Must a Cause be Earlier than its Effect?[1]

John T. Roberts, draft of November 12, 2013
(submitted to Gargnano Conference)

1. The Task at Hand

It was formerly a well-received principle that a cause cannot be later than its effect. But today many philosophers and physicists think that backwards causation is at least conceptually possible and might even be nomologically possible. One of the most important reasons for this reversal of opinion is the fact that general relativity – our best theory of space-time structure – allows for the existence of closed time-like curves. More generally, it allows for the possibility that by following some time-like or light-like curves in the forwards-pointing direction, you can end up in the temporal past of where you started. I will call curves like this 'backwards-looping' curves. It is widely held today that backwards-looping curves could make time-travel possible, and more generally, make it possible to send causal signals to the past. If the correct theory of space-time structure is like general relativity in the relevant respect, then, it seems that a cause can, after all, be later than its effect.

But the kind of backwards causation that is made possible in this way is subject to a certain limitation. An event that lies on a backwards-looping curve can be later than one of its own effects, but it must also be earlier than that same effect. For on the usual way of understanding what 'earlier' and 'later' mean in the context of general relativity, one space-time point is earlier than a second just in case the first is connected to the second by a forward-directed time-like or light-like curve. So, any cause that manages to be later than its effect by exploiting backwards-looping curves must also be earlier than its effect. Let's call any case where a cause is both earlier and later than its effect a case of *weak backwards causation*. It is another matter whether *strong backwards causation* is possible – this would be a case where a cause is later than its effect and fails to be earlier than its effect. I believe that today it is a widely held view that if general relativity – or any other theory much like it – is true, then weak backwards causation is possible, but strong backwards causation is not.[2] You can cause things in your own past, but only if it is

[1] Thanks to Melissa Schumacher, Jennan Ismael, John Carroll, Heather Gert, and David Faraci.

[2] Some evidence for my claim here is provided by the fact that the term "causal past" and "causal future" are standardly used to refer to those regions of spacetime containing points that can be reached from a given point by a past-directed or future-directed time-like or light-like curve, respectively (see e.g. Earman 1992). This terminological convention suggests that an event can only be a cause of those events that can be reached from it by traveling along a future-directed time-like or light-like curve, which is equivalent to saying that *a* causes *b* only if a is earlier than *b*. Another piece of evidence is provided by the fact that the standard "causal past" and "causal future" are standardly used to refer to those regions of spacetime containing points that can be reached from a given point by a past-directed or

possible to get there from here by traveling along a forward-directed time-like or light-like curve.

Another widely held position – which doesn't have anything particularly to do with general relativity – has to do with the relation between causation and practical rationality. It says that it can be rational for an agent to employ means M towards end E only if M might be a cause of E, as far as the agent knows. Even if M is highly positively statistically correlated with E, it cannot be rational to bring about M with the aim of bringing about E, if it is known that M is not a cause of E. (So for example, even though being rich is correlated with smoking expensive cigars, it is not rational to smoke expensive cigars in order to get rich.) This idea is of course at the heart of causal decision theory.

Putting all this together, we have three widely held views:

> (1) If general relativity – or any other space-time theory that permits backwards-looping curves – is true, then weak backwards causation (i.e. causation in which the cause is both earlier and later than the effect) is physically possible.
>
> (2) If general relativity – or any other space-time theory that permits backwards-looping curves – is true, then strong backwards causation (i.e. causation in which the cause is later, and not earlier, than the effect) is not physically possible.
>
> (3) It is rational for an agent to use some means M for the purpose of realizing end E only if, as far as that agent knows, M might be a cause of E.

I am going to argue that it is impossible for all three of these claims to be true, so we must reject one of them. I don't pretend to know which one we should reject, but in the final section I will try to motivate my opinion that (3) is the least vulnerable and should not be rejected. If I am right about that, then we have here an argument that the structure of space-time makes weak backwards causation possible only if it also makes strong backwards causation possible. The strangeness of this conclusion is illustrated by the example I will discuss in the following section.

2. A Magic Trick

---

future-directed time-like or light-like curve, respectively (see e.g. Earman 1992). This terminological convention suggests that an event can only be a cause of those events that can be reached from it by traveling along a future-directed time-like or light-like curve, which is equivalent to saying that *a* causes *b* only if a is earlier than *b.* Another piece of evidence is provided by the fact that in the standard terminology of space-time physics, a 'causal curve' is a curve that is either time-like or light-like. These linguistic conventions are evidence that it is widely assumed that causes must e earlier than their effects, but by themselves they provide no reason to think this widespread assumption is true.

Suppose that you have access to a time-like curve that loops from Friday back to Wednesday, but you do not have access to any such curves that loop back to any time earlier than Wednesday.[3] It might then seem then that on Friday, you might still be able to do something about what takes place on Wednesday, but there is nothing you can do about what happened on Tuesday or Monday; Tuesday and Monday are in a part of space-time that you cannot get at anymore. Suppose that you bought a lottery ticket on Monday, and the drawing was held on Tuesday – but let's suppose you didn't find out which ticket won. It appears that by the time Wednesday comes, there's nothing more you can do to influence your chances of winning. On Friday, you will be able to travel back to Wednesday, but you will never be able to travel back to any time earlier than the lottery drawing. I claim that nevertheless, there are actions you can take on Wednesday, Thursday, and Friday that will ensure that you have won the lottery.

Here is what you need to do: On Wednesday morning, commit yourself to the following course of action. Tomorrow, on Thursday, you will make a video recording of yourself reciting all of the possible winning numbers. (Obviously, this plan will be easier to carry out if it is a lottery with a sub-astronomical number of possible winning numbers.) You do not commit yourself to saying these numbers in any particular order, but you definitely commit yourself to saying them all while the video camera is running. You will make sure that there is a good digital clock in the frame of the video recording. You will give the completed video recording to your assistant. You will then instruct your assistant to check the newspaper to find out what the winning number in Tuesday's lottery drawing was, then watch the video recording and make a note of the exact time at which you said the winning number, and finally, send this note back to you on Wednesday, via the time-like path that loops back from Friday to Wednesday. Later today (on Wednesday), you will of course receive this message from your assistant. You will keep it with you. Tomorrow (Thursday), when you are recording yourself reciting all the numbers, you will make sure to say the number on your own ticket at the time indicated in your assistant's message.

Given the setup, your task and your assistant's task are relatively easy to carry out. All you have to do is receive a message, make a video of yourself saying a bunch of numbers, and make sure that you say one particular number at one particular time. All that your assistant has to do is read the lottery results in the newspaper, watch a boring video, notice the time at which a certain event happens in the video, write that time down correctly, and send it in a message. Nothing here is particularly difficult. But if you do your job correctly, and your assistant does his job correctly, then it follows logically that you have won the lottery. For if the

---

[3] This might be because you have a time machine that is subject to the 'Mallett limitation': It cannot be used to travel to any time earlier than the time when it was itself first turned on. (The physicist Ronald Mallett (2007) has described a mechanism for a time machine subject to this restriction.) But the condition could be satisfied even if you had no time machine at all, but knew about a wormhole connecting a spacetime location on Friday with one on Wednesday.

assistant does his job correctly, then the time mentioned in his message is the time at which you say the winning number, and if you do your part correctly, then the time mentioned in his message is the time at which you say your own number. Therefore, if you both do your parts correctly, then your number is the winning number.[4]

What is more, you can know ahead of time that this strategy will work if carried out correctly. Since you are committed to saying every possible winning number in the video recording, it is guaranteed that you will say the winning number, and so it is guaranteed that your assistant will be able to find a time at which you say the winning number. This provides the answer to the obvious question, "But why do you have to say all the possible winning numbers? Why not just say your own number at the time that your assistant sent you, and forget about the rest?" The answer is that unless you definitely commit yourself ahead of time to saying them all, you have no good reason to think that your assistant will be able to carry out his task correctly.

Of course, something might go wrong. But the fact remains that, given the availability of the Friday-to-Wednesday time-channel, you can carry out this plan at will, every part of the plan is in principle easy to carry out, and if all parts are carried out correctly then success is guaranteed. So it looks like we have here an *effective strategy* for winning the lottery, every component of which is carried out at a time that is later, and not earlier, than the drawing. It seems that if you are in the situation I have described, and you are in need of money, you have excellent reason to carry out the plan I have described as a means to winning the lottery.

3. The Argument

Recall assumptions (1)-(3):

(1) If general relativity is true, then weak backwards causation (i.e. causation in which the cause is both earlier and later than the effect) is physically possible.

---

[4] Obviously, it will take a long time to read all the possible winning lottery numbers aloud. But David Faraci has pointed out a shortcut: Just read each of the possible digits as many times as there are digits in a lottery number, then have your assistant write down a list of times: First, one time at which you said the first digit in the winning number, then one time (not identical to the first) at which you said the second digit in the winning number …. then finally, a time (not identical to any of the other times on the list) at which you said the last digit of the winning number. You have to say each digit multiple times because of the possibility that one digit will occur more than once, and you need to have your assistant plan to write down a different time for each digit in the winning number because you cannot know before carrying out the plan which, if any, of the digits in the winning number match each other.

(2) If general relativity is true, then strong backwards causation (i.e. causation in which the cause is later, and not earlier, than the effect) is not physically possible.

(3) It is rational for an agent to use some means M for the purpose of realizing end E only if, as far as that agent knows, M might be a cause of E.

If assumption (1) is true, then it should be physically possible for a situation with the same structure as the one I described above to take place. Your assistant could exploit the possibility of weak backwards causation to get a message back in time from Friday, say, to Wednesday. Apart from sending the message backwards through time, nothing else in the magic trick presents any serious difficulties at all, so it should be physically possible to carry out the magic trick.

As I pointed out above, if you have the opportunity to use this trick, and you are in need of funds, it is obviously rational for you to use it. It is a good strategy for winning the lottery. So, by assumption (3), it follows that as far as you know, the activities of yourself and your assistant while carrying out the plan might be a cause of your having won the lottery. This should still follow even if we stipulate that you know that general relativity is true, and that you know that there is no time-like curve looping back from any time on Wednesday or later back to any time on Tuesday or earlier. But if you know those things, then you know that it is physically impossible for your actions from Wednesday to Friday to be temporally earlier than your lottery victory. Yet, it is consistent with everything you know that those actions are a cause of your lottery victory. Therefore, it is physically possible that your actions are a cause of your victory, even though those actions are later and not earlier than your victory. Therefore, given general relativity, strong backwards causation is possible.

We have just shown that if assumptions (1) and (3) are true, then assumption (2) must be false; in other words, it is impossible for all three assumptions to be true.

Obviously, there is nothing special about my lottery trick here; it provides the template for a whole class of methods of bringing about past states of affairs. There is also nothing particular special about general relativity here: All that's important about it is that it permits weak backwards causation. So we can generalize the conclusion: If assumption (3) is true, then weak backwards causation is possible only if strong backwards causation is possible; the widely-held belief that backwards causation is possible but causal influence must always propagate along time-like or light-like curves in the forward-pointing direction is mistaken.

4. Is This Really Backwards Causation?

My argument obviously depends crucially on the intermediate conclusion that in the magic trick described in Section 2, the actions of your and your assistant between Wednesday and Friday could be a cause of your winning the lottery. There are a number of ways someone might resist my argument for this lemma.

One worry is that it is far from clear what the effect is supposed to be in this causal relation. You and your assistant are said to cause your winning of the lottery – but when did this effect occur? When you bought our ticket on Monday? When the winning number was selected on Tuesday? Both?

Fortunately, we don't need to settle this question here; it's enough to establish that some event occurring no later than Tuesday can be caused by an event that occurs later than Tuesday, even though there are no forward-directed time-like or light-like pathways leading back to Tuesday from later times. One possibility, though, is that the effect is neither the event of your buying a ticket with the particular number you did, nor that particular number's winning, but instead the coincidence. Compare: If I put on a blindfold and randomly select a mated pair of socks from a vat containing only mated pairs of matching socks, put the two socks in two different envelopes and mail them to two different addresses, then a certain coincidence will occur: Similarly-colored socks will arrive at the two addresses. My decision to draw from a vat containing only correctly matched pairs of socks seems to be a cause of the coincidence, though it is perhaps neither a cause of a pink sock arriving at the one address, nor of a pink sock arriving at the other. In the present case, similarly, perhaps our carrying out the plan causes the coincidence between out purchase of a ticket on Monday and the drawing of the winning number on Tuesday, without being a cause either of our buying the number we bought nor of the drawing of the particular number that was drawn.

A more serious worry about the argument is that when you and your assistant carry out the magic trick, what the two of you do logically necessitates your winning the lottery, but causes do not logically necessitate their effects. So, whatever is going on here, it seems, it could not a case of backwards causation.

However, even if the correct carrying-out of the plan fails to cause you to win the lottery for this reason, there are still causes in the picture. There is your deciding to carry out the plan; your writing a note to your assistant giving him his instructions and at the same time writing a note to yourself reminding yourself of what to do; your trying to carry out the plan. These events do not logically necessitate your winning the lottery. But they do seem to be steps taken in an effective strategy for bringing it about that you do win the lottery. So if assumptions 1 and 3 are true, it still follows that strong backwards causation is consistent with general relativity.

Another objection runs as follows: Obviously, on Wednesday morning, either you have won the lottery or you have not. If you have not, then obviously, something will go wrong with your plan, for it is logically impossible for you to carry your plan out correctly and not win the lottery. In other words, you will not be able to correctly carry out your plan unless you have already won the lottery. It is not the case that your carrying out the plan causes you to have won the lottery; rather, your having won the lottery is a necessary condition for the possibility of your carrying out the plan.

As formulated, this objection commits a modal fallacy: 'It is impossible that both ~P and Q; therefore, unless P, it is impossible that Q.' Perhaps it is possible to reformulate the idea behind this objection without committing this particular fallacy. Suppose that it is. In that case, the argument shows not only that there is no

backwards causation in the case described in section 2, but also that there is no backwards causation in ordinary, garden-variety cases of time travel. Suppose that you exploit a closed time-like curve to visit the Jurassic period. Is your travelling through the time-loop a cause of the appearance of a human being during the Jurassic? Well, if no human being had ever shown up during the Jurassic, then obviously you would not succeed in travelling through the time-loop. So, by the same reasoning used by the objection, it is not the case that your travelling through the time-loop causes your appearance in the Jurassic; rather, your appearance in the Jurassic is a necessary condition for the possibility of travelling through the time-loop. If closed time-like curves really do permit genuine weak backwards causation, then this objection must fail somehow, and so (presumably) does the parallel objection to the claim that there is backwards causation in the lottery case. So it appears that the objection undermines the conclusion that strong backwards causation is possible only if it also shows that weak backwards causation is impossible; that is, it undermines the argument against assumption (2) only if it provides a reason to reject assumption (1). So my main conclusion, which is that one of (1)-(3) must be false, is not undermined.

Let's consider one final objection to the claim that in the magic trick, you and your assistant cause your lottery victory. This objection alleges that the actions of you and your assistant on Wednesday, Thursday and Friday could not be among the causes of your winning the lottery, since that already has a sufficient cause that occurred entirely before the end of Tuesday – namely, the complete state of the universe on Monday morning, which determines what ticket number you will buy and also what number will be drawn. So it couldn't also be caused by what you do between Wednesday and Friday.[5] That would be a case of mysterious causal overdetermination.

But would it really be so mysterious? Let A be the state of the universe on Monday morning, let B be your actions between Wednesday and Friday, and let C be your lottery victory. So we have here a case where A is a sufficient cause of C, and B is also a sufficient cause of C. That can makes it look as if there must be some sort of weird pre-established harmony, making sure that the different sufficient causes don't get in each other's way. But that's so only if A and B are causally independent of one another. And in the case at hand, they aren't: Assuming determinism, the total state of the universe on Monday morning is a sufficient cause of your actions between Wednesday and Friday, so the structure we have here is this: A is a sufficient cause of B, B is a sufficient cause of C, and A is a sufficient cause of C. That's just the structure of a chain of sufficient causes; there's nothing mysterious about that at all.

5. Is It Really Rational to Try to Use This Magic Trick?

The argument of section 3 also depends crucially on my claim that in the situation described it would be rational for you to employ the magic trick as a means

[5] For this objection, I am grateful to Melissa Schumacher.

of winning the lottery. There are some reasons why someone might doubt this claim.

One objection is as follows. Given what you know, you should find it extremely unlikely that the plan will succeed. For you know that the plan will succeed if and only if your lottery number is the winning number, but you also know that the objective chance of your number being drawn was extremely low. Therefore, you should know that the objective probability of your being able to carry out the plan successfully is also very low. Even if it seems unlikely that you and your assistant will fail to carry out the relatively easy tasks prescribed by your plan, it is far more unlikely that you have the winning lottery ticket. So it would be foolish of you to put this plan into action for the sake of trying to win the lottery.

On closer inspection, though, it is not so obvious that what this objection alleges is true. When you conditionalize on one part of your background information – namely that you carry out your plan successfully only if you win the lottery, and the lottery is a fair one with thousands of possible winning numbers of which yours is but one – then it seems you should assign a very low probability to the proposition that your plan will succeed in bringing about its end. But when you conditionalize on a different part of your background information – namely that you and your assistant are both quite competent, that the plan calls upon each of you to carry out only very easy tasks, and that correct carrying out of the plan logically guarantees success – then it seems that you should assign a very high probability to the proposition that your plan will succeed. When you take into account all of your background information at the same time, it is far from clear how likely you should find the prospect of success.[6]

Moreover, even if this objection succeeds against the case as described in section 2, that case can be modified in a way that gets around the problem. Make the lottery be one with very few tickets – four, say. Then the probability that you will fail to win the lottery is 0.75. If you and your assistant are suitably skillful, the probability that you will carry out your plan correctly can easily be higher than that. In this version of the case, there is not even an initial appearance that you should find it unlikely that your plan will work. (It may seem improbable that you will win this lottery, but it is even more improbable that you and your partner will fail to do your jobs right – exactly the reverse of how things seemed in the original case.) But in this case, too, we have an effective strategy for bringing it about that you won a lottery that was finished on Tuesday, consisting entirely of actions taken outside the

---

[6] It might be thought that the Principal Principle (see e.g. Lewis 1980) settles this question, and that it settles it by saying that your credence should be quite low, since you know that the objective chance of your winning is quite low. But the Principal Principle does not apply here: You have information that is inadmissible with respect to whether you won the lottery or no – namely, the information that you and your assistant are prepared to carry out the magic trick, and that you will carry it out correctly iff you have won the lottery. This information is inadmissible because it has evidential bearing on whether you have won the lottery, but no evidential bearing on what the chance of your winning the lottery is.

temporal past of Tuesday. So this case would be enough to establish that, if backwards causation is possible at all, then strong backwards causation is possible.

In my opinion, the most threatening objection to the argument of section 2 I have seen is the following. Suppose that you carry out the strategy, and you do win the lottery. Well, then, on Monday you bought a ticket with number N, and on Tuesday N was selected as the winning number, so even if you hadn't bothered to carry out the strategy, you still would have won the lottery. Your efforts will have been superfluous. But the following seems intuitively obvious:

> The Superfluity Principle: If you know that a strategy S for achieving some goal G is such that, in any possible case in which you carry out S and achieve G, your carrying out S is superfluous (in the sense that you would still have achieved G even if you had not carried out S), then it cannot be rational for you to use S for the sake of G.

After all, a strategy to which this principle applies is one that will never be used in any circumstances in which it is needed. Such a strategy will not ever do anyone any good. And a strategy that you know will never do anyone any good cannot be a strategy that it is rational to employ. Right?

No; the Superfluity Principle is false. To see why, consider a different case: I am informed by a source whom I know to be a reliable informant about the future (maybe a god, maybe a psychic, maybe a seasoned time-traveler) that throughout the rest of my life, whenever I buy a book, a friend or loved one will coincidentally have just bought a copy of the very same book as a present for me and will give it to me within the hour. So, the strategy of buying a copy of a book in order to come to possess a copy of that book satisfies the antecedent of the Superfluity Principle: I know that whenever I use it successfully, it will turn out to have been needless. Nevertheless, it can obviously be rational for me to employ this strategy. I want a copy of *Encyclopedia Brown Foils the NSA*; I've been dropping hints all over the place, but nobody has bought me a copy yet; I keep waiting and waiting; I tell myself that there's no point in buying a copy, because as soon as I do, someone else will give me one; still, time goes by, and no one buys me a copy. I begin to realize that the evidence strongly suggests that nobody will ever buy me a copy, unless I buy one for myself. Under the circumstances, it is obviously rational for me to go buy a copy, even though I know that my doing so will then immediately prove to have been needless. Anyone in this situation who remained bookless for his whole life simply because of his unwillingness to sin against the Superfluity Principle would clearly be a fool. Therefore the Superfluity Principle is false.

In the lottery case, the magic-trick strategy is rational in spite of the Superfluity Principle for the same reason: You know that if you carry it out successfully, then it will turn out that you would have won the lottery anyway, but the only way within your power of making sure that you are in a case where you would have won the lottery anyway is to carry out the strategy.

6. Conclusion

The argument shows that one of (1), (2), and (3) must be false. I find it very implausible that (3) is the false one. (3) expresses a relation between causation and rational agency that seems to provide one of the main reasons—and perhaps the primary reason – why we have ever been interested in distinguishing between causation and mere correlation in the first place. So to let go of this connection is to let go of much of the point of having a concept of causation at all. It would be to modify our concept of causation in a way that would amount to changing the subject.

If I'm right about that, then we have to let go of either (1) or (2). What this means is that we have to reject the widespread view I mentioned back at the beginning: Namely, that if a theory like general relativity is true, then it is possible for there to be backwards-in-time-causation, but only when the cause is earlier than the effect as well as later than it. That seems not to be so: Either backwards causation is not possible at all, or else strong backwards causation is just as possible as weak backwards causation.

Works Cited:

Earman, John (1992): *World Enough and Space-Time*, MIT Press.

Lewis, David (1980): "A Subjectivist's Guide to Objective Chance," in R. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, University of California Press.

Mallett, Ronald (2007): *Time Traveler*, Basic Books.