Hume, Causation and Counterfactuals

Joshua Anderson
Department of History and Philosophy,
Virginia State University
1 Hayden Drive, Box 9070, Petersburg, VA 23806, USA
Email: janderson@vsu.edu

Abstract:

What is offered here is an interpretation of Hume's views on causation. While it might not be literally Hume's view, it is certainly consistent with Hume, and is probably what Hume should say on causation, in light of recent developments in science and logic. As a way in, it is argued that the considerations that Hume brings against rationalist theories of causation can be applied to counterfactual theories of causation. Since, counterfactuals, possible worlds and modality were not ideas that would have been overly familiar to Hume, some supplementation of Hume's arguments will be necessary.

Keywords: causation, counterfactuals, David Hume, David Lewis

In both *A Treatise of Human Nature* and *An Enquiry Concerning Human Understanding* David Hume famously argues against the then current rationalist conceptions of causation. Hume believes that since there is no impression of a power causal influence or necessary connection, rationalist understandings of causation cannot be correct. Instead, Hume argues, what one does have is the constant conjunction of cause *c* and effect *e* and the expectation that *e* will follow *c*.

In this article I will argue that the considerations that Hume brings against rationalist and powers theories of causation can be applied to present-day counterfactual theories of causation. That is not to say that Hume is right. Rather, if his arguments are effective against the rationalist and powers theories, then they are effective against counterfactual theories. Now, counterfactuals, possible worlds and modality—despite the influence of Leibniz—were not ideas that would have been overly familiar to Hume. Thus, some reconstruction and supplementation of Hume's arguments will be necessary.

Actually, this article serves a larger purpose; I aim to offer an understanding of Hume's theory of causation. By having Hume engage the counterfactual theory of causation, a better understanding of Hume's ideas regarding causation, generally, can be had. The understanding of causation offered here, certainly, might not have been Hume's. Instead it is what Hume should say about causation in light of recent developments in science and logic. One final point should be made clear; the counterfactual theory, which I will suggest that Hume must reject, is a characterization of the view, but I feel this is a legitimate approach sense the rationalist conception of causation, which Hume actually criticizes, is a characterization of that view. Moreover, the subtleties of the counterfactual

theory—such as "centering," and more detailed understandings of pre-emption—are certainly granted, but the more "metaphysics" that is built into the system, would seem, at least prima facie, to be objectionable to Hume. Thus, it is not necessary to do complete justice to counterfactual theories.

This article will progress in the following way. First, there will be a brief overview of the argument that Hume presents in *An Enquiry Concerning Human Understanding* against the rationalist/powers notion of causation. Then, there will be a presentation of a generic version of a counterfactual theory of causation. Next, Hume's argument will be restructured in order to show that the counterfactual theory of causation is problematic in a way similar to rationalist and powers based theories of causation, relative to Hume's argument, and potential objections that a counterfactual causation theorist might raise will be addressed. Finally, there will be a discussion of Hume's understanding of causation and how it does relate to counterfactuals, since Hume does at one point "define" causation counterfactually.

1. Hume on causation

Before embarking on a reconstruction of Hume's argument, and its implications for a counterfactual theory of causation, it is important to get clear on what Hume's actual argument against a rationalist/powers notion of causation is. To that end, in this section of the article I will be presenting Hume's argument regarding causation as it was put forward in *An Enquiry Concerning Human Understanding*. Hume's arguments are aimed at particular rationalist conceptions of causation that involve necessary connections, and/or, in Hume's words, "secret powers." It should be noted that Hume does not always clearly distinguish the notion of causation as one based on necessary connections, and one based on causal powers. So, in this article, depending on the context, the terms "rationalist" and "powers" will be used, roughly, interchangeably.

For Hume all objects of human reasoning can be divided into one of two categories, either relations of ideas or matters of fact. The former pertain to analytic and/or *a priori* truths and "are discoverable by the mere operation of thought, without dependence on what is anywhere existent in the universe" (Hume 1975, 25). Matters of fact pertain to synthetic and/or *a posteriori* truths that are actually about the world and one comes to know them based on sense impressions—whether directly or indirectly. What is important, however, is that "[t]he contrary of every matter of fact is still possible; because it can never imply a contradiction, and is conceived by the mind with the same facility and distinctness, as if ever so conformable to reality" (Hume 1975, 25).

So, if one wants to know how one comes to the idea of cause, effect and causation, it must be the case that, for Hume, it is either a relation of ideas or a matter of fact. Hume gives several reasons why the notion of causation cannot be based on a relation of ideas or *a priori* reasoning, but they all boil down to essentially that "[w]hen we reason *a priori*, and consider merely any object or cause, as it appears to the mind, independent of all observation, it never could suggest to us the notion of any

distinct object, such as its effect; much less, show us the inseparable and inviolable connexion between them" (Hume 1975, 31).

Since causation cannot be established *a priori*, if it can be established at all, it must be done *a posteriori*—i.e. there must be some sense impression to establish the relation between cause and effect. Notice, though, that since causation cannot be established by pure *a priori* reasoning a strong rationalist conception of causation has already been excluded. There is hope for something like an Aristotelian/powers theory of causation. However, Hume points out that there is never a sense impression of a power such that would provide the content for the idea of a causal relation. While one does have impressions of "actual motion of bodies ... but as to that wonderful force or power, which would carry on a moving body for ever in a continued change of place, and which bodies never lose but by communicating it to others; of this we cannot form the most distant conception" (Hume 1975, 33).

Hume believes at this point he has successfully argued that rationalist and powers theories of causation fail because one cannot form "an idea of [a] power or necessary connexion" which would underwrite such a theory of causation (Hume 1975, 73). He believes that his point is firmly established by the fact that if one did have an idea of a power or necessary connection then the first time one observed a cause one would be able to anticipate the effect, but the fact is that that is not something that is done. Instead, it is only after repeated instances of "similar" causes followed by "similar" effects that one comes to believe that there is a causal relation between the two.

> All events seem entirely loose and separate. One event follows another; but we never can observe any tie between them. They seem *conjoined*, but never *connected*. And as we can have no idea of any thing which never appeared to our outward sense or inward sentiment, the necessary conclusion *seems* to be that we have no idea of connexion or power at all (Hume 1975, 74).

To be consistent with his empiricist theory of meaning, if it is in fact the case that there is not an idea that corresponds to the causal relation, Hume would have to conclude that causal-talk would be meaningless. However, Hume does not draw such a conclusion; rather he tries to figure out what one could mean when using causal-talk. In essence, Hume puts forward his own "theory of causation".

For Hume, causation amounts to the constant conjunction of two events combined with an expectation that one event will follow the other. So, although "[w]e suppose that there is some connexion between them; some power in the one, by which it infallibly produces the other, and operates with the greatest certainty and strongest necessity" it is really just a custom, habit or feeling (Hume 1975, 75). Further, "[t]his connexion, therefore, which we *feel* in the mind, this customary transition of the imagination from one object to its usual attendant, is the sentiment or impression from which we form the idea of power or necessary connexion" (Hume 1975, 75). In other words, it is the feeling that provides the content for the idea of causation. Thus, to Hume's mind anyway, causation has been vindicated, deflated as it may be; at least it is not meaningless.

The purpose of this section of the article was to introduce Hume's arguments against rationalist/powers understandings of causation. There are four important points that need to be taken away from the discussion thus far. First, for Hume, rationalist conceptions of causations are not adequate by the mere fact that one cannot come to an idea of a necessary connection between cause and effect by *a priori* reasoning. Second, since causation is not a "relation of ideas" it must pertain to "matters of fact" and is therefore, radically contingent. By "radically contingent" is meant that for any matter of fact, there is no contradiction in assuming its opposite, or for any matter of fact its denial is always possible. Third, empirically based powers theories of causation are not adequate since there is no impression of any power that would underwrite the causal relation. Finally, Hume presents a deflationary account of causation where the causal relation is understood as a feeling, or expectation, in the mind of the observer that the occurrence of a particular event (cause) will be followed by particular event (effect), and the expectation comes about because of the constant conjunction of similar causes and effects in the past.

2. Counterfactual theories of causation—a characterization

In this section of the article, I will be presenting a rough and generic version of a counterfactual theory of causation. David Lewis is perhaps the best known proponent of a counterfactual theory of causation, so I will be basing my generic version on his work. It is worth noting that Lewis sees himself following Hume based on the fact that in the *Enquiry* Hume states:

> we may define cause to be *an object followed by another, and where all the objects similar to the first are followed by objects similar to the second*. Or in other words *where, if the first object had not been, the second never had existed* (Hume 1975, 76).

Lewis takes Hume to be presenting two different understandings of causation here—a point I will return to below. Those who focus on the first definition—i.e. the part before "Or in other words"—endorse what Lewis takes to be a regularity based theory of causation. A regularity based theory of causation maintains that "a causal succession is supposed to be a succession that instantiates a regularity" (Lewis 1986, 156).

Lewis, however, believes that such regularity based theories run into all sorts of problems, and "that prospects look dark" for these types of theories (Lewis 1986, 160). That is to say, Lewis does not believe that regularity based theories will be able to adequately explain causation. Instead, Lewis believes that a more promising alternative is to take as a point of departure Hume's second definition—if it is a second definition—and defend a counterfactual understanding of causation. So, "we may define a cause to be an object followed by another, and [...] where, if the first object had not been, the second never had existed" can be understood to be claiming that where the first object, the cause *c*, obtains and the second object, the effect *e*, obtains, *c* is the cause of *e* just in case if *c* had not obtained *e* would not have obtained—formally: (*c* is the cause of *e*) ↔ (~*c* □→ ~*e*) (Hume 1975, 76). However, it should be noted that if *c* and *e* do not obtain, and one still thinks that there is some

sort of causal relation between *c* and *e*, then *c* is the cause of *e* just in case if *c* had obtained *e* would have obtained—formally: (*c* is the cause of *e*) ↔ (*c* □→ *e*). Lewis goes on to fill out his definition of causation, with the following caveats: 1) he will only be discussing event causation. 2) his analysis is only meant to explain particular cases. 3) he is concerned with causation broadly construed, i.e. what it is to be *a* cause. 4) that he is only discussing causation under determinism, and what Lewis means by "determinism" is that the laws of nature generally obtain (Lewis 1986, 161-163).

The first thing that Lewis needs to do is to give the truth conditions for counterfactuals generally, then explain counterfactual dependence, and, finally, explain causal dependence. To explain the truth conditions for counterfactuals it is worth quoting Lewis at length.

> Given any two propositions A and C, we have their *counterfactual* A □→ C: the proposition that if A were true, then C would also be true. The operation □→ is defined by a rule of truth as follows A □→ C is true (at a world *w*) iff either (1) there are no possible A-worlds (in which case A □→ C is *vacuous*), or (2) some A-world where C holds is closer (to *w*) than is any A-world where C does not hold. [... or more simply] A □→ C is nonvacuously true iff C holds at all the closest A-worlds (Lewis 1986, 164).

Counterfactual dependence, for Lewis, is, roughly, the truth conditions for a particular counterfactual generalized over classes of A-like and C-like propositions. Thus, the class of C-like propositions depend counterfactually on the class of A-like propositions where no two propositions in the A-like class are compossible and where "if all the counterfactuals A_1 □→ C_1, A_2 □→ C_2 ... between corresponding propositions in the two [classes] are true [... or more simply:] whether C_1 or C_2 or ... depends (counterfactually) on whether A_1 or A_2 or ..." (Lewis 1986, 165).

Finally, causal dependence is roughly equivalent to counterfactual dependence regarding particular events. So, whether or not a particular effect *e* occurs depends on whether or not a particular cause *c* occurs—remembering how Lewis has defined counterfactual dependence. To be clear, *c* and *e* are not propositions but they can be paired with corresponding propositions; "[t]o any possible event *e*, there corresponds the proposition O(*e*) that holds at all and only those worlds where *e* occurs [...and] O(*e*) is the proposition that *e* occurs" (Lewis 1986, 166). Therefore, causal dependence "consists in the truth of two counterfactuals: O(*c*) □→ O(*e*) and ~O(*c*) □→ ~O(*e*)" (Lewis 1986, 166-167). More formally, Lewis' point can be represented as: (*e* causally depends on *c*) ↔ ((O(*c*) □→ O(*e*)) & (~O(*c*) □→ ~O(*e*))). Notice that if *c* and *e* do occur, then the left side of the conjunct is automatically true, so whether or not *e* causally depends on *c* depends on the truth of the right side of the conjunct, and similarly regarding the case when *c* and *e* do not occur—i.e. the right side of the conjunct is automatically true, et cetera. However, for simplicity, one can disregard the O-predicate and then (*e* causally depends on *c*) ↔ ((*c* □→ *e*) & (~*c* □→ ~*e*)), which is basically Lewis' take on Hume's "definition", noted above, with the important difference that Lewis is talking about causal dependence, and not causation, per se. However, Lewis does think that "[c]ausal dependence among actual events implies causation [... but not] the converse [, and this is important because, for

Lewis c]ausation must always be transitive; causal dependence may not be; so there can be causation without causal dependence" (Lewis 1986, 167).

Bringing everything together, in a general way, a counterfactual theory of causation can be expressed thusly:

(e causally depends on c) ↔ ((c □→ e) & (~c □→ ~e))
((c □→ e) & (~c □→ ~e)) is true ↔ (c □→ e) is true and (~c □→ ~e) is true
c □→ e is true (nonvacuously) ↔ (for any world w) all the closest worlds (to w) where c is true e is also true
~c □→ ~e is true (nonvacuously) ↔ (for any world w) all the closest worlds (to w) where ~c is true ~e is also true

More simply, and generically, though, the counterfactual theory can be stated formally as (c is the cause of e) ↔ (c □→ e)—irrespective of the possible-world truth conditions.

3. A Humean argument against counterfactual theories

In this section of the article I will present several arguments, on Hume's behalf, that will demonstrate the inadequacy of a counterfactual theory of causation. Most importantly, I will show that the same types of arguments that Hume uses against rationalist and powers theories of causation can be applied to counterfactual theories of causation. Then, I will consider and respond to some possible objections the counterfactual theorist could make.

To begin, Hume argues that rationalist and powers theories of causation cannot be true. First, rationalist conceptions of causation are inadequate by the mere fact that one cannot come to an idea of a necessary connection between cause and effect by *a priori* reasoning. Second, since causation is not a "relation of ideas" it must pertain to "matters of fact" and, therefore, for any matter of fact its denial is always possible—thus, there can be no necessary connection. Third, empirically based powers theories of causation are inadequate since there is no impression of any power that would provide content for the idea of a causal relation.

Now the rough and general version of a counterfactual theory of causation is: c causes e, just in case, c □→ e, but notice that because causation is a matter of fact the denial is always possible. Thus, it is possible that c could obtain, and e would fail to obtain—i.e. if c were to happen, e might not happen, or formally: c ◊→ ~e. It is clear that Hume entertained such a possibility when he discusses billiard balls in the *Enquiry*.

> When I see, for instance, a Billiard-ball moving in a straight line towards another; even suppose motion in the second ball should by accident be suggested to me, as the result of their contact or impulse; may I not conceive, that a hundred different events might as well follow from that cause? May not both these balls remain at absolute rest? May not the first ball return in a straight line, or leap off from the second in any line or direction? All these suppositions are consistent and conceivable. Why then

> should we give the preference to one, which is no more consistent or conceivable than the rest? (Hume 1975, 29-30)

Further, by definition—in fact Lewis' definition[1]—$(\Phi \mathrel{\Box\!\!\rightarrow} \Psi) \leftrightarrow \sim(\Phi \mathrel{\Diamond\!\!\rightarrow} \sim\Psi)$, or in the cause and effect language that is being used $(c \mathrel{\Box\!\!\rightarrow} e) \leftrightarrow \sim(c \mathrel{\Diamond\!\!\rightarrow} \sim e)$. Yet, Hume believes that $c \mathrel{\Diamond\!\!\rightarrow} \sim e$, and, in fact, because $c \mathrel{\Diamond\!\!\rightarrow} \sim e$ is true, the rationalist conception of causation is false. So, if $c \mathrel{\Diamond\!\!\rightarrow} \sim e$ is true, then $c \mathrel{\Box\!\!\rightarrow} e$ cannot be true, by a simple application of biconditional modus tollens. Hence, it cannot be the case that *c* causes *e*, just in case, $c \mathrel{\Box\!\!\rightarrow} e$, if one wants to maintain causal talk.

Formally, the above argument goes as follows:

1) *c* causes *e* (assumption, one wants to maintain causal talk)
2) *c* causes $e \leftrightarrow (c \mathrel{\Box\!\!\rightarrow} e)$ (assumption, the counterfactual theory of causation)
3) $(c \mathrel{\Box\!\!\rightarrow} e) \leftrightarrow \sim(c \mathrel{\Diamond\!\!\rightarrow} \sim e)$ (true by definition)
4) $c \mathrel{\Diamond\!\!\rightarrow} \sim e$ (assumption, Hume's argument against the rationalist theory of causation)
5) $\sim(c \mathrel{\Box\!\!\rightarrow} e)$ (3, 4 $\leftrightarrow$MT)
6) ~(*c* causes *e*) (2, 5 $\leftrightarrow$MT)
7) (*c* causes *e*) & ~(*c* causes *e*) (1, 6 &I)

However, (7) is a contradiction, thus one of the assumptions must be false since everything else follows logically. Both Hume and the counterfactual theorist seem committed to the truth of (1), i.e. they both want to keep causal talk. Hume certainly believes that (4) is true, and the counterfactual theorist must believe so as well since if he or she did not, then there would be no real reason to put forward an alternative—to, for example, the rationalist—theory of causation. Therefore, (2) has to be rejected. So, for the same reason that the rationalist theory of causation is inadequate, for Hume, the counterfactual theory of causation must also be inadequate, for Hume. In other words, if one thinks that Hume's arguments against the rationalist theory of causation are good arguments, then, *eo ipso*, one is committed to rejecting the counterfactual theory of causation. There may be a worry that I am conflating issues of physical and logical necessity/possibility here. However, I address that concern below when I discuss possible objections.

Besides the main argument, just put forward, there seem to be two additional problems with the counterfactual understanding of causation. The first problem is suggested by Jaegwon Kim, and that is that a counterfactual theory of causation is too broad. In other words, counterfactual dependence captures a broad range of non-causal dependencies. Although Kim's concerns are not, necessarily, Hume's, they do illustrate an additional inadequacy of the counterfactual theory, and therefore lend credence to the main arguments contention that, from a Humean perspective, the counterfactual theory of causation is not an adequate theory of causation. The second problem is that the counterfactual theorist, and in particular Lewis, is inappropriately "breaking apart" the definition that Hume puts forward in the *Enquiry*. It is important to draw attention to this second

problem since Lewis uses Hume's definition to motivate his counterfactual theory of causation. If Lewis' interpretation of Hume is incorrect, then he is not justified in appealing to Hume's authority, which, in turn, gives one reason to suspect the adequacy of a counterfactual conception of causation, from Hume's perspective, at least *prima facie*.

As to the first problem, Kim gives four different types of cases where a counterfactual dependence, of the kind that Lewis maintains, cover instances that are not causal, at least not causal in the way that one normally thinks about causation. Of the four types, three seem particularly problematic and these are: (i) cases of logical/analytic dependence, (ii) when one event is a constituent aspect of another event, and (iii) when one event determines another, but not causally (Kim 1993, 205-206). Kim gives the following examples for each: (i) "If yesterday had not been Monday, today would not be Tuesday." (ii) "If I had not written 'r' twice in succession, I would not have written 'Larry'." (iii) "If my sister had not given birth at *t*, I would not have become an uncle at *t*" (Kim 1993, 205-206).

The reason that the overly broad application of counterfactual dependence is problematic is that it requires "piling on epicycles" to make the theory work (Lewis 1986, 160). What is meant here is that not only does an adequate counterfactual theory of causation have to explain genuine cases of causation, but it also has to have something above and beyond the causal explanation that can, in a principled way, distinguish genuine cases of causation from mere cases of counterfactual dependence.

As to the second problem, it is important to be reminded of what Hume's "definition" of causation is. Hume states that "we may define cause to be *an object followed by another, and where all the objects similar to the first are followed by objects similar to the second*. Or in other words *where, if the first object had not been, the second never had existed*" (Hume 1975, 76). Lewis believes that Hume is actually putting forward two distinct definitions, one regarding a sort of regularity principle and one, the part following "in other words", as a counterfactual theory. Lewis does not believe that the latter is a mere restatement, or clarification, of the former; he believes that the latter "propose[s] something altogether different" (Lewis 1986, 160). Yet, it is unclear why that would be the case—i.e. why the latter is something altogether different, and not a mere restatement, or clarification, of the former.

The fact of the matter is that Hume says "in other words". Now it seems fair to assume that Hume would mean what is normally meant by "in other words," which is that Hume is articulating the same principle in two different ways and not putting forward something altogether different, as Lewis maintains. It seems reasonable to assume that had Hume been putting forward two different principles he would have said as much, for example

> we may define cause to be *an object followed by another, and where all the objects similar to the first are followed by objects similar to the second*. Or [another possible way we could define cause is that] *where, if the first object had not been, the second never had existed*" (Hume 1975, 76).

However, one does not need to just split hairs with Lewis over Hume's wording to get the very point under consideration here across—i.e. that Hume is putting forward one definition and a restatement thereof, and not two distinct definitions. Lewis maintains that the regularity theory of causation—a theory based on the first half of Hume's definition—is problematic for at least three reasons. First, a regularity theory cannot adequately differentiate between genuine causes and cases of confusion between cause and effect, where it is not that *c* is the cause of *e*, but where it is actually that *e* is the cause of *c*. Second, a regularity theory cannot adequately distinguish cases of genuine cases of causation and epiphenomenon. Finally, a regularity theory cannot adequately handle cases of preemption (Lewis 1986, 160). The problem is that the counterfactual theory of causation struggles with the exact same issues, as Paul Horwich has pointed out: "What Lewis said about regularity analyses is now a fair assessment of the counterfactual approach" (Horwich 1993, 216). Kim draws a similar conclusion when he states that: If we compare the classical regularity theory with Lewis's account [...] it is by no means clear that the latter fairs better than the former" (Kim 1993, 207).

Now, it is by no means decisive that just because a counterfactual theory of causation is left wanting, in the same way that a regularity theory of causation is, leads to the conclusion that Hume is not putting forward two distinct understandings of causation. However, coupling the fact that the two theories are left wanting in the same way with the fact that Hume says "in other words" instead of, for example, "or, alternatively", gives one reason to believe that Hume is not putting forward two distinct understandings.

I now turn to some objections that the counterfactual theorist might have regarding the argument I have presented against the counterfactual theory of causation. First, the counterfactual theorist might contend that the argument fails to appreciate the truth conditions for counterfactuals that Lewis puts forward. Second, that the argument does not take seriously the caveats that Lewis puts forth before he even begins his presentation of a counterfactual theory of causation.

It is not the case that the above argument fails to appreciate the truth conditions for counterfactuals, but the introduction of possible worlds does complicate things. So, even if the truth conditions are granted, it is hard to see how one could make a principled distinction regarding closest worlds. As Alan Hajek has argued "the connection between similarity and the truth-conditions for counterfactuals is far less straightforward than has been widely assumed" (Hajek 2018, 14). His point is that for any ordering for determining similarity of worlds, which are almost always based on intuitions, it is susceptible to counter-examples that lead to unintuitive results. It is for that very reason that the argument put forward here has tried to steer clear of such a potential quagmire.

All that really needs to be granted is the mere possibility that c & $\sim e$ is true for $c \mathrel{\square\!\!\rightarrow} e$ to be false. Further, since the mere possibility that c & $\sim e$ is true is one of Hume's reasons for rejecting the rationalist conception of causation, then the same should hold true regarding the counterfactual theory—that is assuming that one believes that Hume is right on that count. In other words, if one grants that a rationalist conception of causation is false because of the mere possibility that c & $\sim e$,

then, one must accept that there are implications for counterfactual theories of causation based on the mere possibility that *c* & ~*e*. It should be noted that the claim here is a bit more nuanced than Hume's objection to the rationalist conception of causation. There Hume seemed to be talking about the logical possibility that *c* & ~*e*. Here—and this is a reconstruction—it is maintained that it is physically possible that *c* & ~*e*. Thus, even if the counterfactual theorist maintains that the would-counterfactual ($c \Box\!\!\rightarrow e$) is a connection of natural necessity, there is the physical possibility that *c* & ~*e* —i.e. there is a probability greater than zero *c* & ~*e* can naturally or physically obtain—which would make the might-counterfactual ($c \Diamond\!\!\rightarrow \sim e$) true and the argument put forward goes through. This issue, the natural/physical possibility that *c* & ~*e*, will become clearer in responding to the next objection.

It should be noted, Hume's actual argument against the rationalist seems to be making a mistake. Hume seems to moving from an epistemic modal—"for all I know the billiard ball could head in any direction"—to a metaphysical conclusion. What is important here is that, the might-counterfactual is a metaphysical, in some sense, modal—the billiard ball might head off in any direction. Thus, the claims made here, are not only consistent with Hume, but actually improve his argument since they can be applied directly to the rationalist understanding of causation without the modal confusion.

The second objection the counterfactual theorist could put forward is that the argument has not taken into account the fact that Lewis has limited himself to discussing causation in a deterministic world. By determinism Lewis means that "the prevailing laws of nature are such that there do not exist any two possible worlds which are exactly alike up to some time, which differ thereafter, and in which those laws are never violated" (Lewis 1986, 163). Therefore, it is not the case, if the laws of nature hold at all the closest worlds, that it is physically possible that *c* & ~*e*.

The problem is, though, that a Humean understanding of natural laws is, qualitatively, not different from a Humean understanding of causation. Thus, assuming the laws of nature at the outset would appear to be, possibly, question-begging. More importantly, the current best science—quantum mechanics, for example—maintains that the laws of nature as we know them are indeterministic and probabilistic, which again if there is a greater than zero probability that *c* & ~*e* then the might-counterfactual holds. "And it isn't just the canonical quantum mechanical examples—radioactive decay, spin measurements on a particle in a Stern-Gerlach apparatus, and so on—that are indeterministic. The indeterminism reaches medium-sized dry goods (and even oversized wet ones), just less obviously so" (Hajek 2018, 7). Hajek gives a billiard ball example to drive the point home, which is not that different from Hume's for establishing that a rationalist conception of causation is untenable.

> Two billiard balls colliding may approximate a deterministic system, but even they are not immune from quantum mechanical indeterminism. One ball might spontaneously tunnel through the other, to China, or to the North Star—incredibly unlikely, to be sure but possible. Thus I cannot truly say "if the cue ball were to hit the 8 ball, the 8 ball would begin rolling" (Hajek 2018, 7).

Finally, Lewis has argued elsewhere that he believes that the counterfactual theory of causation does function in indeterministic settings (Lewis 1986). Thus, the caveat can be ignored, even by Lewis' own light.

However, it should be noted that even determinism would not eliminate the chanciness that is required to make the might-counterfactual true. Hajek points out that a prime example occurs in statistical mechanics—a deterministic system—with Maxwell's demon, but "[t]he point generalizes to other deterministic systems. For every set of initial conditions in which the cue ball hits the 8 ball and each follows an expected trajectory, there is a nearby initial condition in which the balls behave anomalously" (Hajek 2018, 20). And yet again, these are just the types of reasons that Hume did, or would, give to refute the rationalist conception of causation.

To be clear, I am taking Hume in his most skeptical mood here. Thus, holding the laws of nature fixed runs afoul of Hume's arguments against induction. Second, even granting that the laws of nature hold, because of quantum indeterminism and deterministic chanciness—discussed above—the might-counterfactual still holds, and the argument goes through. The onus would be on the counterfactual theorist to explain why, and in what sense, the might-counterfactual does not hold, in a non-question-begging way.

What Lewis is doing with his similarity-of-worlds truth conditions is to block logically possible outcomes from acting as defeaters for the would-counterfactual. What the claim here is, and in Hume's spirit, is to suggest that the *actual* chanciness of natural laws opens up the modal neighborhood which allows the might-counterfactual to be true, even if highly unlikely. Further, appealing to some sort of nomic dependence that holds between cause and effect, and which would make the would-counterfactual true, is *prima facie*, a response not available to the counterfactual theorist, as understood in this article. Lewis, for example is quite clear that "[i]t is essential to distinguish counterfactual and causal dependence from [...] *nomic dependence*" (Lewis 1986, 167).

Granted, in his less skeptical mood Hume allows for natural necessity, which, for example, plays an important role in his discussion of miracles (Hume 1975, 109-131). But, again, given that "natural necessity," understood as the laws of nature holding, is probabilistic and chancy the might-counterfactual will almost always be true making the would-counterfactual false. Finally, it is not claimed that counterfactual theories of causation are false, or that Hume is right. The point of this article is to demonstrate that the types of arguments that Hume uses against rationalist and powers conceptions of causation can be applied to counterfactual theories. The counterfactual theorist might have responses, just as a rationalist or powers theorist might have responses, but those are issues that extend beyond the scope of this article.

4. Hume on causation—redux

With all the foregoing in mind, the fact of the matter is that causal-talk is used, and that Hume did "define" causation using counterfactual-like language. However, if what has been said to

this point is correct, then it cannot be the case that Hume actually was putting forward a counterfactual theory of causation, as such. In this section of the article I will discuss what Hume is actually doing by suggesting that causation can be "defined" counterfactually.

The first thing to remember is how Hume believes one can form an idea of anything. For Hume all of one's ideas are derived from some impression or sentiment, "and where we cannot find any impression, we may be certain that there is no idea" (Hume 1975, 78). Thus, after Hume has argued that the traditional theories of causation—and by extension a counterfactual theory—have no correlated impressions or sentiments that would make them true, Hume is left with two options. Hume can either claim that all causal language is completely meaningless, or he can try and discover an impression or sentiment from which causal-talk could be derived. Hume opts for the latter and puts forward his definition of causation.

Hume actually words his definition in various ways, the two that have already been discussed, but he also says that one can define cause as "an *object followed by another, and whose appearance always conveys the thought to the other*" (Hume 1975, 77). It is this final definition that seems to provide the key to understanding what Hume means by asserting, what Lewis sees as, the counterfactual and regularity definitions of causation. So, in a causal situation there is an event followed by another event, and the reason that one claims that the first causes the second is that one has a feeling, or expectation; upon witnessing the first event one anticipates the second. Further, it is the "feeling" or "expectation" or "anticipation" in the mind that is the impression or sentiment that gives meaning to the idea of a cause.

Hume then explains how it is that one comes to have the feeling/expectation.

> In all single instances of the operation of bodies or minds, there is nothing that produces any impression, nor consequently can suggest any idea, or power or necessary connexion. But when many uniform instances appear, and the same object is always followed by the same event; we then begin to entertain the notion of cause and connexion. We then *feel* a new sentiment or impression, to wit, a customary connexion in the thought or imagination between one object and its usual attendant; and this sentiment is the original of that idea we seek for [i.e. causation] (Hume 1975, 78).

Hume's point is that after many uniform experiences of constant conjunction between a cause and an effect, one begins to expect that there is some connection between the two, and it is that feeling of expectation, based on the many uniform experiences of constant conjunction, that provides the content for the idea of causation.

It is the feeling of expectation that helps make sense of the counterfactual definition of causation that Hume puts forward, and why it is not distinct from the regularity definition. So, "*an object followed by another, and where all the objects similar to the first are followed by objects similar to the second*," is how one comes to have the feeling of expectation, and "*where, if the first object had not been, the second never had existed*," helps explain what the feeling of expectation is (Hume 1975, 76).

Now, it seems that a counterfactual theory of causation would seem to make some sense, for Hume. To be clear, it is not the case that (*c* causes *e*) ↔ (*c* □→ *e*), but something more like: "When I say that *c* causes *e*, I basically mean that when I observe that *c*, I *expect* that *e would* follow." Further, the possible worlds semantics can also be explained in a similar way. Instead of saying that in all the nearby possible worlds where *c* obtains, *e* also obtains, it would be something more like: "Based on my experience, if the world continues as I expect it to, then whenever *c* obtains, *e* also does."

So, there is an additional caveat to the original thesis of this article. If a counterfactual theorist does not have anything stronger in mind than this psychological/linguistic explanation of causation mentioned above, then it does not seem that Hume would find it problematic. If they do, then it would seem that Hume would reject those stronger types of counterfactual theories of causation, and he would do so for reasons quite similar to the reasons he has for rejecting rationalist and powers based theories of causation.

5. Conclusion

This article has been an exploration of Hume and counterfactual theories of causation. It was shown that the types of arguments that Hume gives to reject rationalist and powers based theories of causation can be extended to counterfactual theories of causation. The article does not claim that Hume's understanding of causation is correct, or that counterfactual theories of causation are false. The purpose was just to demonstrate that Humean-type arguments are available to critique counterfactual theories of causation. In other words, if one believes that Hume's arguments are *in fact* successful against rationalist and powers based theories of causation, it would seem, then one is committed to a rejection of counterfactual theories of causation, *prima facie*. It may be that there are good reasons for holding the counterfactual theory, or for that matter a rationalist or powers based theory, of causation, but if one does then one needs to be able to answer the Humean type of arguments presented here. Finally, by engaging a more contemporary understanding of causation, a more complete understanding of Hume on causation can be had. While what was offered here might not be literally Hume's view, it is certainly consistent with Hume, and is probably what Hume should say on causation, in light of recent developments in science and logic.

Endnotes:

1. See for example, Lewis (1973).

References:

Hajek, Alan. "Most Ordinary Counterfactuals are False." accessed September 12, 2018, https://docplayer.net/32738-Most-counterfactuals-are-false-alan-hajek.html.

Horwich, Paul. "Lewis's Programme." *Causation*. Ed. Ernest Sosa et al. Oxford: Oxford University Press, 1993. 208-216.

Hume, David. *David Hume Enquires Concerning Human Understanding and Concerning the Principles of Morals*. Ed. L.A. Selby-Bigge. Oxford: Clarendon Press, 1975.

Kim, Jaegwon. "Causes and Counterfactuals." *Journal of Philosophy* 70, no. 4 (1993): 570-572.

Lewis, David. *Counterfactuals*. Cambridge, MA: Harvard University Press, 1973.

Lewis, David. *Philosophical Papers: Volume II*. Oxford: Oxford University Press, 1986.