While for Warmbrod evaluating one piece of discourse ((1)) relative to one context (c') and evaluating another piece of the same discourse ((2) and (3)) relative to a different context (c'') is impossible, for von Fintel it is quite natural. This is so because von Fintel does not take it that holding the context fixed is a necessary condition for staying in the same discourse. Where Warmbrod sees a change of discourse due to context-shifting, von Fintel sees just context-shifting, dismissing the claim that one cannot change context if he wants to stay in the same discourse:

[T]here is no argument here against the idea that the context can and does change over the course of simple pointful discourses. The proper diagnosis would seem to be that \dots the set of worlds quantified over properly expands, but that over the course \dots it cannot shrink.²⁰

Now, one might object that this context-shifting that dynamic semantics find to be so natural is something Warmbrod was very much against. This could mean that it would be wrong to put Warmbrod and von Fintel 'under the same hat', as their theories seem principally incompatible. That this difference is not as harsh as it may seem is secured by von Fintel's distinguishing of 'dynamic' entailment' and the 'notion that we use to assess logical arguments'. He notices that "[f]or logical argumentation, we take speaker to be committed to

¹⁹ *Ibid.* 143

²⁰ *Ibid.* 131

a stable context" and that "[s]omeone who makes a logical argument gives an implicit promise that the context is not going to change during the argument".²¹

Thus, if (Hoover) was to be a logical argument, it fails, as it can only work if we overlook the context-shifting underlying our 'first-hand' evaluation of the relevant conditionals. Once we have a stable context, 'modal horizon' or 'accessibility relation', which is necessary for the evaluation of a logical argument, (Hoover) fails as a counterexample, as it poses no threat for the transitivity of counterfactuals.

Once again, this strategy of insisting on quantifying the same possible worlds when evaluating counterfactuals (at least for as long as they are to be considered a 'logical argument') can be applied to other cases of alleged nonmonotonicity of counterfactual conditionals. What is important for our purposes is that we have equipped ourselves with a principal way of dealing with the alleged counterexamples to transitivity of counterfactuals. That being done, we move on to the problem of causation.

²¹ *Ibid.* 143

3. Lewis' Causation and transitivity

When it comes to causation, Lewis offers two different theories, both analyzing causation in terms of counterfactuals, understood according to his theory of counterfactuals.

In Lewis' earlier theory he starts from the notion of *counterfactual dependence*: in a conditional A > B we have that B counterfactually depends on A. But, A and B stand for propositions while, for Lewis, it is not propositions that stand in causal relations but events that are both actual and distinct. Solution is not hard to find. For every event, there is a proposition stating that that event has occurred, which makes that proposition true only in those worlds in which the event in question is actual. Thus, events (a, b, c...) will simply be represented by propositions (A, B, C...) which claim that these events have occurred i.e. are actual.22

The next step is causal dependence. Where a and b are distinct possible events, bcausally depends on a if and only if the propositions that claim or deny that those events occurred B, ~B counterfactually depend on propositions A, ~A. In other words, b causally depends on *a* if and only if whether b occurs depends on whether a occurs. Then, to establish a causal dependence of b on a, it is required that counterfactual conditionals A > B and $\sim A > B$ ~B are true. Whenever a and b are not actual events, conditional ~A > ~B will have an "automatically" true antecedent and consequent, so whether b causally depends on a or not will be determined by the counterfactual A > B. On the other hand, whenever a and b are actual events, conditional A > B will be trivially true, so whether b causally depends on a or not will be determined by the counterfactual $\sim A > \sim B$. For Lewis, the second case is the relevant one, as he is trying to provide a theory that deals with actual causation.²³

²² Lewis 1986a, "Causation": 166 ²³ *Ibid.* 166 - 167

So, Lewis writes that "If c and e are two actual events such that e would not have occurred without c [\sim C > \sim E], then c is a cause of e" but he "rejects the converse".²⁴ This means that for an event c to cause event e, truth of conditional $\sim C > \sim E$ is sufficient, but not a necessary condition. The reason Lewis rejects 'the converse' and does not take the truth of underlying conditionals to be the necessary condition for causation is that he believes that causation must always be transitive. Since, according to Lewis' theory of counterfactuals these conditionals are not transitive, failure of transitivity of counterfactuals would mirror in the failure of transitivity of causation, and that is something Lewis wants to avoid. Therefore, there may be cases in which we have causation, but no causal dependence. Lewis thus reaches his final definition of causation:

Let c, d, e, \ldots be a finite sequence of actual particular events such that d depends causally on c, e on d, and so on throughout. Then this sequence is a causal chain. Finnaly, one event is a cause of another iff there exists a causal chain leading from the first to the second.²⁵

Of course, this causal chain may have only two members, but no less then two: one could not say that an event is a cause of itself because it does not satisfy the criteria that only distinct actual events enter into causal relations, since no event is distinct from itself.

So far, we have seen how Lewis, starting from "non-transitive" counterfactual conditionals and counterfactual dependence, through causal dependence and causal chain, reaches the concept of causation which "must always be transitive". So much about Lewis' first theory. A detailed presentation of it will save us some space-time in presenting his later theory.

In Lewis' second theory *influence* takes over the causal dependence, and *chain of* stepwise influence takes over the causal chain. Namely, where a and b are distinct actual events, an event a influences and event b if and only if there is a significant scope a1, a2, ...

²⁴ Ibid. 167 ²⁵ Ibid.

of different and not too distant alterations of an event *a* (including the actual alteration *a*) and if there are b1, b2, ... alterations of an event *b*, at least some of which differ, such that had *a1* occurred, *b1* would have occurred, and had *a2* occurred, *b2* would have occurred, and so on.²⁶ Roughly speaking, and event influences another event if and only if relevantly small alterations of the first event are systematically followed by alterations of the second event throughout possible worlds. But now influence is defined in terms of counterfactuals, which means that we reach the transitivity of causation via "by-passing" once again. Thus, an event *a* is a cause of an event *b* if and only if there exists a "chain of stepwise influence" from *a* to *b*.²⁷

This new theory may at first strike one as a bit complicated, but the main idea is a rather simple one – 'wiggling' causing events should have consequences for the events caused. The apparent complexity is a small price to pay considering what is gained by defining causation in such a way. Namely, Lewis' first theory provides us with clear direction when we have that *whether* an event occurs counterfactually depends on *whether* some other event had occurred. What is gained by new theory is that causation now takes into consideration not only whether events occur, but also, *how* and *when* they occur, so we are provided with patterns of counterfactual dependence from *whether, when* and *how* the causing event occurred, to *whether, when* and *how* the effect event occurs.²⁸

It is obvious that according to both of Lewis' definitions causation is a transitive relation. For Lewis, it "must always be transitive"²⁹. It is interesting, however, that Lewis does not argue for this claim and, in general, there are not many authors of the field do argue for it. Transitivity of causation is usually taken to be somewhat obvious, which is one of the reasons why it is difficult to offer a principal explanation of counterexamples to transitivity of

²⁶ Lewis 2004: 91

²⁷ Ibid.

²⁸ Ibid.

²⁹ Lewis 1986a, "Causation": 167

causation. Literature is actually filled with such counterexamples, and these typically share the same structure, and are designed according to the same pattern. For these reasons, I will be dealing with only one such counterexample (bomb), as the same argumentation can be applied to any other counterexample I have encountered so far.³⁰

Here's the story: An assassin places the bomb in front of my door. My friend, luckily enough, sees the bomb and is capable of defusing it, and he does so, saving my life. If causation is invariably transitive (and what other transitivity is there?), then assassin's placing of the bomb saved my life. We have the following three claims:

(4) My friend's finding the bomb assassin planted is a cause of me being alive.

- (5) Assassin's placing the bomb is a cause of my friend finding it.
- (6) Assassin's placing the bomb is a cause of me being alive.³¹

Here we are inclined to consider (4) and (5) true, while dismissing (6) as false. Therefore, if we are to save the transitivity of causation, we must offer some sort of an explanation of our intuitions in this, and relevantly similar cases.

Lewis lists this example together with some other typical counterexamples to transitivity of causation and dismisses them all, maintaining that transitivity holds. Namely, Lewis believes that (6) is true and he offers three kinds of reasons as to why we are inclined to think of (4)–(6) as of an intransitive sequence (reasons are adjusted to the example presented):

Firstly, an assassin who plants the bomb in front of my door does end my life, but causes it as well. One version of my life is prevented, and another is actualized. It is wrong to think that just because an assassin pretends to end my life, he is unable to cause it.

³⁰ Some of these counterexamples are even easier to deal with than the one I deal with here.

³¹ The example is taken from Lewis 2004: 97; a similar example and an interesting analysis can be found in Maslen 2004: 350-2

Secondly, events like bomb placing usually lead to death, not life. It is wrong, however, to confuse what is usually the case with what is the case in this particular sequence of events.

Thirdly, we notice that placing of the bomb did not have any consequences for me being alive in the sense that I would have been equally alive had the bomb never been planted in the first place. Still, my being alive in both cases does not mean that my life is not caused by the planting of the bomb. There are two relevant scenarios in which I am alive. One is that in which bomb has been planted but defused, and in other there is no bomb to begin with. In both cases, my being alive is caused, one way or another.³²

It seems, however, that the three reasons Lewis offeres are not the happiest way out of the problems these counterexamples pose. As Lewis himself notices, in dismissing these counterexamples he acts as a historian who, tracing a certain causal chain concludes, with no hesitation, that what comes in the end of the chain is caused by what precedes it.³³ If that is the case, then it seems that Lewis' approach to the problem begs the question – transitivity of causation is presupposed in the answer establishing the very same transitivity.

The problem is that the transitivity of causation is achieved in a rather *ad fiat* way within Lewis' theories. It is simply built into the definitions of causation, because for Lewis, causation just "must always be transitive". We are not provided with some arguments as to why this is so, and we are not provided with a strategy of dealing with the counterexamples such as (bomb). This leaves the transitivity of causation somewhat unfounded and in need of a stronger basis, if it is to resist the force of counterexamples.

³² Lewis 2004: 98-9

³³ *Ibid.* 99

4. Transitivity of Counterfactuals and Causation

My goal is to connect the two different critiques of Lewis' theories i.e. to connect the problem of transitivity of counterfactuals to that of transitivity of causation. The guiding idea here is to try to provide a basis for the transitivity of causation by arguing for the transitivity of underlying counterfactuals. It is clear that Lewis' reductive account(s) of causation will, combined with transitive counterfactuals, result in transitive causation. The question is whether such a combination will yield answers to the counterexamples to the transitivity of causation, i.e. whether it will provide the tools for explaining away such counterexamples.

In order to explore this possibility, we need to apply the 'strict implication theories argumentation to the counterfactuals that (bomb), as a typical counterexample to the transitivity of causation, translates to.³⁴ Thus, we start off with the counterfactuals that determine the truth value of (4), (5) and (6). These are:

(7) Had my friend not found the bomb assassin planted, I would be dead.

(8) Had the assassin not planted the bomb, my friend would not have found it.

(9) Had the assassin not planted the bomb, I would be dead.

Now, we can only deny causation of transitivity if we consider sequences such as (7)–(9) intransitive. If we think that (7)–(9) is transitive, and that it is the adequate translation of (4)–(6), there is no room for us to deny the transitivity of causation. I take it that the 'translation' from causal statements to the counterfactual ones is adequate, i.e. in accordance with Lewis' theories of causation.

It is clear that according to Stalnaker-Lewis theory of conditionals, (7)–(9) is not transitive. According to this theory we would say that (7) is true since in the possible worlds

 $^{^{34}}$ As both Lewis' theories of causation would evaluate (bomb) ((4), (5) and (6)) in the same way, I will be using Lewis' earlier theory for reasons of simplicity.

closest to our world in which the assassin plants the bomb and my friend does not find it, the bomb blows me up. Therefore, it is true that had my friend failed to find the bomb I would have been dead.

Because in the closest possible worlds in which the assassin does not plant the bomb there is no bomb to be found by my friend, (8) is true as well. Possible worlds in which there is some other bomb for my friend to find are more distant than the worlds in which assassins' not planting the bomb means that there is no bomb at all.

However, (9) is false since in the closest worlds in which the assassin does not plant the bomb there is no reason for me to die a premature death. This does not mean that (6) corresponding to (9) is false because Lewis maintained that there can be causation without counterfactual dependence (causal chain may just be enough). Still, it is hard to see what else could justify sticking to (6) other than Lewis', in this case artificial, 'causal chain'.

Now, if transitivity (7)–(9) would secure the transitivity of causation, and we have a theory that claims that counterfactuals are 'invariably' transitive, then wouldn't that solve the problem? Indeed, if seemingly intransitive sequences of counterfactuals can be dispelled of their apparent intransitivity by exposing the subtle context-shifts, then (7)–(9) must be transitive, and so does (4)–(6) i.e. (bomb). What needs to be done, then, is to analyze (7)–(9) in the very same way we analyzed (Hoover) i.e. within the strict implication theories. Here it goes.

In order for (7) to be true we assume that the assassin has planted the bomb. This means that in all the accessible worlds it is true that the assassin has planted the bomb. But then there are no accessible worlds in which the bomb has not been planted, so (8) and (9) have impossible antecedents, which makes these conditionals either (I) trivially true or (II) parts of a different discourse from the one in which (7) is uttered, as they require a different

model in which they could be non-trivially true or false. So, we have that according to 'static' strict implication theories conditionals (7), (8) and (9) are either (I) all true or (II) are not part of the same discourse, and therefore cannot form an argument. One way or another, we have no counterexample to transitivity.

If we were to go for a dynamic strict implication theory such as von Fintel's, we would have that (7) is true in the context c', while (8) requires a context update and broadening the modal horizon. Once that is done and we let the no-bomb-planted worlds to enter our modal horizon we are in context c'', in which (8) is true. Since (9) requires no further adding of the possible worlds, it can be evaluated in c'', a context in which it is false. However, we have that (7) and (8) are true while (9) is false only because we have evaluated them according to different sets of accessible worlds, i.e. in different contexts. Since for analyzing a logical argument it is crucial to hold context fixed, (7)–(9) does not form a logical argument, as (7) is evaluated in a different context from that in which (8) and (9) are evaluated in. Therefore, (7)–(9) cannot be used as a counterexample to transitivity of counterfactuals, which means that once again we have no threat of intransitivity of counterfactuals.

If we were to swap the order of premises we would have:

- (8) Had the assassin not planted the bomb, my friend would not have found it.
- (7) Had my friend not found the bomb the assassin planted, I would be dead.
- (9) Had the assassin not planted the bomb, I would be dead.

Now, according to the strict implication theories, (8) determines the accessibility relation. For (8) to be true, my friend must not find the bomb assassin planted in the worlds in which the assassin had not planted the bomb at all. This should not be too much of a task for him. Not finding the non-existing objects seems like a rather easy thing to do. What is required, then, is that among the accessible worlds there are some worlds in which the assassin has not planted

the bomb. For the accessibility relation to be *normal* for the given discourse, there must be at least one such world among the accessible ones. But why, then, would we consider (7) true? For (7) to be true I must end up dead in all the accessible worlds in which my friend fails to find the bomb. But why would I end up dead in the world in which the assassin had not planted the bomb to begin with? Thus, (7) is to be dismissed as false which means that (9) being false is not a problem at all. Once again we have one true premise (8), one false premise (7) and a false conclusion (9). The transitivity prevails.

According to dynamic strict implication analysis, (8) broadens the modal horizon as to include the possible worlds in which the assassin has not planted the bomb among the accessible worlds. Once again, we take it that my friend fails to find some non-existing bombs, and we have that (8) is true. But now that our modal horizon is broadened to the extent that it includes among the accessible worlds those in which there is no bomb planted, it is not true that I die in all the worlds in which my friend fails to find the bomb. As we have said, there are worlds in which he fails to find it because it is not there to begin with. Thus, (7) is false when uttered after (8), as it can only be true in a 'smaller' modal horizon. Once (8) has introduced more distant worlds into consideration, (7) comes out false. Conditional (9) being false then makes no problems, as we have, as with Warmbrōd's 'static' approach, one true premise (8), one false premise (7) and a false conclusion (9). Same as before, no threat for transitivity of counterfactual conditionals.

Thus we have analyzed the (bomb) in the same way we analyzed (Hoover), presenting the example in two different orders, and finding that in both cases there is no counterexample to the transitivity of counterfactuals. But it is not the order of the premises that is crucial. The point is rather that the first order ((7), (8), (9)) demonstrates the inadequacy of the accessibility relation that is not normal for the given discourse. In the framework of Warmbrōd's theory it does not satisfy the requirement that for every antecedent of the discourse there is at least one antecedent-world among the accessible ones.³⁵ In the framework of von Fintel's theory, it ((7), (8), (9)) fails to be a logical argument due to context-shifting.

The second order ((8), (7), (9)) demonstrates that not even with the normal accessibility relation do counterexamples achieve their goal, as one of the premises always turns out to be false. In other words, once we evaluate all counterfactuals according to the same set of possible worlds, in a sufficiently broad modal horizon, we find that one of the premises -(7) – comes out false, which means that there is no counterexample to the transitivity of counterfactuals. The big question here is – what could all this mean for (bomb) and transitivity of causation?

Well, if we were to use the accessibility relation of the (7), (8), (9) sequence, we would have to say that (4), (5) and (6) are either (I) all true, or (II) are not part of the same discourse. Accepting (II) seems very unintuitive. What we have is one of the cases in which Warmbröd's technical term 'dicourse' appears to be somewhat artificial when applied to the ordinary language. While he has provided some convincing examples where in a certain conversation or a discourse there are in fact more then one discourse (meaning more then one model), it is not always clear which parts of a conversation can be evaluated together and which not -(4), (5) and (6) may seem to fit the same context and we may want to evaluate them together. von Fintel would go on to claim that while (7), (8) and (9) are a part of the same discourse, they cannot be treated as a logical argument for as long as they are not evaluated according to the same set of worlds or accessibility relation or modal horizon. This means that if we want to treat (7)–(9) as a logical argument, we need to evaluate them in a single fixed context. For yon Fintel, evaluating these conditionals in the context in which not

³⁵ Warmbrod 1981: 282

all antecedents are provided with adequate possible worlds is unnatural, so the only other way to go is the 'normal' accessibility relation and a broader modal horizon. The intuition that we are talking about the one pointful piece of discourse is maybe stronger in the (4)–(6) case then in the (7)–(9) case, so it runs against either Warmbröd's theory of counterfactuals, or against Lewis' theory of causation.

If we were to not go as far as to claim (II) and stick to (I) we would have that all (4), (5) and (6) are true. This is so because the underlying counterfactuals are true, and that is sufficient for causation. This seems to be the closest we have gotten to principally explaining why (6) is true and why (bomb) is transitive. However, the big problem here is that we have that (8) and (9) are trivially true. While we have no problem accepting (5) as true, we do with accepting (6), but in this case they are both equally ill-founded as they get their truth value from conditionals that are trivially true. Is one supposed to claim that (5) and (6) are also trivially true? That they are trivially true in this particular context? Can causal statements even be trivially true? What would that even mean? To make things worse, counterfactuals are trivially true because they have impossible antecedents, but antecedents here stand for some events. Now, all counterfactual with such an antecedent would be trivially true, meaning what? That the event corresponding to the antecedent can be said to be the cause of any event, or maybe all events? The problem here is that anything with the same causing event and corresponding antecedent will be true:

(6*) Assassin's placing the bomb in front of my door is a cause of the global economic crisis.

Now, (6^*) is false if anything is, but corresponding to (6^*) we have (9^*) stating that:

(9*) Had the assassin not planted the bomb in front of my door, there would have been no global economic crisis. Conditional (9*) will be trivially true if there are no accessible worlds in which the assassin has not planted the bomb, which is the case if we evaluate (9*) in the context and accessibility relation imposed by (7) alone. Since (9*) would be true in such a context, (6*) would be true as well, as truth of underlying counterfactuals is sufficient for the truth of the relevant causal statements. It is clear that this is not at all a happy way to go, and that evaluating (7)–(9) in such a narrow modal horizon is not the right thing to do.

Claiming that (7)-(9) and (4)-(6) are not single pointful pieces of discourse seems highly implausible, and saying that (4)-(6) cannot form a logical argument due to a contextshift in (7)-(9) does not seem to resolve our problem. We do not see any context-shifting in (4)-(6) per se, and what changes when (5) is uttered after (4). What one would like to have at this stage is an analysis of (4)-(6) in which they are evaluated according to a single context and where it is showed what goes wrong. Offering no such analysis leaves one with a feeling that problem is not actually solved. In any case, it seems that the non-normal accessibility relation, or a narrow modal horizon, used in (7), (8), (9) case is not at all a happy solution we hoped for. It creates more problems and greater ones then those we started from.

On the other hand, if we were to try with the normal accessibility relation or a broader modal horizon used in the (8), (7), (9) sequence, we would say that while (5) and (8) are true, (4), (7), (6) and (9) are false. According to Lewis, we are not forced to dismiss a causal claim just because the events that the claim is about do not exhibit causal dependence, i.e. we do not have to claim that (4) and (6) are false just because (7) and (9) are. However, the whole point of applying the strict implication theories is to achieve a more direct route from counterfactuals to causation, and if we want to use the falsity of (9) to explain why we consider (6) false, and (7) is false for the very same reasons as (9), then it would be inconsistent not to judge that (4) is false because (7) is. Furthermore, if (8) and (7) were to establish a *causal chain* from the antecedent of (8) and (9) to the consequent of (7) and (9),

then we would be right, according to Lewis, in saying that (6) is true no matter that (9) is false. The problem here is that if we say that (7) is false, we cannot say that (4) is true, as we haven't got the established causal chain that would justify such a claim.

Thus, it seems that we should deny (4), and only because of (7), which claims that in all the accessible worlds in which my friend fails to find the bomb I end up dead, is false. We said that in order for (8) and (9) to be non-trivially true or false there needs to be among the accessible worlds at least one world in which the assassin does not plant the bomb. In such a world, my life is not threatened, and thus (7) is false, which means that (4) "My friend finding the bomb is a cause of me being alive" is false. Now, although I would have been alive had there not been for the assassin in the first place, once the assassin planted the bomb, we consider finding and defusing of that bomb as a cause of me staying alive. Our intuitions are very clear and strong and clear in such a case; certainly stronger than those that lead us to dismiss (7) as false. Thus, it seems that the normal accessibility relation used in (8), (7), (9) does not match our intuitions well. It requires us to dismiss what is, if anything is, a true causal statement.

Conclusion

We have seen how Stalnaker-Lewis' theory interprets counterfactuals and how it deals with the counterexamples to the transitivity of such conditionals by accepting and explaining them. On the example of Warmbröd's theory, we have seen how such counterexamples can be explained away and dismissed within the theories that take counterfactuals to be strict implications. Dealing with von Fintel's theory, we have seen how Stalnaker-Lewis example fails within the dynamic strict implication approach, and why one is somewhat attracted to accepting such counterexamples. According to both Warmbröd and von Fintel, for as long as we hold the context fixed while evaluating counterfactual conditionals i.e. for as long as we evaluate these conditionals according to a fixed accessibility relation or in the same model, there is no danger of intransitivity. Stalnaker and Lewis, however, are not denying that while evaluating a certain argument we should hold the context fixed. It is just that these two types of theories interpret context and context-sensitivity of counterfactuals in a different way i.e. they have different criteria as to when new model is required.³⁶

Once Lewis' counterfactual theories of causation were introduced, we were able to analyze causation in terms of counterfactuals. We have pointed out the fact that conditionals determining the truth value of causal statements from the counterexamples to the transitivity of causation can be analyzed in exactly the same way as conditionals constituting the counterexamples to the transitivity of counterfactuals. Thus we have applied to the

³⁶ Broggard and Salerno (2008: 45) stress out that it would be very odd that counterfactuals are exempt from the rule that the context should be held fixed when evaluating an argument, suggesting that this is the case with Stalnaker's and Lewis' theory. However, it is not that these theories dismiss such a requirement but that they do not mean the same thing by context as strict implication theories do. For an interesting analysis of the Broggard and Salerno paper see: Cross 2011

counterfactuals (bomb) was translated to, the same argumentation that was applied in the (Hoover) example.

But once we have tried, after arguing for the transitivity of counterfactuals that (bomb) translates to, to apply the results to the original counterexample, we got into all sorts of trouble. Whichever accessibility relation or fixed modal horizon we tried to use, we ended up unable to successfully relate our claims about counterfactuals to causal statements we started with, due to some unacceptable consequences. In one case we had to deal with 'trivially true causal statements', while in the other case we would have to deny an intuitively very plausible causal claim. While in the first case we ended up with nonsense and contradictions, dismissing a *prima facie* plausible claim (second case) does not necessarily have such severe consequences. Some intuitively plausible proposition may be dismissed in the light of sufficiently strong argumentation, but it seems that in this case we are very far from any such argumentation.

The only option left is to say that the counterexamples to the transitivity of causation fail to constitute logical arguments because of the underlying context shifts. The problem with such a claim is that we see no traces of context-shifts when dealing with causal statements, but only when dealing with counterfactuals. Thus, although this seems like the least problematic option for those who would like to have both strict implication counterfactuals and Lewis' reductive theories of causation, it seems that these two types of theories do not piece well together.

My point was to show that there is a strong connection between the problems of the transitivity of counterfactuals and those of the transitivity of causation, if we are to analyze causation in terms of counterfactuals. By revealing the common structure of these two types of counterexamples, I have shown such a connection. In addition, it was demonstrated why what seemed to be a natural path to take in order to defend the transitivity of causation is not

31

the road that reaches the required goal. Thus, since combining Lewis' theories of causation with the strict implication theories of counterfactuals leads to a dead-end, it seems that one should, in order to avoid such a position, avoid combining the two theories. One should, then, give up either on theories of strict implication, or on Lewis' theories of causation.

We have seen some evidence which could make one be more inclined to the strict implication theories of counterfactuals as they seem to be more in line with natural language and detect the subtle context-shifts Stalnaker-Lewis' theory leaves unattended. However, this seems to suggest a more general question of how tight of a connection one should have between counterfactuals and causation. Subtle context-shifts, accessibility relations, modal horizons etc. seem to be quite natural topics when dealing with counterfactuals, but not something we usually relate to causation. There is interplay of philosophy of language and metaphysics here, and it may be one that does not bring much to either side. In any case, whatever the connection between counterfactuals and causation should look like, not even transitive counterfactuals can provide us with a strong basis for transitive causation. The solution, if there is one, lies somewhere else.

Bibliography

Broggard, B. and Salerno, J.

2008 "Counterfactuals and Context", Analysis 68, p. 39-46

Collins J, Hall N. and Paul, L. A.

2004 "Counterfactuals and Causation: History, Problems, and Prospects" in Causation and Counterfactuals, eds. J. Collins, N. Hall, and L. A. Paul, Cambridge: The MIT press, p. 1-57

Cross, C.

2011 "Comparative world similarity and what is held fixed in counterfactuals", Analysis 71, p. 91-96

Đordevic, V.

"Similarity and Cotenability", Synthese, forthcoming

Lewis, D.

1973 Counterfactuals, Cambridge, Harvard University Press.

1973a "Causation" Journal of Philosophy 70, str. 556-67. (Reprinted in Lewis 1986a)

1986a Philosophical Papers vol. II, Oxford University Press

1986b "Events" in Lewis 1986a p. 241-69

1986c "Postscript to 'Causation'" in Lewis 1986a p. 172-213

2000 "Causation as Influence" Journal of Philosophy 97, p. 182-97

2004 "Causation as Influence" – a more elaborate version, in Causation and Counterfactuals, eds. J. Collins, N. Hall, and L. A. Paul, Cambridge: The MIT press, p. 75-106

Lowe, E. J.

1990 "Conditionals, context and transitivity", Analysis 50, p. 80-87

33

Maslen, C.

2004 "Causes, Contrasts, and the Nontransitivity of Causation" in Causation and Counterfactuals, eds. J. Collins, N. Hall, and L. A. Paul, Cambridge: The MIT press, p. 341-358

Nute, D and Cross, C.

2001 "Conditional Logic" in Handbook of Philosophical Logic Vol. IV, ed. Gabbay,

D. Reidel Publishing Company, Dordecht, Holland, p. 1-98

Ostojić, J.

A draft of a PhD dissertation "Conditionals, Context and Knowledge"

Stalnaker, R.

1968 "A theory of conditionals", in Studies in Logical Theory American Philosophical Quarterly Monograph Series 2, ed. N. Rescher, Oxford: Blackwell, p. 98–112

von Fintel, K.

2001 "Counterfactuals in a Dynamic Context." in Ken Hale: A Life in Language, ed. Michael Kenstowicz, Cambridge: MIT Press, p. 123-152

Warmbröd, K.

1981 "Counterfactuals and Substitution of Equivalent Antecedents", Journal of Philosophical Logic 10:2, p. 267-289

Wright, C.

1983 "Keeping track of Nozick", Analysis 43, p. 134-140