Journal of the American Philosophical Association (2026) 91–109 © The Author(s), 2025. Published by Cambridge University Press on behalf of the American Philosophical Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.
DOI: 10.1017/apa.2025.10021

Causal Loops and Direct Self-Causation

ABSTRACT: *Causal loops are circular chains of causally related events: each link causes others which in turn cause* it. *Not only are causal loops widely accepted as coherently conceivable; some are also provably self-consistent as well as seeming genuinely possible according to currently accepted laws of physics. On the common assumption that causation is transitive, each link in any causal loop would wind up causing itself; but the idea of self-causation is pretty much universally rejected as incoherent. A popular attempt to resolve this dilemma distinguishes "direct" from "indirect" self-causation: the direct variety, which operates without the aid of causal intermediaries, is claimed to be impossible even if the indirect variety isn't. I argue against this attempted resolution on the grounds that causal loops themselves, unlike the links that compose them, should be viewed as directly self-caused; so indirect self-causation via causal loop is possible only if direct self-causation is as well. An important consequence is the availability of groundbreaking solutions to several longstanding puzzles in philosophy of mind.*

KEYWORDS: causation, causal loops, self-causation, *causa sui*, free will, libertarianism, consciousness, self-awareness

1. The Dilemma

We tend to think it impossible for any event[1] to cause itself: a bomb's exploding cannot cause that very bomb to explode. Although *part of* a temporally extended explosion can cause later parts, such "piecemeal" causation is what David Lewis calls self-causation "only in a derivative sense" (1986a: 173). Disbelief in self-causation reaches as far back as Aristotle (1984a: §8.4) and Hume (1748: Part I), perhaps most forcefully captured by Nietzsche: the "*causa sui*" is "absurd", a "self-contradiction", and a "perversion of logic" (1886: §21). Recent examples are too numerous to mention: Chris Weaver (2018: 110) "can hardly find anyone who rejects the formal asymmetry of causation", and asymmetry entails irreflexivity. However, so-called "causal loops"—circular chains of events each link of which causes others which return the favor—would prove otherwise, were such things possible. On the common assumption that causation is transitive—which Ned Hall notes is widely taken to be a "bedrock datum, one of the few indisputable a priori insights we have into the workings of the concept" (2000: 198)—events causing other events that return the favor wind up causing themselves. Even if one does not

[1] This article takes events to be the fundamental causal relata, but the same issues could be restated in terms of causation between things, properties, facts, or any other contending entities.

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

agree to transitivity as a general principle (see for instance McDermott 1995 and Hitchcock 2001), there is a clear sense in which any link in a causal loop should be listed in a full recounting of its own causal history.

The most-discussed kind of causal loop is generated by a person or object traveling back in time; but any kind of backward causation (i.e. in which a cause postdates its effect) can do the trick. Alternatively one can imagine a causal loop grounded in "instantaneous" (aka "simultaneous") causation, as Quentin Smith describes: "a state Sl being caused by another state S2, with the existence of S2 being simultaneously caused by S1" (1999: 581). Such loops are coherently conceivable to whatever extent backward or instantaneous causes are; and while instantaneous causation is only occasionally endorsed, "[c]ompatibility with backwards causation is a requirement that many philosophers now demand" from any potentially successful analysis of the causal relation (Dowe 2000: 188). Although our imaginations are notoriously fallible judges of logical possibility, Echeverria et al. (1991) found support for the logical possibility of at least some causal loops by proving the mathematical self-consistency of a pool ball falling into a wormhole after being knocked into a pocket, traveling back in time and back onto the table, and knocking its future self into the pocket. Perhaps most importantly, the *nomo*logical possibility of at least some causal loops follows from an unadulterated version of Einstein's overwhelmingly empirically corroborated general theory of relativity (GR). GR tells us that spacetime can bend and so can create "closed timelike curves" (CTCs) through which a causal chain could connect to its own earlier links: Gödel (1949) proved full CTCs to comply with the constraints GR places on possible spacetime structures, and was followed by Tipler (1974) and many others. Granted, some physicists speculate that GR should be modified in ways that would rule out causal loops: most famous along these lines is Stephen Hawking's (1992) "chronology protection conjecture". However, such speculation is rarely fueled by any empirical shortcomings on GR's part: again, the precision of GR's predictions of astronomical observation is unprecedented.[2] Instead, such speculation tends to be fueled by an independently motivated aversion to causal loops (or, as an anonymous reviewer helpfully pointed out, to other controversial entities like wormholes, Gott strings, or Ori-Soen time machines). Despite directly contradicting many of our commonsensical ideas about space and time, GR is widely trusted as revealing features of our world we didn't used to believe possible: the actual warping of spacetime in the way predicted by GR was first confirmed by Eddington in 1919 and continues to pass increasingly precise tests to this day (e.g. Touboul et al. 2022, Abuter et al. 2020, Takamoto et al. 2020, and Collett et al. 2018). In fact, GR opens the possibility of causal loops in a second way by allowing Einstein-Rosen bridges (aka wormholes) through spacetime. And although certain types of CTCs (such as ones arising from the particular spacetime structure imagined by Gödel) evidently do not exist in our actual universe, we've found no empirical evidence against the existence of wormholes.

In any form, causal loops bring self-causation in their wake. The event of Echeverria's pool ball falling into the wormhole, for instance, must be listed

[2] Also granted (as an anonymous reviewer pointed out), some recent attempts to adapt GR into the quantum realm are less friendly toward CTCs (e.g. Barrow & Dąbrowski 1998, Dyson 2004).

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

among the sum total of events making up its own causal history: the ball falls in at least partly because it travels back in time and collides with its later self. Which raises a dilemma: at least some instances of self-causation are coherently conceivable, provably self-consistent, and compatible with currently accepted laws of physics; yet still we recoil at the idea. At the end of this article I'll mention two of many groundbreaking benefits the idea of self-causation could have for philosophy of mind; unfortunately, the most common response to the above dilemma is to deny that causal loops entail self-causation in its most impactful sense. I will argue against that response.

2. Direct vs. Indirect Self-causation

The common response to the above dilemma, which I am going to argue against, begins by distinguishing "indirect" from "direct" self-causation. Links in causal loops are only *indirectly* self-caused, in that they cause themselves by causing other events which return the favor; while directly self-caused events (if there could be such things) would cause themselves without any outside help. With this distinction in hand, one can allege that direct self-causation is impossible even if (though) indirect self-causation isn't:

> [E]ven if it is true, as some have argued, that there could be causal loops, it is surely impossible for any event to be the immediate cause of itself. (Tooley 1990: 191)

One might in particular think that direct (what Tooley calls "immediate") self-causation is more objectionable than indirect self-causation generated through closed timelike curves (CTCs)—since CTCs are scientifically respectable according to GR. Toby Friend, for instance, hopes to reconcile the "natural assumption that events are not '*causa sui*'" with the acknowledgment that GR "may ultimately force us to relax our belief that nothing could possibly be *causa sui*", by noting that

> … the self-causing systems made available with closed timelike curves will still be mediated by temporal-like distance — i.e. where the cause exists in its own future lightcone. (Friend 2019: 5072, n. 14)

David Lewis, perhaps unsurprisingly, pioneered a version of this strategy. By distinguishing causation from "causal dependence", and identifying the former as the "ancestral" of the latter (the ancestral of any relation R is R's transitive closure: any elements related by a chain of R relations are related by R's ancestral), Lewis allowed for self-causation via causal loops despite maintaining that "events which are not distinct cannot stand in causal dependence" (Lewis 1986b: 259). What I'm calling a "directly" self-caused event would have to causally *depend* on itself, since such causation wouldn't work by way of external causal intermediaries; but given Lewis's definition of causal dependence, that's impossible. As a consequence:

> [N]o event can be self-caused unless it is caused by some event distinct from it. (Lewis 1986a: 213)

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

Thus Lewis claims to be able to accommodate causal loops, each of whose constituent links is only indirectly causally related to itself, while reinterpreting our intuitive aversion to self-causation as an inaptly phrased aversion to *direct* self-causation:

> [N]o event depends causally on itself […] However, I do allow that an event may cause itself by way of […] closed causal loops […] Thus I have taken care not to rule out the sort of self-causation which appears in time-travel stories that I take to be possible. (Lewis 1986a: 212–213)

It is worth noting that Lewis's proprietary counterfactual analysis of causal dependence is irrelevant here: any competing analysis (e.g. in terms of regularities, probabilities, interventions, causal processes, or what have you) could similarly distinguish what could be called the "root" variety of the causal relation from its ancestral, and rule out direct self-causation by stipulating that nothing can be its own "root" cause.

The rest of this article will argue against the above (call it a "Lewisian") strategy for staving off commitment to direct self-causation. My thesis will be that that strategy is ultimately untenable, because causal loops themselves should be considered directly self-caused—so indirect self-causation is possible only if the direct variety is as well. I'll begin by defending that thesis for a special category of causal loops, namely ones that aren't caused by any external events. I'll then defend the unrestricted thesis by way of arguing that every causal loop is caused by each of its individual links. I'll finish by describing two of many groundbreaking consequences the possibility of direct self-causation might have for philosophy of mind, if mental events in particular can directly cause themselves.

3. First Argument: Even Spontaneous Loops Are Caused

Note first that, in addition to being composed of smaller constituent events, causal loops arguably qualify as events in their own right—and as such are either caused or not. Events needn't be instantaneous things (if instantaneous events are even a thing): a war is as much of a happening as any of its constituent battles, each of which is as much of a happening as any of its constituent bullet-firings (and so on). And since events can take time, then since some timelines apparently can loop backward onto themselves, some events can presumably span loopy timelines. This would be compatible with Lewis's idea of events, for instance, as classes of spacetime regions (1986b: 244). Echeverria's pool ball falling into the pocket, traveling back in time, dropping onto the table, and knocking its future self into that same pocket, for instance, constitutes what we might call a "loopy" event. Accordingly, if causal loops count as events, it makes sense to ask, of any particular loop, whether (like most events) it has any causes—or whether, instead, it is entirely uncaused (like, say, radioactive decays[3]). Note that this is an entirely different question from asking, as

[3] Apparently radioactive decays *are* caused—by "vacuum fluctuations". However, vacuum fluctuations are in turn uncaused.

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

for instance Lewis (1976) does, whether any causal loop can be given an "ultimate explanation"—in the sense of our being able to list all of the causes of all of its causes, or perhaps identify a unique cause from which all of its others result. Ordinary—e.g. externally determined—events certainly are caused, even if they are on a par with random quantum events with respect to lacking any such "ultimate explanation" (as Lewis maintains; see also Meyer 2012).

So-called "open" loops bear causal relations to events outside themselves. The loop consisting of the pool ball's trajectory is "open", because among its causes is my previously having hit the ball with a pool cue. To borrow for a moment Lewis's counterfactual test for causal dependency, if I hadn't first hit the ball with a cue then it wouldn't have wound up in the proper location to collide with its time-traveling self and get knocked into the wormhole. My arguments in the upcoming section will apply equally to such open loops; but I want to start by arguing in this section for self-causation in the special case of "closed" loops (meaning loops which bear no causal relations to anything outside themselves). In particular we need only consider loops that have no external *causes* (whether or not those loops have any external effects): call such at-least-partially closed loops "spontaneous", in the Kantian sense. To avoid unnecessary complications, we can ignore loops containing any link that is itself uncaused, like a radioactive decay: a spontaneous causal loop, in the sense I mean it, is composed of links each of which is caused by—but *only* by—links in that loop. If I can show that that kind of causal loop is self-caused, then the Lewisian strategy for avoiding commitment to direct self-causation fails (because, again, spontaneous loops have no external causes, while indirectly self-caused events cause themselves by way of causing other events which return the favor; so if spontaneous loops are self-caused then they must be *directly* so). My argument runs as follows:

1. Spontaneous loops have no external causes.
2. Spontaneous loops are caused.

Conclusion: Spontaneous loops are self-caused.

Premise 1 is definitional, and in the upcoming section I'll argue that my conclusion follows from 1 and 2. I'll argue now for premise 2.

History speaks against premise 2, since causal loops are typically described as *un*caused. Andrei Lossev and Igor Novikov (1992) coined the term "Jinn" for entities born from causal loops, and they speak of all Jinn as being uncaused. Although that seems to me obviously false for things like *links in* loops—which by assumption are caused by other links, and hence caused *simpliciter* (even if, again, they lack any "ultimate explanation" in Lewis's 1976 sense)—the claim is both more common and admittedly more compelling when restricted to entire loops as wholes. Lewis, for instance, finds it obvious that "causal loops that arise in time travel" are "uncaused and inexplicable": "There is simply no answer" to the question "Why did the whole affair happen?" (1976: 148) Serguei Krasnikov (2001: n. 5) labels as "acausal" all of Lossev and Novikov's "Jinn", explicitly including Harry Harrison's (1967) "self-sufficient loops". Ryan Wasserman agrees that in any loop that's not externally

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

caused (which I've called "spontaneous") "we seem to have an example of a causal loop without a cause." (2018: 159) Ulrich Meyer accepts that "to request a 'full' explanation of causal loops, is to ask for something that is impossible"—where a "full" explanation provides "a *causal* reason" why that thing occurs (2012: 262–263). Meyer accepts that causal loops nonetheless "admit all the explanation one could reasonably ask for" (ibid.: 260), since each of their *links* is caused; but again, I am asking only whether causal loops themselves *have causes*, and his answer to *that* question is "No". (For instance, Meyer compares a causal loop to an electron that has always existed.) If, as all of these authors believe, causal loops (at least of the spontaneous variety) are entirely uncaused, then they cannot be *self*-caused—as I claim them to be.

But let me ask: what, in general, qualifies any particular causal chain as being caused rather than uncaused? An example of an *un*caused causal chain would be a series of causally connected microphysical events the first of which is a random radioactive decay: that temporally extended event (beginning with the decay) is itself presumably uncaused, since its first link is. In contrast, consider an otherwise similar causal chain that undeniably *does* have a cause: the toppling of a domino chain, when someone pushes the first domino with their finger. What makes that toppling event as a whole qualify as caused rather than uncaused? Without needing to dive into the relative merits of regularity, counterfactual, or other theories of causality in general, we can simply note that there is an event (the finger-pushing) that causes all of the constituent dominoes to fall; and it does so by causing one domino (the first) to fall, which in turn causes all the rest. Whereas in the case of the randomly-triggered microphysical chain, nothing at all causes the radioactive decay which constitutes the chain's initial link. So I propose the following principle:

> P1. A particular causal chain is caused if there is an event that causes each of that chain's constituent links. In particular, if there is an event that causes one of the chain's links which causes the rest.

Crucially, P1 is satisfied for causal loops—even ones I'm calling "spontaneous", which by definition have no external causes. Point to any link in any particular causal loop, and you'll be pointing to an event that causes every link in that loop (including, crucially, that very link itself); furthermore, whichever link you point to causes all of the others by causing one in particular (namely, whichever link is the next one in line), and that next one causes all of the rest. Therefore even spontaneous causal loops have causes.

Am I overstepping by applying a commonsensical principle about causal *chains* to causal *loops*? I don't think so. First of all, causal loops *just are* causal chains—and unlike our microphysical chain, the first link of which is a random radioactive decay, causal loops are chains in which each and every link is caused. Granted, causal loops are *loopy* causal chains: unlike any ordinary chain, their last link causes their first (instead of causing some event outside the chain)—or equivalently, their first link is caused by their last (instead of being caused by some outside event). I'm not denying that this makes causal loops unique among causal chains: of course they boast certain properties that ordinary chains do not (I will claim that one of those properties is

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

being self-caused). But that doesn't mean they stop behaving certain other ways we can reasonably expect all causal chains to behave. Suppose that any particular marching band is beloved as long as someone in the world loves every member, and suppose that anyone who marches in the past footsteps of any particular band member winds up loving them. Any marching band followed for long enough by someone who is also marching will wind up being beloved, on the assumption that the follower will eventually march in every member's past footsteps. But even a marching band with zero followers will wind up being beloved if it forms a circle, assuming that in that case each member will *themselves* walk in every member's past footsteps—including, crucially, their own. In either case, every band member, and hence the band as a whole, is beloved: it makes no difference that in the latter case all the love comes "from the inside". Analogously, P1 talks only about whether or not a particular causal chain *is caused*, i.e. has at least one cause in the world. It makes no difference to *that* question whether or not that cause comes "from the inside".

Consider once again a domino chain. Instead of someone pushing the first domino with their finger, the triggering cause of the ensuing toppling-event—by which I mean whatever causes the first domino in the chain to fall, and hence is a cause of all of them falling—could have been Merlin's waving a magic wand, or God's proclaiming that the first domino must fall … or even, in a suitably curved spacetime, the *final* domino's falling. Suppose that last possibility were in fact the case: the toppling dominoes form a causal loop. Why should the status of the entire toppling-event as caused rather than uncaused be in any way adversely affected, just because the first link is caused by the last link instead of by the movement of someone's finger? The fact that there is at least one event that causes every link in the ordinary chain-toppling, instead of the first of those links occurring randomly, is plausibly what makes the entire toppling-event as a whole count as caused; and the loopy domino chain satisfies that same condition. Granted, again, the ordinary chain-toppling is caused by events outside itself (in this case a finger-pushing), whereas the toppling of the loopy chain isn't: but *what* causes either event is irrelevant to *whether* they are caused at all.

Does it matter that in the case of the ordinary domino chain, a unique event stands out (namely, the finger-pushing) as allowing the causal chain in question to satisfy P1 and hence qualify as caused; whereas in the loopy case, each and every one of the loop's links individually can fulfill that role? The short answer is "No": having more than one candidate cause does not make an event *un*caused, no matter how else one might analyze the situation. Imagine if two people's fingers simultaneously pushed the first domino in an ordinary domino chain: no matter which events one might or might not consider to be a cause of the chain's subsequent toppling, saying that it has *no* cause is out of the question. Many philosophers prefer "individualism" regarding such causal overdetermination, holding that each of the two finger-pushings qualifies as a separate cause of the toppling (see Schaffer 2003: 26 for a list, including Schaffer himself), while others lean toward "collectivism", holding that only the combination of the two pushings as a whole qualifies as a cause (e.g. Loeb 1974, Mellor 1995, Hausman 1998; see Schaffer again). Regardless, *no one* would say that the toppling has no causes at all. So the mere fact that causal loops satisfy P1 thanks to the contribution of each and every one of its links individually cannot mean that causal loops are uncaused (whatever else one takes it to mean).

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

Could one argue that what qualifies the ordinary domino chain's toppling as caused is that its *temporally first* link is caused—while the loopy toppling has no such first link? In fact such a loopy causal chain presumably would—and most causal loops of the kind typically discussed *do*—have a temporally first link: in particular what I've called "temporal" causal loops that work through backward causation, for instance through closed timelike curves (CTCs), typically have a unique link that is objectively earliest (see for instance Wasserman 2018: 161–162). And even if the current argument worked *only* for that kind of temporal loop—and not also, say, for non-temporal loops of the kind we previously saw imagined by Smith (1999), which work through instantaneous causation among simultaneous links and hence have no earliest link—it would still follow that *some* causal loops are caused as wholes. Since we'll soon see that it is a short step from there to the conclusion that any such loop is *directly self*-caused, and since the Lewisian cannot accept any instances of direct self-causation at all, the Lewisian, at least, cannot take this route.

Might one resist the move from premises 1 and 2 to my conclusion by suggesting that spontaneous loops are caused only by their proper parts? In that case spontaneous loops would not strictly speaking be *self*-caused, in the sense of being caused also by their whole selves. In the upcoming section I will argue against that possibility: if a causal loop is caused by each and every one of its constituent links, then it is also caused by its whole self. However, for now note simply that the Lewisian in particular cannot take this route, because they cannot accept part-whole causation any more than they can accept direct self-causation. Wholes do not causally *depend on* their proper parts—since Lewis defines "causal dependence" as denoting a relation that can hold only between distinct events, and

> … events are distinct if they have nothing in common: they are not identical, neither is a proper part of the other, nor do they have any common part. (Lewis 1986a: 212)

Nor are wholes connected to their proper parts by any *chain of* such causal dependence; so parts cannot cause the wholes in which they appear. A causal loop, therefore, cannot be caused by any of its constituent links. In fact Lewis (1986b: 259) expressly opposes part-whole causation just as strongly as he opposes *direct* self-causation: in a response to Jaegwon Kim's (1973) objection to to a counterfactual theory of causation, Lewis touts the fact that his account allows us to avoid judging someone's writing"rr" as a cause of their writing "Larry" (even though the latter wouldn't have occurred if the former hadn't, which is a sufficient condition for causal dependence between any events that can possibly instantiate the dependence relation). As I already mentioned, we'll see in the upcoming section why Lewis is wise to resist part-whole causation, given his desire to resist direct self-causation—since the path from the former to the latter is fairly quick.

To finish my argument: spontaneous loops are caused, but not by anything outside themselves; so they must be self-caused. Yet they cannot be *indirectly* self-caused, since, again, by definition spontaneous loops have no external causes. So, since spontaneous loops are self-caused, but not indirectly, they must be *directly* self-caused.

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

4. Second Argument: From Part-Whole to Direct Self-Causation

My second argument that causal loops are directly self-caused applies to any variety of causal loop—spontaneous or externally caused, or otherwise open or closed. The meat of the argument establishes that causal loops are self-caused *simpliciter* (we'll see afterward why it follows that they are *directly* so):

1. Causal loops are caused by each of their constituent links.
2. If (1), then causal loops are self-caused.

Conclusion: Causal loops are self-caused.

Premise 1 should be clear from the previous section, since links in a causal loop are events whose causal contribution to that loop make it count as caused. To see this more clearly, return to our ordinary domino chain and ask: what makes the person's pushing the first domino with their finger a cause of the temporally-extended chain-toppling event? Again without diving into various theories of causality, we can simply note that the finger-pushing plausibly counts as a cause of the chain-toppling as a whole because it causes all of the individual dominoes to fall. And it does so by causing one domino in particular to fall, which in turn causes all the rest. So I propose the following modification of our previous principle P1:

> P2. A particular event is a cause of a particular causal chain if that event causes each of that chain's constituent links. In particular, if that event causes one of the chain's links which causes the rest.

Now consider any *loopy* causal chain: call it "ABCDE", such that in addition to the event named by each letter causing the next, E causes A. P2 tells us that each and every link in ABCDE causes the loop as a whole: A, for instance, causes each of the other labeled links as well as itself; and it does so by virtue of causing B, which in turn causes all of the others. B, likewise, causes each of the other links as well as itself, by causing C which causes the rest; and so on for each and every link. So on a commonsensical view of what is required for any particular event to count as one of the causes of any particular causal chain, links in loopy chains qualify as causes of the wholes in which they appear (that's my premise 1).

Is this endorsement of part-whole causation unacceptable? As I already mentioned, Lewis touts his ban on part-whole causation as a way to avoid absurdities such as having to judge someone's writing "rr" as a cause of their writing "Larry" (1986b: 259). But there is an obvious way to avoid such absurdities without barring part-whole causation across the board: don't permit part-whole causation in any case in which the part does not cause *every part of* the whole. The person's writing "rr", after all, is not a cause of their writing *each and every one of* the letters in "Larry". Even Lewis might regard the person's writing "rr" as a cause of their writing the final "y" (since those are distinct events, and the latter wouldn't have happened if the former hadn't); but at the very least the person's writing "rr" should not be considered one of the causes of their previously writing "La", since this is presumably an ordinary situation which involves

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

no backward causation. And since their writing "rr" is not a cause of their writing the first two letters, it cannot be a cause of their writing of the name considered as a whole. But imagine, instead, that the person's writing "rr" were not *only* a cause of their writing the final "y" (as it might actually be judged to be), but also, thanks to some fortuitous backward causation, a cause of their previously writing "La" *and* (hence) of their writing "rr" itself. Wouldn't we, in that case, count the person's writing "rr" as among the causes of their writing the name as a whole? What more would be required, for any particular event to count as a cause of the name-writing event considered as a whole, than for that particular event to cause the person's writing of each and every one of the five constituent letters? Yet precisely that kind of relation obtains between any link in a causal loop and the loop as a whole: each link causes not only all of the others, but also itself. So under this commonsensical condition on part-whole causation, we should count each link in any particular causal loop as causing the loop in which it occurs—while that same condition explains quite reasonably why someone's writing "rr" (in any ordinary situation) should not be counted among the causes of their writing "Larry".

Granted, part-whole causation is decidedly strange. But that strangeness, I believe, is nothing beyond the strangeness of indirect self-causation via causal loops. Once you accept the possibility of causal loops, plausible ideas about the causal relation force you to accept (as Lewis does) the reality of indirect self-causation; and once you accept that, it seems to me, equally plausible ideas should lead you to also accept part-whole causation holding between links in causal loops and the loops in which they occur. What seems to me even stranger would be to insist that an event that causes every link in some particular causal chain does not count as causing the entire chain itself. And the Lewisian is committed to precisely that uncomfortable position: each and every link in any causal loop causes not only the rest of that loop, but also itself; yet it cannot be regarded as causing the entire loop considered as a whole—despite the fact that causal loops *just are* the sum total of their constituent links. So the Lewisian view not only fails to recognize the crucial difference between causal loops and uncaused processes like ones beginning with radioactive decays (as I stressed in section 3); it also fails to recognize the seemingly undeniable causal relation borne by links in loops to the loops in which they occur.

Historically, granted, part-whole causation is shunned almost as universally as self-causation (though see Friend 2019). And it is easy to understand why, once one sees how quickly we can move from the former to the latter. My second premise delivers that move:

> 2. If causal loops are caused by each of their links, then they are self-caused.

In general, of course, one should be wary about moving from facts about proper parts of a thing to conclusions about the whole. As the so-called "fallacy of composition" reminds us, every proper part of a thing can have properties lacked by the whole of which they are a part (trivial example: *being a merely proper part*). But when discussing *causal* properties in particular, as I am doing, such a move seems less risky.

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

In particular, the needed inference is of a kind that generally appears reasonable when moving from causal facts about spatiotemporally contiguous events to similar facts about the temporally extended whole in which those events constitute temporal parts. For instance: if every discernible segment of a years-long war is among the causes of the enemy's eventual surrender, then surely the war in its entirety is a cause as well. There is no fallacy of composition at work here: causal chains are composed of their links, so to say that something is caused by each link in a chain *just is* to say that the chain as a whole causes it. At the very least, such an inference is intuitively compelling; so, I claim, is premise 2. Good reason would have to be provided to *reject* that inference, independent of pointing out that some structurally similar inferences would be fallacious in other contexts.

Imagine that our ordinary domino chain's final domino falls onto a light switch and turns on the light. The toppling of the domino chain as a whole, in that case, is presumably among the causes of the light's turning on. And again without diving into theories of causality in general, the reason why is fairly clear: each and every one *of* the dominoes' toppling is a cause of the light turning on. And they are so because the light's turning on is caused by the final domino's toppling, which is caused by all of the prior ones—and causation is transitive. So I propose the following counterpart to principle P2:

> P3. A particular causal chain is a cause of a particular event if each one of that chain's constituent links is a cause of that event. In particular, if the event is caused by one of the chain's links which is caused by the rest.

Given our previous result that causal loops are caused by each of their constituent links (and our earlier assumption that causal loops are events), P3 gives the result that causal loops are self-caused. Our previous result was that loops are caused by each of their links; now we add that whatever is caused by each of a thing's parts is also caused by the whole.

Assuming we stick to causal chains that qualify as temporally extended events (which I don't see as a significant ask), P2 and P3 can be combined in an illuminating way. Call the temporally-extended toppling of our domino chain's first half "C1" and the toppling of its second half "C2": plausibly, C1 causes C2. The reason is a combination of our previous reasons for P2 and P3: every one of the first half's dominoes' toppling is a cause of every one of the second half's dominoes' toppling. In particular, the second half's first domino topples all of the others in its half, while the first half's last domino is toppled by all of the others in *its* half; and the first half's last domino topples the second half's first domino. So combining P2 and P3, we get the following principle:

> P4. Causal chain C1 is a cause of causal chain C2 if all of C1's constituent links cause all of C2's. In particular, if a link in C1 that is caused by all of its other links causes a link in C2 which causes all of *its* other links.

Perhaps unsurprisingly by now, P4 is satisfied when both "C1" and "C2" name one and the same *loopy* causal chain—say, the causal loop we previously discussed,

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

ABCDE. E causes A, while being caused by all of the links A–D leading up to it; and A causes all of the links B–E following it. Of course the same reasoning applies to any other name you care to call that same loop: CDEAB cause *its*elf since B, which is caused by CDEA, causes C, which causes DEAB; EABCD causes *its*elf since D, which is caused by EABC, causes E, which causes ABCD; and so on. Since, then, the same intuitive conditions that ordinarily qualify one causal chain as causing another are satisfied when those two chains are numerically identical due to forming a causal loop, causal loops should qualify as self-caused.

Note, finally, that the above argument established that causal loops are self-caused by considering only causal relations that are internal to those loops, while an event can be *indirectly* self-caused only with outside help; so a consequence of the above argument is that causal loops are *directly* self-caused. The Lewisian, again, must resist this conclusion, since anything that caused itself directly would have to causally *depend* on itself; and again, Lewis defines causal dependence in such a way as to allow only mutually distinct events to be so related. But in so doing, his definition fails to countenance the undeniably cause-like relation ABCDE bears to itself — which in all respects appears to be a full-fledged case of causation, albeit a unique one due to its direct reflexivity. Intuitively, again, all that's needed for a causal relation to obtain between two causal chains is for the last link in one to cause the first link in the other; and though that description needs a bit of massaging when the two chains in question are numerically identical due to a single chain forming a causal loop, the relevant relation arguably holds. No matter what you name causal loop ABCDE (whether "BCDEA", "CDEAB", or any other variant), the first-named link is caused by the last. Or if we forego talking about "first" and "last" links: since causal loops are causal chains that are caused by each and every one of their own constituent links, and since whatever every link any particular causal chain causes is also caused by that chain as a whole, causal loops directly cause themselves.

5. Non-loopy Self-causation

Some version of my reasoning seems to be employed by Quentin Smith (1999) with respect to the "non-temporal" causal loop I mentioned him imagining earlier, in which three lone particles' simultaneous states instantaneously cause one another: state A instantaneously causing simultaneous state B, which instantaneously causes simultaneous state C, which instantaneously causes A. Because all proper parts of such a loop both cause and are caused by all others, it seems practically irresistible to call the whole *self*-caused instead of uncaused:

> If the universe at t = 0 is a, b and c, and a, b and c are each caused to begin to exist by something internal to the universe, it follows that the universe is caused to begin to exist, but not by anything external to the universe. The universe is self-caused (Smith 1999: 581)

Granted (as an anonymous reviewer helpfully pointed out), the mere fact that each proper part of some particular thing is "caused to begin to exist by something internal to" that thing cannot by itself qualify the thing as self-caused: consider a

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

universe consisting of a single unbroken causal chain extending infinitely far into the past. Even though each and every event in such a universe is caused "by something internal to the universe" (namely, by each and every event preceding it), it sounds wrong to call such a universe self-caused (or even caused at all, as a whole), since it extends infinitely far into the past. Smith's universe consisting of a non-temporal causal loop, however, is crucially different: at least one event can be found in Smith's universe (pick any event you choose) which causes *each and every* event in that universe (including itself); whereas no such event can be found in the universe extending infinitely far into the past. This difference was taken into account in my phrasing of principle P1 in section 3. So regardless of whether instantaneous causation is possible, Smith's inference illustrates that mutual causal relatedness among all parts of a whole plausibly entails causal relatedness of the whole to itself. In the commonly imagined case of causal loops involving time travel or other backward causation we find causation between temporally separated parts, whereas in Smith's variety the parts are separated only spatially; but arguably if the implication from indirectly self-caused links to direct self-causation of the entire loop holds in latter case, it should also hold in the former.

If you feel resistant to Smith's inference—from the premise that all links in a loop are caused by all others to the conclusion that the loop as a whole is self-caused—let me ask: what more could one demand from a directly self-caused event, than that each and every one of its proper parts both causes and is caused by all the others? Again, by mutually causing one another all such parts thereby cause themselves as well as the whole; and whatever every part causes, the whole arguably does as well. Perhaps you could demand that a directly self-caused event must have *no* proper parts, i.e. be indivisible? That would be an extremely strong demand of course, since it is debatable whether indivisible events are even a thing. But even if they are, we don't demand causation *in general* to hold among only such indivisible relata; so why should we demand differently for *self*-causation?

Perhaps, instead, you think that the only kind of event deserving of the title "directly self-caused" would be an event that caused itself *without* its proper parts causing one another. Call that "non-loopy" self-causation: a direct variety of self-causation that doesn't depend on the existence of causal loops at all. I don't doubt that non-loopy self-causation is what the phrase "self-causation" most naturally brings to mind. When Spinoza called God "*causa sui*", for instance, I doubt he had in mind a necessarily composite being each of whose proper parts both causes and is caused by the others. Nor do I deny that such non-loopy self-causation would be markedly different from the indirect variety exhibited by links in loops, as well as from the direct variety exhibited by entire loops themselves: unlike those latter two, instances of non-loopy self-causation would not be explainable by pointing out the occurrence of any causal loops. And non-loopy self-causation might, I admit, turn out to be, unlike the loopy variety, physically impossible: although GR shows several ways loopy self-causation could occur in accordance with the known laws of physics, nothing similar has ever been shown for any non-loopy instance.

But at the very least, the apparently real physical possibility of causal loops goes a long way toward removing any reasonable resistance one might feel to the idea of non-loopy self-causation. If I'm right that causal loops should count as directly self-

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

caused, then we've found something that differs crucially from both *un*caused events (like radioactive decays) and events caused only by other events (like most ordinary happenings): causal loops aren't uncaused (which distinguishes them from radioactive decays), yet at the same time that causal status doesn't depend on their being caused by anything else, i.e. by anything other than themselves (which distinguishes them from ordinary happenings). Non-loopy self-causes would differ from causal loops only in *how they accomplish* their direct self-causation—namely, in some yet-unexplained way which doesn't require the existence of any causal loops. But we've already admitted the remarkable step: the self-causation, independently of how it might be achieved. The surprising conceptual leap is from the common belief that self-causation is logically impossible to the realization that it isn't, and in fact that it appears to be genuinely possible in accordance with well-established laws of physics. Causal loops already force that leap. My argument that causal loops are directly self-caused puts a cherry on that point; but in contrast to those initial leaps, it seems to me like a relatively minor step to accept non-loopy self-causation, were we ever to find reason to.

Moreover, the question of whether non-loopy self-causation is possible is irrelevant to the question of whether causal loops themselves should count as self-caused, and hence to whether direct self-causation is possible. What we're debating —what the Lewisian response to causal loops is designed to rule out, and what I'll claim can help solve puzzles in philosophy of mind—is the possibility of an event causing itself without external intermediaries: that is the proper definition of "direct self-causation", and that is what I've argued causal loops prove possible. Regardless of whether or not such directness can ever be achieved non-loopily.

6. Future Prospects

You might wonder if much is at stake in this debate. What's the big difference between calling causal loops *self*-caused and calling them *un*caused? In fact there is a huge amount at stake: I'll finish by mentioning just two of the many groundbreaking benefits I see the possibility of direct self-causation having for philosophy of mind in particular.

Perhaps most dramatically, new life can be breathed into the age-old fight to save a genuinely libertarian belief in free will. That fight seeks to vindicate your feeling that when you freely decide to do something like raise your arm, your decision is neither externally causally necessitated—i.e. determined by other events—nor to any degree random. As Epicurus famously expresses this feeling:

> [N]ecessity destroys responsibility and chance is inconstant; whereas our own actions are autonomous. (Epicurus 1925: §133)

However, it has long been taken for granted that no event of any kind—decisions included—can have a causal status *other* than those two: namely, either determined by other events or at least to some degree random. So the Epicurean challenge generates what Mark Balaguer (2014: 24) calls the "random or predetermined" argument against free will, leading countless philosophers to what Richard Double (1997)

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

calls the "no-free-will-either-way theory": *no matter what* we ever discover about the causal structure of our brains, free will as we feel we have it cannot exist.

A response voiced in various forms, from Plato's *Phaedrus* to the present, is that not everything that happens must be either externally determined or random: instead, there is a third causal status events may have. The best-known alternative, tracing to Reid (1788), is that free decisions are caused by the deciding agent without being caused by any *events*; however, even founding agent-causalist Richard Taylor admits that such irreducible "agent-causation" is "strange indeed, if not positively mysterious" (1963: 52). For anyone who wants to avoid that mystery, an attractive alternative with an even longer history—though one rarely mentioned nowadays—is that the third causal status is that of being *self*-caused. Philo of Alexandria speaks of a "spontaneous and self determining mind"; both Aristotle (1984b: Book I) and Aquinas (1912) maintain that "the free is the cause of itself"; Francisco Suárez (1856–78) calls free acts "inherently self-willed," such that "when the will elects [...] it elects to elect," which "does not happen through acts distinct from those having been directed"; and so on. My own proposal is that a free decision to raise one's arm is a spontaneous but directly and intentionally self-caused mental event with the complex conjunctive content *that my arm rises and I make this very decision* (where "intentional causation" is causation by intentional content). Whatever the details, if at least some of the decisions we make can cause themselves then the "random-or-predetermined argument" fails: in which case our core feeling of freedom stands a chance of being non-delusional. However, the decline of theism and the rise of scientific and otherwise naturalistic thinking has seen the idea of self-causation fall out of favor. While Galen Strawson (1994: 15) recognizes that "the heart of the free will debate" revolves around the possibility of a "*causa sui*", he admits failure in his attempts to defuse what he calls his "basic argument" against free will (keep in mind that Strawson takes free will to be necessary for moral responsibility):

> (1) Nothing can be *causa sui* — nothing can be the cause of itself. (2) In order to be truly morally responsible for one's actions one would have to be *causa sui*, at least in certain crucial mental respects. (3) Therefore nothing can be truly morally responsible. (Strawson 1994: 5)

Note that a decision that is only *indirectly* self-caused cannot deliver the needed kind of mental process, since any such event will be determined by other events—namely, by whichever events mediate it's causing itself. For a fascinating illustration of this point, see Boris Kment's (2017) example of Twin-Fred, "the self-creating time traveler" and "human *causa sui*" (though Kment invents the example for a different purpose): by virtue of being intentionally self-caused, Twin Fred's decisions fulfill all the criteria for freedom ordinarily demanded by even the most stringent libertarians; nonetheless, those decisions are still clearly unfree, as a result of being only *indirectly* self-caused and hence clearly determined by other events. Self-causation stands a chance of vindicating libertarianism only if free decisions are *directly* self-caused.

The possibility of direct self-causation also impacts several longstanding puzzles about consciousness. Note for instance that insofar as consciousness carries with it

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

direct and immediate introspective awareness of one's own mental states, it appears to be an *intrinsic* feature of states that have it. On the one hand, our conscious mental states cannot be separated from our awareness of them: a conscious twinge of pain, for instance, cannot occur without its owner being directly and immediately introspectively aware of it. Whereas most events, including even many mental ones (e.g. unconscious beliefs and desires), can occur without anyone being aware of them. And conversely, introspective awareness of pain cannot turn out to be a mere hallucination; whereas most states of awareness (e.g. your current visual experience of this page) could. Acts of introspective awareness which underlie consciousness, in short, seem to *just be* the conscious states we thereby introspect: your pain *just is* your introspective awareness of it, and your introspective awareness of pain just is your pain itself. Accordingly, there is a venerable tradition of viewing consciousness as reflexively self-directed—extending back at least to Aristotle ("knowing, perceiving, believing, and thinking are always of something else, but of themselves on the side." 1984b: Book 12.9) and popularized by Franz Brentano (every conscious act "includes within it a consciousness of itself" 1874: 153–154). Nowadays such views are often categorized as "self-representational" theories of consciousness (e.g. Kriegel and Williford 2006).

However, the Aristotelian/Brentanian theory ("pure" self-representationalism, to use a helpful label by Gennaro 2006) has rarely been taken seriously, even by those who claim to endorse it: "For in point of fact perception is not of itself, but of something else besides the perception that is necessarily prior to the perception" (Aristotle 1984b: Book 4.5). In the words of prominent self-representationalist Uriah Kriegel: "[T]he mental state in question does not actually represent itself. At most, we can say that one part of it represents another part" (Kriegel 2005: 48). A central reason for this nearly universal abandonment of the (pure) Aristotelian/Brentanian theory is that introspection is commonly taken to be an internal form of perception, and perceptual experiences grant us awareness of their objects only by way of being causally impacted by them. You can see the sun, for instance, only if the sun itself is among the causes of your visual experience. Which means that for any mental event to be numerically identical with a direct introspective awareness of it—i.e. to be a reflexive awareness of itself—it must *directly cause* itself. And that, of course, has always been assumed impossible. Aristotle's own reason for ultimately accepting that "perception is not of itself", for instance, is that "that which effects change is prior by nature to that which is changed." Kriegel voices a modern version of the same argument:

> [T]he causal relation is anti-reflexive; therefore, no brain state can bear the causal relation to itself; therefore, no brain state can represent itself. Call this the Argument from Physical Implausibility. If sound, it would refute the Neo-Brentanian account of consciousness. (Kriegel 2003: 483)

To say that the theory of consciousness as reflexively self-directed awareness has therefore been viewed unfavorably would be an understatement:

> There can be no doubt that the theory of Aristotle and Brentano is the purest nonsense imaginable. (Bell 1990: 20–23; quoting Findlay 1933)

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

Which unfortunately leaves us unable to account for the kind of introspective awareness each of us experiences as intrinsic to our own conscious states.

If, however, causal loops are directly self-caused—as I've argued above we should judge them to be—then because causal loops not only are coherently conceivable, but also seem genuinely possible in a way explained by currently accepted laws of physics, direct self-causation should be so regarded as well. This injects fresh hope into our fights to resolve the two above puzzles. On the one hand, an Aristotelian/Brentanian theory of conscious events as reflexively self-aware finally stands a chance of being true, even on a naturalistic view of the mind as wholly consisting in neurophysiological brain processes. And on the other hand, our core feeling of genuinely libertarian free will might likewise be vindicated by the idea that our decisions literally "make themselves". Perhaps we can even begin searching for evidence either for or against such causal structures occurring in our brains, in hopes of either confirming or disconfirming those theories. Of course one might understandably be skeptical of the existence of neurophysiological causal loops; but I see no reason to think them impossible. More importantly, my arguments open the possibility that either of the above two theories could be true even in the absence of such loops—in which case, consciousness and/or freedom would inhere in neurophysiological instances of what I previously called "non-loopy" self-causation.

It goes without saying that these last proposals have been woefully brief: they will require vastly more fleshing-out to command serious attention. But as I already mentioned, they are only two among many applications of the idea of direct self-causation I foresee in philosophy of mind. I cannot describe those further potential applications here; but they are all nonstarters if direct self-causation is rejected *a priori*, as historically it always has been. I offer this article in order to remove that roadblock, in hopes of paving the way for such fruitful future investigations.[4]

ANTHONY E. NEWMAN
anewman@alum.mit.edu

References

Abuter, R., Amorim, A., Bauböck, M., et al. (2020). Detection of the Schwarzschild precession in the orbit of the star S2 near the Galactic centre massive black hole. *Astronomy & Astrophysics* 636: L5.

Aristotle (1984a) *Physics*. In Jonathan Barnes, *The Complete Works: The Rev*. Oxford Translation. Princeton, N.J.: 315–446.

Aristotle (1984b) *Metaphysics*. In Jonathan Barnes, *The Complete Works: The Rev*. Oxford Translation. Princeton, N.J.: 1552–1728.

Aquinas, Thomas (1912) *Summa Theologica*. In B. Rand (ed.), *The classical psychologists: Selections illustrating psychology from Anaxagoras to Wundt*. Houghton, Mifflin and Company: 138–146.

Balaguer, Mark (2014) *Free Will*. MIT Press.

Barrow, John and Dąbrowski, Mariusz (1998) 'Gödel Universes in String Theory'. *Physical Review* D58: 103502.

Bell, David (1990) *Husserl*. London: Routledge.

[4] Thanks to Editor-in-Chief Heather Battaly, an anonymous Associate Editor, and several previous anonymous reviewers for helpful comments.

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

Brentano, Franz (1874) *Psychology From an Empirical Standpoint*. Routledge.
Collett, T. E., Oldham, L. J., Smith, R. J., Auger, M. W., Westfall, K. B., Bacon, D., Nichol, R. C., Masters, K. L., Koyama, K., & van den Bosch, R. (2018) 'A precise extragalactic test of General Relativity'. *Science* 360 (6395): 1342–1346.
Double, Richard (1997) 'Misdirection in the Free Will Problem'. *American Philosophical Quarterly* 34 (3): 357–366.
Dowe, Phil (2000) *Physical Causation*. New York: Cambridge University Press
Dyson, Lisa (2004) 'Chronology Protection in String Theory'. *Journal of High Energy Physics* 24 (3).
Echeverria, Fernando, Gunnar Klinkhammer, and Kip Thorne (1991) 'Billiard balls in wormhole spacetimes with closed timelike curves: Classical theory'. *Physical Review* D44 (4): 1077.
Epicurus (1925) 'Letter to Menoeceus'. In Diogenes Laertius and R. D. Hicks, *Lives of Eminent Philosophers*. Harvard University Press: 187–193.
Findlay, John Niemeyer (1933) *Meinong's theory of objects*. Oxford: H. Milford.
Friend, Toby (2019) 'Can parts cause their wholes?' *Synthese* 196: 5061–5082.
Gennaro, Rocco J. (2006) 'Between pure self-referentialism and the extrinsic HOT theory of consciousness'. In Uriah Kriegel and Kenneth Williford (eds.), *Self-representational Approaches to Consciousness*. MIT Press.
Gödel, Kurt (1949) 'An Example of a New Type of Cosmological Solutions of Einstein's Field Equations of Gravitation'. *Reviews of Modern Physics* 21 (3): 447–450.
Hall, Ned (2000) 'Causation and the Price of Transitivity'. *Journal of Philosophy* 97 (4): 198–222.
Harrison, Harry (1967) *The Technicolor time machine*. NY: Garden City.
Hausman, Daniel (1998) *Causal Asymmetries*. Cambridge: Cambridge University Press.
Hawking, Stephen W. (1992) 'The chronology protection conjecture'. *Physical Review*. D46: 603–611.
Hitchcock, Christopher (2001) 'The Intransitivity of Causation Revealed in Equations and Graphs'. *Journal of Philosophy* 98 (6): 273.
Hume, David (1748) *An Enquiry Concerning Human Understanding*. Oxford: Clarendon Press, 2006.
Kim, Jaegwon (1973) 'Causes and counterfactuals'. *Journal of Philosophy* 70 (17): 570–572.
Kment, Boris (2017) 'Free Will and Ultimate Explanation'. *Philosophical Issues* 27: 114–130.
Krasnikov, Serguei (2001) 'The time travel paradox'. *Physical Review* D65 (6): 064013–064020.
Kriegel, Uriah (2003) 'Consciousness as intransitive self-consciousness: Two views and an argument'. *Canadian Journal of Philosophy* 33 (1): 103–132.
Kriegel, Uriah (2005) 'Naturalizing Subjective Character'. *Philosophy and Phenomenological Research* 71 (1): 23–57.
Kriegel, Uriah and Williford, Kenneth (eds.) (2006) *Self-representational Approaches to Consciousness*. Bradford.
Lewis, David (1976) 'The Paradoxes of Time Travel'. *American Philosophical Quarterly* 13 (2): 145–152.
Lewis, David (1986a) 'Postscripts to "Causation"'. In *Philosophical Papers, Volume* II. New York, Oxford University Press: 172–213.
Lewis, David (1986b) 'Events'. In *Philosophical Papers, Volume* II. New York, Oxford University Press: 241–270.
Loeb, Louis (1974) 'Causal Theories and Causal Overdetermination'. *The Journal of Philosophy* 71: 525–544.
Lossev, Andre and Igor Novikov (1992) 'The Jinn of the Time Machine: Non-Trivial Self-Consistent Solutions'. *Classical and Quantum Gravity* (9): 2309–2321.
McDermott, Michael (1995) 'Redundant Causation'. *British Journal for the Philosophy of Science* XL: 523–544.
Mellor, D. H. (1995) *The Facts of Causation*. London: Routledge.
Meyer, Ulrich (2012) 'Explaining causal loops'. *Analysis* 72 (2): 259–264.
Nietzsche, Friedrich (1886) *Beyond Good and Evil*. Vintage.
Plato (1952) *Plato's Phaedrus*. Cambridge: Cambridge University Press.
Reid, Thomas (1788) *Essays on the Active Powers of Man*. Garland.
Schaffer, Jonathan (2003) 'Overdetermining causes'. *Philosophical Studies* 114 (1–2): 23–45.

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press

Smith, Quentin (1999) 'The reason the universe exists is that it caused itself to exist'. *Philosophy* 74 (4): 579–586.
Strawson, Galen (1994) 'The impossibility of moral responsibility'. *Philosophical Studies* 75 (1–2): 5–24.
Suárez, Francisco (1856–78) *Opera omnia*. Paris: Louis Vivès.
Takamoto, M., Ushijima, I., Ohmae, N., Yahagi, T., Kokado, K., Shinkai, H., & Katori, H. (2020) 'Test of general relativity by a pair of transportable optical lattice clocks'. *Nature Photonics* 14 (6): 411–415.
Taylor, Richard (1963) *Metaphysics*. Englewood Cliffs, N.J., Prentice-Hall.
Tipler, Frank (1974) 'Rotating cylinders and the possibility of global causality violation'. *Physical Review*. D9, Issue 8: 2203–2206.
Tooley, Michael (1990) 'Causation: Reductionism versus realism'. *Philosophy and Phenomenological Research* 50: 215–236.
Touboul, P., Métris, G., Rodrigues, M., Bergé, J., Robert, A., et al. (2022) 'MICROSCOPE Mission: Final Results of the Test of the Equivalence Principle'. *Physical Review Letters* 129 (12): 121102.
Wasserman, Ryan (2018) *Paradoxes of Time Travel*. Oxford: Oxford University Press.
Weaver, Christopher Gregory (2018) *Fundamental Causation: Physics, Metaphysics, and the Deep Structure of the World*. Routledge.

https://doi.org/10.1017/apa.2025.10021 Published online by Cambridge University Press