The International Conference on Interdisciplinary Humanities and Communication Studies
DOI: 10.54254/2753-7064/4/20220848

Backwards Time Travel and Backwards Causation

Zhao Guo

California University of Berkeley, Berkeley, CA,USA
guozhao07@berkeley.edu

Abstract: Is it possible to travel back in time? Objections against backwards time travel frequently focus on the impossibility of changing the past. However, there is a deeper problem for backwards time travel. It seems that if backwards time travel were possible, the time traveler could causally affect the past, even if she could not alter the past. Is it possible to causally affect the past? David Lewis famously argues that backwards causation is possible as long as we understand causal dependence in terms of counterfactual dependence rather than temporal precedence. In this paper, I argue that this Lewisian line of reasoning is untenable. Lewis' counterfactual theory of causation faces the so-called 'problem of effects'. To solve this problem, he either has to reintroduce temporal precedence back into the definition of causation or stipulate that the closest possible world needs to be historically the same as the actual one. Either way, the possibility of backwards causation is immediately ruled out.

Keywords: time travel, philosophy of time, causation, counterfactuals, David Lewis

1. Introduction

Is it possible to travel back in time? Objections against backwards time travel usually concern the logical impossibility of changing the past. If backwards time travel is possible, then a traveler, say, Anne, could go back in time, kill her grandfather and thereby cause herself never to exist (by causing her father never to exist). But this generates a contradiction: if Anne never comes into exists, then she cannot travel back in time to kill her grandfather. This is the infamous 'Grandfather Paradox'. Arguably, any act of changing the past would generate a similar paradox: under the assumption that there is only one past, to change the past means to make something both happen and not happen at some time in the past, which seems to involve a straightforward contradiction. Philosophers disagree over whether the possibility of backwards time travel implies the possibility of changing the past. However, there is a deeper problem for backwards time travel. It seems that if backwards time travel were possible, the time traveler could causally affect the past, even if she could not alter the past—that is, even if she could not make what has happened not have happened or make what did not happen have happened, she still could causally bring about what has happened to happen. Therefore, even if the possibility of backwards time travel does not imply the possibility of changing the past, it seems at least imply the possibility of backwards causation. But is backwards causation possible? I argue it is not.

2. Backwards Time Travel Involves Backwards Causation

First, a few words about the relationship between backwards time travel and backwards causation. To motivate the idea that the possibility of backwards time travel implies the possibility of backwards causation, suppose Dr. Who travels back in time from 2050 to 1950 in his TARDIS. The clothes he wears is from 2050 and may have an impact on the fashion of 1950. Dr. Who remembers the knowledge he gains in 2050 and can use the knowledge to invent thing in the age of 1950, etc… All these cases are cases of backwards causation. What is said of Dr. Who can be generalized to all possible time travelers: The present can influence the past through the mediation of the time traveler.

More importantly, there is a sense in which backwards time travel must involve backward causation. What makes the person before the time-travel, say, Dr. Who in 2050, and the very same person as the person after the time-travel, say, Dr. Who in 1950? There must be some causal continuity between Dr. Who in 2050 and Dr. Who in 1950. As Lewis emphasizes[1], the properties of the stage of the person after the time-travel must be causally dependent upon the properties of the stage of the person before the time travel. But this means that whenever someone travels back in time, there must be at least one instance of backwards causation: the properties of the stage of this person before the time travel (e.g. the stage of Dr. Who in 2050) is causally dependent upon the stage of this person after the time travel (e.g. the stage of Dr. Who in 1950).

3. Hume and Lewis on Causation

Is backwards causation possible? At least, the idea of causing something backwards in time is very much counter-intuitive: we tend to think that the cause must precedes the effect, but if so, backwards causation is simply logically contradictory [2]. Consider Hume's famous account of causation: A CAUSE is an object precedent and contiguous to another, and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other [3].

According to Hume [3], it is built into the definition of causality that the cause is 'precedent' to the effect. But in backwards causation, the effect is precedent to the cause. Therefore, if Hume's account of causation is correct, backwards causation is directly ruled out.

In order to defend the possibility of backwards causation, one must find some non-temporal account of causation. The standard non-temporal account is David Lewis' counterfactual theory of causation. According to Lewis[4], causal dependence is just a form of counterfactual dependence: an event e causally depends on another event c just in case if c had occurred, e would have occurred and if c had not occurred, e would not have occurred[4]. He then defines causation in terms of causal dependence: an event c is a cause of an event e just in case there is a chain of causal dependence between c and e.[4]. Lewis's definition of causation only appeals to chains of counterfactual dependence without resorting to any temporal order. Therefore, if Lewis' definition of causation is correct, then there seems to be nothing incoherent in the idea of backwards causation.

However, I will point out that Lewis' counterfactual theory faces what is known as "the problem of effects". I shall argue that a straightforward way to deal with this problem is to go back to Hume and introduce temporal precedence back into the definition of causation, but in this way the possibility of backwards causation is by definition excluded. I will also examine Lewis' own response to this question, and point out that this response, while circumventing the "the problem of effects", would still rule out the possibility of backwards causation. Therefore, whether we are to get around "the problem of effects" by going back to Hume or by following Lewis' own line of thought, the possibility of backwards causation should be ruled out, which entails that the possibility of traveling back in time should be ruled out as well.

4. The Problem of Effects

In his 1973 paper “Causation” Lewis presents a classic characterization of the “problem of effects” Suppose that c causes a subsequent event e, and e does not also cause c. (I do not rule out the possibility of closed causal loops a prior, but this is a different case). Suppose further that, given the laws and some of the actual circumstances, c could not have failed to cause e. It seems to follow that if the effect E had not occurred, then its cause c would not have occurred. We have a spurious reverse causal dependence of c on e, contradicting our supposition that e did not cause c [4]. The root of this problem of effects lies in the fact that, in general, causation is asymmetric — that is, in general, if event c is the cause of event e, then e is not the cause of event e (this does not mean that causal loops are absolutely impossible)— but counterfactual dependence can very easily be symmetric—that is, if event e counterfactually depends on event c, then based on the given natural laws and conditions of reality, c usually also counterfactually depends on e. But if so, a problem immediately arises: it is very likely that in some cases, the causal relationship is asymmetric but the counterfactual dependence is symmetrical. This is in conflict with Lewis’ counterfactual theory of causation, because according to Lewis, if an event is counterfactually dependent upon another, then the latter is the cause of the former, but if there is a two-way counterfactual dependence between two events, then it implies that there must be a two-way causal relationship between them. This means that in many cases where the causal relation is one-way, Lewis’s account will mistakenly predicts that the causal relation is two-way.

Lewis doesn’t provide a specific example, but it's not hard to think of such an example. Here’s one: I smashed a very brittle window with a hammer and it shattered. My action of smashing the window with a hammer is event c, and the event of the window being broken is event e. c is the cause of e, but e is not the cause of c. However, given the laws of physics in the real world, if e had not occur, then c would not have occurred either, so c is counterfactually dependent upon e. According to Lewis’ counterfactual theory, this means that e is a cause of c, but by our stipulation of the example, e is not a cause of event c. Therefore, counterfactual dependence is not sufficient for causation.

In order to solve this problem, we need to realize that counterfactual dependence by itself is not sufficient to distinguish the cause from the effect, because there can very easily be a two-way counterfactual dependence between the cause and the effect. An straightforward way to solve this problem is this: we need to presume that the effect not only needs to be counterfactually dependent upon the cause, but also needs to be temporally subsequent to the cause. But this straightforward solution reintroduces the temporal relationship into the definition of causation, and thereby excludes the possibility of backwards causation.

In “Causation,” Lewis argues that the “problem of effects” does not undermine his counterfactual account because, for him, the reverse counterfactual conditional is not true. That is to say, if c is the cause of e, then if e had not occurred, c would still have occurred. Lewis argues that the possible world in which e does not occur but c occurs is the closest world to the actual world where e and c both occur. According to Lewis’ theory of counterfactual conditionals, if c occurs in a possible world where e does not occur and if this possible world is the closest one to the actual world, then the counterfactual conditional “if e had not occurred, c would still have occurred” is true. Therefore, a reverse counterfactual dependence does not exist.

However, we must ask: why is the possible world where e does not occur but c occurs is “closer” to the actual world than a possible world where neither e nor c occurs? Lewis’ answer is as follows:

To get rid of an actual event e with the least over-all departure from actuality, it will normally be best not to diverge at all from the actual course of events until just before the time when e takes place. The longer we wait, the more we prolong the spatiotemporal region of perfect match between our actual world and the selected alternative[4].

The kernel of Lewis' solution is that he requires that under normal circumstances, to determine whether c would still have occurred under the counterfactual condition that e had not occurred, one needs to consider a possible world that is completely the same as the actual world before e takes place, but in this possible world, c must occur: because in the actual world c occurs before e takes places and this possible world is completely the same as the actual world up until e. Therefore, Lewis concludes, even if e had not occurred, c would still have occurred. This strategy circumvents the "problem of effects" because it shows that cases where there appear to be bidirectional counterfactual dependence is actually unidirectional.

While Lewis' strategy successfully circumvents the "problem of effects", it automatically excludes the possibility of reverse causation. This is because, according to Lewis' counterfactual theory of causation, if a later event c is the cause of an earlier event e, then we need to consider whether e would have occurred if c had not occur. However, if in order to determine whether this counterfactual conditional sentence is true, we have to consider a possible world that is completely the same as the actual world until c occurs, then in such a possible world, e must occur, because in the actual world e occurs before c and this possible world is the same as the actual world up until c occurs. So even if c had not occurred, e would still have occurred. Therefore, there is no counterfactual dependence between c and e. This means that no temporally prior event can be counterfactually dependent on a temporally posterior event, and this means that, according to Lewis' theory of causation in conjunction with its solution to the "problem of effects", backwards causation is impossible.

5. Conclusion

In conclusion, Lewis' theory of causation faces a dilemma: he can either resolve the "problem of effects" by introducing Hume's chronological relationship back into the condition of causation or circumvent the problem by stipulating that the closest possible world needs to be historically the same as the actual world. No matter which strategy Lewis adopts, the possibility of backwards causation will inevitably be ruled out in the end.

References

[1] *Lewis, D. (1976). "The Paradoxes of Time Travel", his Philosophical Papers, vol. 2, Oxford: Oxford University Press, 1986,pp.73-74.*

[2] *Mellor, D. H. (1998). Real Time II. London: Routledge.*

[3] *Hume, D. (2007) A Treatise of Human Nature, edited by David Fate Norton and Mary J. Norton. Oxford: Oxford University Press.*

[4] *Lewis, D. (1973). "Causation", in his Philosophical Papers, vol. 2, Oxford: Oxford University Press, 1986,pp.166-171*