Ratio (new series) XXVIII 2 June 2015 0034-0006

INTRINSIC CAUSATION IN HUMEAN SUPERVENIENCE

Daniel Kodaj

Abstract

The paper investigates whether causation is extrinsic in Humean Supervenience (HS) in the sense that *being caused by* is an intrinsic relation between token causes and effects. The underlying goal is to test whether causality is extrinsic for Humeans and intrinsic for anti-Humeans in this sense. I argue that causation is typically extrinsic in HS, but it is intrinsic to event pairs that collectively exhaust almost the whole of history.[1]

Key words: causation, counterfactual dependence, extrinsicality, Humean Supervenience, intrinsicality, relations, totality events.

Arguably, the debate between Humean and anti-Humean views of causation concerns a local link between token causes and effects. Anti-Humeans believe in a causal connection that is, in some sense, wholly present in the locality of token cause/effect pairs (e.g. because it depends on second-order universals, transfer of energy etc.), while Humeans believe that causal relations obtain in virtue of widespread regularities, so that we have to go beyond the immediate locality of two event tokens to see if one caused the other. For the anti-Humean, the fact that *C* caused *E* depends only on how *C* and *E* stand to each other, whereas for the Humean, it depends on how the world at large is.[2]

One might try to characterize this difference by saying that the causal relation is *intrinsic* to token cause/effect pairs for

[1] I would like to thank the reviewer who helped me fix the main argument and Ferenc Huoranszki, who drew my attention to this topic.

[2] See Menzies (1999: 314–7) for more on this distinction. Contemporary anti-Humeans include Armstrong (1992: ch. 14), Martin (1993), and Tooley (1990); the paradigmatic contemporary Humean is Lewis. Note that Tooley (2003: 390–1) draws the distinction in a different way: in the present terminology, he claims that for Humeans, *being a cause* is an extrinsic property of token causes (likewise for effects). As a result, some theories that I would call anti-Humean are classified as Humean by Tooley (e.g. causation as transmission of energy, Fair (1979), or causation as conservation of quantity, Salmon (1997), cf. Tooley 2003: 418). To resolve this issue, one can distinguish two senses of 'Humean.' In Humeanism-1, *being causally related* is an extrinsic property of cause/effect pairs, and in Humeanism-2, *being a cause* is an extrinsic property of causes (likewise for effects). This paper concerns Humeanism-1. Cf. note 7 on p. 4.

anti-Humeans and *extrinsic* for Humeans. Although intrinsicality is typically associated with properties of particulars, we can extend the notion to tuples:

> Exactly as some properties are just a matter of how the thing itself is, without regard to any relationship to any second thing, so some relations are just a matter of how things stand *vis-à-vis* one another, without regard to any relationship to any third thing. The relation is intrinsic to the pair of *relata*. (Lewis 1999a: 193)

This goal of this paper is to investigate whether the relation *being caused by* is extrinsic to token cause/effect pairs in this sense in David Lewis's Humean Supervenience (HS). If it is, then the hypothesis that extrinsic causation is the mark of the Humean is corroborated, and if it is not, then the hypothesis is falsified. Moreover, whether causation is extrinsic in HS seems an interesting question in its own right, even if we refuse to take HS to be representative of Humeanism in general.

Throughout the argument, I assume that HS comes with an ambient metaphysic that includes Lewis's definition of intrinsicality, his theory of causation, and his modal recombination principle. I adopt Lewis's definition of intrinsicality because it meshes very well with his definition of HS,[3] and I adopt his theory of causation and his modal recombination principle because they have impeccable Humean credentials. I also believe that Lewis was selling HS as part of such a package deal, but I will not argue for this claim. In any case, the ideas in question combine very easily and naturally.

Section 1 reconstructs the definition of extrinsic causation in HS. Section 2 argues that causation is typically extrinsic in HS, and Section 3 argues that causation is nonetheless intrinsic to certain (roughly, very 'big') event pairs in HS. The intended upshot is that the Humean need not abhor intrinsic causation, she just needs to locate it at the level of very big events. My argument, if sound, supports the verdict that for Humeans, causation is intrinsic only to pairs of very big events, while for anti-Humeans, causation can be intrinsic to pairs of relatively small and mid-sized events as well.

[3] See note 8 on p. 4.

1. Causation and intrinsicality in HS

Throughout the paper, I assume that causation is a relation between event tokens,[4] and I take events to be regions of spacetime. (Plus their content, if supersubstantivalism is false.)[5]

Whether causation is intrinsic depends on the intrinsic properties of token cause/effect pairs. If causality is intrinsic, then causation is a 'local affair' that concerns the properties of the cause, the properties of the effect, the relations between cause and effect, and nothing else. If causation is extrinsic, then whether *C* causes *E* depends on a bigger chunk of reality, one that involves something more than *C* and *E* and their relations to each other. My first goal is to regiment this idea using Lewis's account of intrinsicality.

Lewis's official account of intrinsicality is built on the notion of *duplication* (Lewis 1986: 61, Lewis and Langton 1999: 120–21): *F* is an intrinsic property of *x* iff every duplicate of *x* is *F*. As for duplication, *y* is a duplicate of *x* iff *y* and *x* have the same natural properties, while natural properties are those 'sparse,' non-gerrymandered properties that form a complete supervenience base for all truths in a world (Lewis 1986: 60). Intrinsicality, duplication and naturalness form a tight conceptual family for Lewis, and at least one of them must be taken as a primitive.[6] I bracket this problem for the purposes of the present paper, because I am not trying to challenge or amend Lewis's metaphysic; my goal is to see how it handles the intrinsicality of causation.

Suppose that *A* and *B* are particulars, and let '<*A*, *B*>' denote their ordered pair. Then <*A**, *B**> is a *duplicate* of <*A*, *B*> iff *A** is a duplicate of *A*, *B** is a duplicate of *B*, and any two-place natural relation is instantiated either by both pairs or by neither. We can then define relation *R* as intrinsic to <*A*, *B*> iff all duplicates of

[4] Supposing that event types T_C and T_E are intrinsically causally related iff the causal relation is intrinsic to all causally related pairs of T_C-tokens and T_E-tokens, the argument of the paper is easily generalized to type causation.

[5] This conforms to Lewis's definition of events as classes of worldbound regions (1986b: 245) with the extra proviso that we are only considering singletons, so that events can be identified with worldbound regions. The argument can be extended to more numerous classes of regions, and hence to Lewis-events proper, by taking two (non-singleton) events *C* and *E* to be causally related iff for each world *W*, the *W*-member of *C* (if there is one) causes the *W*-member of *E*.

[6] See Lewis and Langton (1999: 120–21) and Lewis (1999b: 112) on interdefinability, Loewer (2004: 185–7) on the objectivity of naturalness, and Sider (1993) for a defense of primitivism.

<*A*, *B*> instantiate *R* (Lewis and Langton 1999: 129).[7] Intrinsic relations survive duplication of the pair of relata. Finally, if *C* and *E* are events, let us say that <*C*, *E*> is *causal* iff *C* causes *E*. We then have the following basic definition:

Intrinsic Causation:
Causation is intrinsic to event pair <*C*, *E*> $=_{df}$
All duplicates of <*C*, *E*> are causal.

Example: Suppose that Carol saved drowning Ed by diving into the vortex. If causation is intrinsic to this event pair, then every duplicate of Carol saves the life of every duplicate of Ed, more precisely, every duplicate of the event pair <Carol's diving, Ed's surviving> is such that (duplicate) Carol's diving causes (duplicate) Ed's survival.

Notice that causation is trivially intrinsic if *being caused by* is a natural relation. Intrinsic relations are those that survive the duplication of the pair of relata, and two pairs are duplicates iff they involve the same natural properties and relations. If the causal relation is natural, then duplicates of causal pairs are by definition causal, hence causation is trivially intrinsic.

To make *Intrinsic Causation* more intriguing, we have to reach for Humean notions. I will now give a brief outline of Lewis's Humean Supervenience (HS), his counterfactual theory of causation, and his counterfactual semantics, then I will reformulate *Intrinsic Causation* in the resulting context. The general idea behind HS is that our world is a mosaic of point-like particulars:

> Humean Supervenience is yet another speculative addition to the thesis that truth supervenes on being. It says that in a world

[7] Lewis (1986: 62) distinguishes internal and external relations: internal relations supervene on natural (monadic) properties of the relata (e.g. *having the same shape*), while external relations supervene on natural *relations* of the relata (e.g. *x* orbits *y*). As Lewis (1999c: 26n16) puts it, internal relations are 'intrinsic to their relata,' external ones are 'intrinsic to pairs.' Lewis and Langton (1999: 129) complicate this terminology by lumping external and internal relations together and calling them *intrinsic* relations, understood as intrinsic properties of tuples or fusions. I use the latter convention, and I define extrinsic relations as relations that are not intrinsic. This convention gives us a name for Lewis's 'surd' category, the category of those relations that are not even external (1986: 62). In the terminology I adopt, these are precisely the extrinsic ones. We can then distinguish the two senses of 'Humeanism' by distinguishing between conceptions where causality is extrinsic to token cause/effect pairs (Humeanism-1) and conceptions where it is not internal (i.e. is either extrinsic or external) to them (Humeanism-2). The present paper concerns Humeanism-1 (cf. note 2).

> like ours, the fundamental relations are exactly the spatiotemporal relations: distance relations, both spacelike and timelike, and perhaps also occupancy relations between point-sized things and spacetime points. And it says that in a world like ours, the fundamental properties are local qualities: perfectly natural intrinsic properties of points, or of point-sized occupants of points. Therefore it says that all else supervenes on the spatiotemporal arrangement of local qualities throughout all of history, past and present and future. (Lewis 1999d: 225–6.)

For present purposes, this pitch can be reduced to the following definitions:

Hume/Lewis worlds:
World *W* is a Hume/Lewis world $=_{df}$

(1) All particulars in *W* are (occupants of) spacetime points (or sums thereof).
(2) All facts in *W* supervene on spatiotemporal relations between, or monadic natural properties instantiated by, particulars in *W*.[8]

Humean Supervenience:
The actual world is a Hume/Lewis world.

The first pair of brackets in (1) indicate that HS has at least two versions, a supersubstantivalist one and a non-supersubstantivalist one, the latter of which is committed to point particles (or something near enough). The quote indicates that Lewis was noncommittal about the choice between these options. To make life easier, I will use 'point' as a neutral term for spacetime points and point particles, without assuming anything pro or contra supersubstantivalism.

We are defining events as spacetime regions (plus their content, if this distinction is relevant), so (2) tells us that the natural properties of an event in HS are the monadic properties and the spacetime relations instantiated by the points that

[8] Because of the invocation of naturalness in its definition, HS combines very easily and economically with Lewis's account of intrinsicality. Once you have Lewis's natural properties, you get his intrinsic properties for free.

constitute the event in question. Consequently, HS-events E and E^* are duplicates iff E contains the same number of points as E^*, the points of E are arranged in the same spatiotemporal pattern as the points in E^*, and the points in E instantiate the same monadic properties as the points in E^*.

All we need to complete the puzzle is a definition of causation. Causation is a form of counterfactual dependence for Lewis. Roughly, C causes E iff E would not have occurred if C had not occurred. More precisely, causation is the ancestral of this relation: C causes E iff there is a chain of such counterfactual dependences running from C to E (Lewis 1986: 23 and 1986c: 164–5). For simplicity, I will only investigate chains with two links. It is straightforward to extend my argument to longer chains. For our purposes, then, the canonical definition of causation is the following:

> *Causation*:
> C causes $E =_{df}$ C and E are nonoverlapping events and if C had not occurred, then E would not have occurred.

There are a number of well-known problems about *Causation*. The most important is that it does not handle preemption cases very well. Suppose that Carol dived into the vortex and saved Ed, but Ed would have been saved by his guardian angel if Carol had not intervened. *Causation* then delivers the verdict that Carol's diving did not cause Ed's survival.[9] Because of puzzles like this, Lewis eventually switched to the principle of counterfactual *influence*, which says, roughly, that C causes E iff small variations in C are counterfactually correlated with small variations in E (Lewis 2000: 190f). (Again, we must take the ancestral to get the official definition.) In the present context, the differences between the earlier and the later theories are not terribly important,[10] so I will opt for the original, more straightforward formulation.

Causation can be refined by importing Lewis's theory of counterfactuals. In a nutshell, his theory says that '$A \mathrel{\square\!\!\rightarrow} B$' ('If A had been the case, then B would have been the case') is true at world W iff worlds where A & B is true are more similar to ('closer to') W than worlds where A & $\sim B$ is true.[11] For example, it is true

[9] See Schaffer (2000) for more devious preemption challenges to Lewis's theory.

[10] Note 15 on p. 11 indicates how my argument can be modified for counterfactual influence.

[11] More precisely, '$A \mathrel{\square\!\!\rightarrow} B$' is true at W iff there is a class Γ of worlds where $A\&B$ is true such that no world where $A\&\sim B$ is true is closer to W, the world of evaluation, than any

here and now that this match would have lit if it had been struck iff worlds where this match is struck and lights are closer to our world than worlds where it is struck but and does not light.

Closeness depends on the similarity of the global distribution of (natural) properties. When looking for worlds that make a counterfactual true at *W*, we first look for worlds that are overall very similar to *W*, then, in this restricted group, we look for worlds that differ from *W* as little as possible except for the truth of the counterfactual's antecedent.[12] For example, when we evaluate the counterfactual about the striking of the match, we first look for worlds where matches are typically lit when struck (provided there is oxygen, it is not raining etc.), just like in actuality, then, in this restricted group, we look for worlds that are as similar to ours as the presence of the striking allows.

We now have all the pieces of the puzzle. In Lewis's system, causation is intrinsic iff duplicates of a causal pair are causal, and a pair of events in *W* is causal iff worlds where both events are absent are closer to *W* than worlds where the first is absent but the second is present:

> *Intrinsic Lewis-Causation*:
> Causation is intrinsic to $<C, E> =_{df}$
> For any $<C, E>$-duplicate $<C^*, E^*>$, located in some world *Z*, worlds where both C^* and E^* are absent are closer to *Z* than worlds where C^* is absent but E^* is present.[13]

If causation is intrinsic to a pair of events in this sense, then wherever we find a duplicate of the pair in modal space, the second event is not present without the first one in nearby worlds. If causation is extrinsic to an event pair, then a duplicate of the pair in some world *W* is such that the second member of the pair occurs without the first in a world near *W*.

world in Γ (Lewis 1973: 16, 50, 1986f: 10). I will express this in a shorthand by saying that worlds where *A*&*B* is true are closer to the world of evaluation than worlds where *A*&~*B* is true. Note that in Lewis's official semantics, '$A \mathrel{\Box\!\!\rightarrow} B$' is true at *W* if *A* is necessarily false (cf. Lewis 1973, clause (1) on p. 16 and illustration (A) on p. 17). I will disregard this possibility, because causal counterfactuals cannot be true in this fashion.

[12] Some of the devils in this passage will be confronted in Section 2 (p. 9). For the classic primer on interworld similarity, see Lewis (1972: 91–5); for the fine print, see Lewis (1986d: 43–8).

[13] Since we are identifying events with worldbound regions, the definiens is not meant to invoke transworld identity, only similarity relations based on natural properties.

Intuitively, one would think that the intrinsicality of causation must be an all-or-none affair in the sense that causation is intrinsic to all causal pairs or to none. As we will see, this is not so, and in any case, the definition does not entail anything like that. It formally allows for the possibility that causation is intrinsic to some pairs of events but extrinsic to others. Let me distinguish these two cases by distinguishing between partial and full intrinsicality:

Causation is Partly Intrinsic:
For some causal pair <*C*, *E*>, causation is intrinsic to <*C*, *E*>.

Causation is Fully Intrinsic:
For all causal pairs <*C*, *E*>, causation is intrinsic to <*C*, *E*>.

We can then define partly extrinsic and fully extrinsic causation as the subcontrary and contrary of these cases, respectively: causation is partly extrinsic iff it is extrinsic to some causal pair, and causation is fully extrinsic iff it is extrinsic to all causal pairs. Partly intrinsic and partly extrinsic causation are logically compatible, fully intrinsic and fully extrinsic causation are not. If causation is partly but not fully extrinsic (or partly but not fully intrinsic), then it is both partly intrinsic and partly extrinsic.

2. Causation is partly extrinsic in HS

I will now argue that causation is partly but not fully extrinsic in HS. Roughly, the claim will be that causation is extrinsic to pairs of small and mid-sized events but intrinsic to pairs of 'very big' events in HS. To show that causation is partly extrinsic in HS, I invoke one of Lewis's most important modal principles, the principle of free recombination:

> I suggest that we look to the Humean denial of necessary connections between distinct existences. To express the plenitude of possible worlds, I require a *principle of recombination* according to which patching together parts of different possible worlds yields another possible world. Roughly speaking, the principle is that anything can coexist with anything else, at least provided they occupy distinct spatiotemporal positions. Likewise, anything can fail to coexist with anything else. (Lewis 1986: 87–8)

With this presuppositon in place, assume that Carol dived into the vortex when Ed was sucked into it, powerless to break free on

his own (=event *C*), and, as a result, Ed made it back to the shore alive (=event *E*). Presumably, this pair of events is causal in the our world. If HS is true and the actual world is a Hume/Lewis world, then it follows that <*C*, *E*> is composed of points. By the principle of recombination, these points are modally separable from each other and from the rest of the world. Let '*D*' denote an event during which (a duplicate of) Carol fails to dive and Ed is drowning as usual, e.g. the event of Carol's lying on the beach while Ed is thrashing helplessly in the vortex. By the principle of recombination, there is a world *X* containing a duplicate of <*C*, *E*> plus a two-way infinite recurrence of duplicates of *D* followed by duplicates of *E*, in the following pattern:

World X:

$t =$	−4	−3	−2	−1	0	1	2	3	4	5	
...	*D*	*E*	*D*	*E*	*C*	*E*	*D*	*E*	*D*	*E*	...
	...		Carol is lying on the beach, Ed is saved		Carol dives, Ed is saved		Carol is lying on the beach, Ed is saved		...		

The recombination principle also guarantees the existence of this world:

World Y:

$t =$	−4	−3	−2	−1	0	1	2	3	4	5	
...	*D*	*E*	*D*	*E*	*D*	*E*	*D*	*E*	*D*	*E*	.....
	...		Carol is lying on the beach, Ed is saved		Carol is lying on the beach, Ed is saved		Carol is lying on the beach, Ed is saved		...		

Now consider whether <*C*, *E*> is causal in *X*. (More precisely, the question is whether the relevant duplicate of <*C*, *E*> is causal, but I will omit this qualification to reduce clutter.)

To see whether <*C*, *E*> is causal in *X*, we look for worlds that are overall very similar to *X* but do not contain *C* at $t = 0$. If, among these worlds, worlds where *E* is nonetheless present at $t = 1$ are not farther from *X* than worlds where *C* and *E* are both absent, then <*C*, *E*> is not causal in *X*. Specifically, if the closest relevant world is *Y*, then <*C*, *E*> is not causal in *X*

The next question is, of course, what makes two worlds close. This question generates one of the toughest problems for Lewis's account of causation. For consider @, the actual world, where, by hypothesis, Carol's diving is the cause of Ed's survival. And take a world @* where Carol lies on the beach while Ed is drowning, yet, after a small jerk, Ed is saved and everything proceeds as in @.

One might claim that, among the worlds with an inactive Carol, @* is the one that is closest to actuality, since it matches actual history almost perfectly. But then '$\sim C \,\Box\!\!\rightarrow \sim E$' is false at @. Hence, *C* does not cause *E* in the actual world, and since the argument generalizes, nothing causes anything ever.

This objection, originally raised by Bennett (1974) and Fine (1975) against Lewis-style semantics for counterfactuals, can be met by stipulating that we evaluate closeness to *W* on the basis of the number of events that violate *W*'s laws, with laws understood as systematic global regularities.[14] In a world where Carol is lying on the beach while Ed is drowning in the vortex, powerless to break free on his own, then, a little bit later, Ed makes it to the shore alive, systematic regularities of the actual world are grossly violated. In contrast, if Carol does nothing and Ed dies (as unaided swimmers who are sucked into vortices are wont to do), then the world may fully conform to our laws even if its later history diverges from ours because of the absence of Ed. If violations of law are very important when it comes to closeness, then @* is not the closest world in terms of the truth value of '$\sim C \,\Box\!\!\rightarrow \sim E$' at @.

To deflect such puzzles in a principled way, Lewis set up the following rules about evaluating closeness of worlds:

> (1) It is of the first importance to avoid big, widespread, diverse violations of law.
> (2) It is of the second importance to maximize the spatiotemporal region throughout which perfect match of particular fact prevails.
> (3) It is of the third importance to avoid even small, localized, simple violations of law.
> (4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly. (1986d: 47–8)

Whether these commandments truly save Lewis's theory of causation is, of course, debatable. (For a nice challenge, see Tooley 2003: 410–11.) But the fix is coherent, and it is part of Lewis's system. I will assume, therefore, that (1)–(4) are set in stone as far as closeness is concerned. I will call them 'the Rules.'

[14] Lewis (1972: 74–7, 1986d: 43–48). On Lewis's Humean/Ramseyan account of laws, see Lewis (1972: 74–5) and Loewer (2004).

Now, by the Rules, *Y* seems to be the world that is closest to *X* in terms of the truth value of '$\sim C \,\Box\!\!\rightarrow \sim E$' at *X*. *Y* does not contain *C*, so *Y* makes the antecedent of the conditional true. *Y* is OK by rules (1) and (3), because no laws (that is, no systematic global regularities) of *X* are violated in *Y*. And *Y* is OK by rule (2) too, because it matches the whole history of *X* almost perfectly, except for $t = 0$. Rule (4) does not seem relevant in this case.

Moreover, it seems impossible to assemble a $\sim C$-world which the Rules declare to be closer to *X* than *Y* is. If, instead of switching *C* for *D*, you tinker more seriously with *X*'s history, you'll violate rule (2) without making any improvement in terms of rule (1), so you won't get a world which is closer to *X* than *Y* is. Hence, the Rules say that *Y* is the world that determines the truth value of '$\sim C \,\Box\!\!\rightarrow \sim E$' at *X*. And since *E* occurs at $t = 1$ in *Y*, '$\sim C \,\Box\!\!\rightarrow \sim E$' is false at *X*. So $<C, E>$ is not causal in *X* even though it is causal in the actual world. Hence, causation is partly extrinsic in HS.[15]

My argument crucially depends on the free recombination principle, which guarantees that worlds like *X* and *Y* exist. The recombination principle is not part of the official definition of HS, so one may claim that I'm mixing HS with extraneous components. But this objection only has force on a very narrow interpretation of HS. The principle of recombination has impeccable Humean credentials, and it also happens to be one of Lewis's signature ideas, so if HS is meant to be a theory that is both Humean and Lewisian, then one is justified to incorporate the principle of recombination into it. Likewise for Lewis's theory of causation. At any rate, this paper uses 'Humean Supervenience' to denote that specific combination of ideas.

3. Causation is not fully extrinsic in HS

One might take the previous reasoning as proof that causation is fully extrinsic in HS. The argument would go roughly like this: We can construct worlds like *X* and *Y* for any given causal pair. Those worlds will make the duplicate in *X* noncausal. Hence, causation is extrinsic to all causal pairs in HS.

To see why this is false, take a world *U* that has a 100-million-year history and contains nothing but a ball of uranium and its

[15] An analogous argument gets rid of counterfactual influence. For any counterfactual about small concomitant variations in *C* and *E*, we can construct a world like *X* where the counterfactual in question will be false because of the relevant analogue of *Y*.

accumulating decay products. Suppose that the half-life is 1 million years and it takes exactly 100 million years for the ball to decay completely. The history of *U* is then the fusion of the following two events:

The history of U

A = History begins, the ball of uranium starts to decay, and 50 million years pass without any other activity.

B = Starting 50 million years after the beginning of history in a state identical to the end of *A*, the ball keeps decaying, nothing else happens for another 50 million years, then the world ends.

We may assume that '$\sim A \mathrel{\square\!\!\rightarrow} \sim B$' is true at *U*: If *A* had not occurred and the first half of history had not contained the beginning of the decay process, then the second half of history would not have contained *B*. For *B* begins with a rather complicated arrangement of decay products, and if *A* is absent, then the presence of these products violate systematic global regularities of *U*, hence, by the Rules, we are propelled far away from *U* in modal space. So <*A*, *B*> is a causal pair in *U* – the first half of history causes the second half of history.

Now consider duplicates of <*A*, *B*>. Evidently, you cannot duplicate <*A*, *B*> without producing a world that is indiscernible from *U*. Since *A* and *B* exhaust the whole history of the world and they rule out any further events, you cannot duplicate <*A*, *B*> without duplicating *U* itself. Hence, all causal relations in *U* are present in worlds which contain a duplicate of <*A*, *B*>. And since <*A*, *B*> is causal, it follows that all duplicates of <*A*, *B*> are causal, therefore causation is intrinsic to <*A*, *B*>. And since nothing rules out the hypothesis that <*A*, *B*> is part of a Hume/Lewis-world, causation is not fully extrinsic in HS.

It might be objected that this result is either trivial or uninteresting. It is trivial if there are no indiscernible worlds. In that case, <*A*, *B*> has no duplicate except for itself, so causation is trivially intrinsic to it. Alternatively, if there are indiscernible worlds, then the result is uninteresting, since indiscernible worlds do not seem to do any real theoretical work (apart from making causation intrinsic to <*A*, *B*>, perhaps). Or so one might claim. Moreover, the reasoning has nothing to do with HS itself – it shows that

causation is intrinsic to pairs like <*A*, *B*> in *any* theory of causation where <*A*, *B*> qualifies as causal.

But we can easily modify the example, removing the threat of triviality and bringing in HS itself. Consider the following three events in *U*:

A = History begins, the ball of uranium starts to decay, and 50 million years pass without any other activity.

C = Starting 50 million years after the beginning of history in a state identical to the end of *A*, the ball keeps decaying, and nothing else happens for another 49 million minus 1 years.

D = Starting 99 million years after the beginning of history in a state identical to the end of the 49 millionth year in *B*, the ball keeps decaying, nothing else happens for another 1 million years, then the world ends.

Let '*CD*' denote the fusion of *C* and *D*. *CD* is part of *B*. Specifically, *CD* contains all of *B* except for a 1-year segment which comes between the end of *C* and the beginning of *D*. By a reasoning analogous to the one about <*A*, *B*>, we may assume that <*A*, *CD*> is causal in *U*. Worlds where *CD* occurs in the absence of *A* are farther from *U* than worlds where both are absent.

Given the principle of recombination, there are infinitely many discernible worlds containing duplicates of <*A*, *CD*>. Hence, it isn't trivially true that causation is intrinsic to <*A*, *CD*> in HS, and it isn't even immediately obvious that all duplicates of <*A*, *CD*> are causal in HS.

Consider a world *V* containing a duplicate of <*A*, *CD*>. To make *V* saliently different from *U*, suppose that in *V*, the gap between *C* and *D* is filled with an event that contains nothing but a duplicate of Earth replaying the year 1944 from our history. In *V*, 99 million minus 1 years of radioactive decay (= *A* + *C*) are followed by the Red Army marching into Poland, D-Day getting under way, Laurence Olivier's *Henry V* opening in London etc. (= the year 1944 from actual history), which, in turn, is followed by another 1 million years of radioactive decay (=*D*).

To evaluate '~ *A* $\Box\!\!\rightarrow$ ~ *CD*' at *V*, we must look for worlds that are close to *V* but do not contain *A*. (We could also look at small variations in *A* for an essentially similar argument about counterfactual influence.) To compare two relevant candidates, take (i) a world *W* which contains nothing but the year 1944, and

(ii) a world *Z* which begins with 50 million years of emptiness, followed by *C*, followed by 1944, and ending in *D*:

The history of V (the world of evaluation)
$A > C > 1944 > D$
The history of W
1944
The history of Z
50 mn ys of emptiness $> C > 1944 > D$

Clearly, the history of *Z* contains big, widespread and diverse violations of the laws of *V*.[16] In *V*, the accumulation of radioactive decay products is systematically correlated with changes in the mass of the uranium ball, but in *Z*, a complex configuration of decay products (=*C*) pops out of nothing, without any prior ground in a systematic decay process. By contrast, *W* does not violate systematic global regularities of *V* in this way. In *W*, the year 1944 appears unexpectedly without any prior ground in anything, just like in *V*, and there is no uranium ball before or after, so the violations that infest *Z* are missing in *W*. By Rule (1), *W* is closer to *V* than *Z* is.

This reasoning does not yet prove that '$\sim A \,\Box\!\!\rightarrow \sim CD$' is true at *V*. To show that it is, we must show that replacing *A* with some other event (instead of a stretch of emptiness, as in *Z*) also results in gross violations of laws. So suppose we replace *A* with *R*, which is a non-empty event. If *R* does not feature a ball of uranium, then worlds containing <*R*, *CD*> will again massively violate the laws of *V*, because the beginning of *CD* won't be nomically grounded in *R*. And if *R* features a ball of uranium that is different from the one in *A* (either in size or in terms of its behavior), then *R* and *CD* again won't mesh, so the laws of *V* will again be violated.

It seems safe to conclude, therefore, that removing *A* from the history of *V* in any way while leaving *CD* in place yields worlds with big, widespread, and diverse violations of the laws of *V*. By Rule (1), these worlds are not close to *V*. So <*A*, *CD*> is causal in *V*. And since there is nothing special about *V* as far as this argument is concerned, it follows that causation is intrinsic to <*A*, *CD*>.

A similar reasoning applies to counterfactual influence: worlds where small variations in *A* are counterfactually correlated with small variations in *CD* are closer to *V* than worlds without such a correlation.

[16] See Lewis (1986e: 55–6) for a note on how to count violations: it is the number of violators (= small irregular events) that matters and not the number of violated regularities.

The general shape of this argument is this: We have two contiguous events which together exhaust almost the whole history of a world, adding up to an 'almost-totality' event that leaves only a relatively small spatiotemporal region undescribed. (In the example, the region is the gap between *C* and *D*.) When we duplicate this event pair, we automatically get a world with the same laws, and if we remove the first event from history, leaving the second in place, then we introduce big, widespread and diverse violations of the original laws, which propels us far away in modal space. As a result, the second event will not occur without the first in close worlds.

Three conditions must be met for this reasoning to work. (i) The two events must be of roughly the same size. For if the first is very small compared to the second, then removing it without removing the second may not yield enough nomic violations to propel us far away in modal space. (ii) The two events must be complex enough. They must involve systematic regularities that are violated a great number of times if we chop off the first half of history. (iii) We must admit a special class of events into HS. Let me elaborate on this point.

A is an event the description of which specifies that nothing precedes *A* and nothing else happens while *A* lasts. *B* and *CD* are events the descriptions of which specify that they occur 50 million years after the beginning of history, nothing follows them, and nothing else happens while they last. *B* and *CD* also impose restrictions on the size of the cosmos (they specify that it lasts 100 million years).

I propose to call events like *A*, *B* and *CD* 'subtotality events.' Subtotality events impose various restrictive conditions on the worlds they are parts of. The main restriction is that nothing else is happening while they last. Other restrictions may include conditions about the spatiotemporal location of the event in question (e.g. *B* and *CD* are 50 million years from the start of history), and/or about the extremal regions of the event in question (e.g. nothing comes before *A* and nothing comes after *B* and *CD*).

My argument for intrinsic causation in HS only works for subtotality events, because without the relevant subtotality conditions, recombination easily yields worlds where the original regularities disappear. For example, let A^{O} and CD^{O} be exactly like *A* an *CD* except for the subtotality conditions (i.e. A^{O} is not necessarily at the beginning of history and it may be simultaneous with other complex events, and similarly for CD^{O}). Suppose we recombine

$<A^O, CD^O>$ with a two-way infinite recurrence of $<Q, CD>$, where Q is an event different from A^O. As shown in Section 2, A^O does not cause CD^O in such a world. So causation is not intrinsic to $<A^O, CD^O>$. Removing the subtotality conditions from A and CD destroys the intrinsic causal connection between them.

To sum up: In HS, causation is intrinsic to pairs of relatively complex, evenly balanced, contiguous subtotality events that together exhaust almost the whole history, and causality is extrinsic to all other causal pairs. Specifically, it is extrinsic to pairs of events at least one of which isn't a subtotality event and to pairs of subtotality events that are simple and/or unevenly balanced and/or relatively small compared to the possible size of the worlds they can be parts of. More simply, what we have is that causation is extrinsic to any pair of simple and/or small events in Humean Supervenience.

As far as I know, subtotality events are not part of Lewis's metaphysic. Indeed, I am not aware of any discussion of them anywhere. So my argument can be challenged on the grounds that it imports alien notions into HS.

To meet this objection, let me distinguish HS*, a metaphysic that allows for subtotality events and is otherwise the same as HS, from HS**, which lacks the concept of subtotality events and is otherwise the same as HS. My conclusion is then the following: Causation is partly but not fully extrinsic in HS* and it is fully extrinsic in HS**. The second claim seems justified because the argument from Section 2 works for all causal pairs that lack subtotality characteristics.

Whether Humean Supervenience is identical to HS* or HS** is an interesting question that I'm not taking a stand on. I only note that the combination of HS with the idea of subtotality events seems coherent.

All in all, the following verdict is justified: In HS, causation is definitely extrinsic to small and/or simple token cause/effect pairs. Whether it is extrinsic to all causal pairs depends on one's policy about subtotality events.

4. Summary

I have argued that causation is partly but perhaps not fully extrinsic in HS, provided we take HS to include Lewis's theory of causality, his counterfactual semantics, and his modal recombination

principle. Section 1 assembled these notions into a definition of intrinsic and extrinsic causation. Section 2 argued that causation is partly extrinsic in HS. Section 3 argued that causality is not fully extrinsic in HS, because it is intrinsic to pairs of complex, evenly balanced subtotality events that together exhaust most of history.

If we take HS to be a good representative of Humeanism about causation, then my argument corroborates the following hypothesis: The crucial difference between Humeanism and anti-Humeanism about causation is that for Humeans, *being caused by* is an extrinsic relation between pairs of small and/or simple event tokens, but for anti-Humeans, it isn't.[17]

Central European University
1051 Budapest, Nádor u. 9, Hungary
kodaj_daniel@student.ceu.hu

References

Armstrong, D. (1992). *A World of States of Affairs*. Cambridge: Cambridge University Press.
Bennett, J. (1974). Counterfactuals and possible worlds. *Canadian Journal of Philosophy*, 4: 381–402.
Fair, D. (1979). Causation and the flow of energy. *Erkenntnis*, 14: 219–50.
Fine, K. (1975). Critical notice: *Counterfactuals*. *Mind*, 84: 451–8.
Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.
—— (1986). *On the Plurality of Worlds*. Oxford: Blackwell.
—— (1986a). *Philosophical Papers II*. Oxford: Oxford University Press.
—— (1986b). Events. In: Lewis 1986a, pp. 241–69.
—— (1986c). Causation. In: Lewis 1986a, pp. 159–72.
—— (1986d). Counterfactual dependence and time's arrow. In: Lewis 1986a, pp. 32–52.
—— (1986e). Postscripts to 'Counterfactual dependence and time's arrow'. In: Lewis 1986a, pp. 52–66.
—— (1986f). Counterfactuals and comparative possibility. In: Lewis 1986a, pp. 3–31.
—— (1999). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
—— (1999a). 'Rearrangement of particles: Reply to Lowe'. In: Lewis 1999, pp. 187–96.
—— (1999b). Extrinsic properties. In: Lewis 1999, pp. 111–5.
—— (1999c). New work for a theory of universals. In: Lewis 1999, 8–55.
—— (1999d). Humean Supervenience debugged. In: Lewis 1999, 224–48.
—— (2000). Causation as influence. *Journal of Philosophy*, 97/4: 182–97.
Lewis, D. and R. Langton (1999). Defining 'intrinsic'. In: Lewis 1999, pp. 116–33.
Loewer, B. (2004). Humean Supervenience. In: J. Carroll (ed.): *Readings on Laws of Nature*, pp. 176–206. Pittsburgh: University of Pittsburgh Press.

[17] I have not shown that causation is intrinsic to small/simple causal pairs for anti-Humeans, but it seems clear that it is, at least on Lewis's conception of intrinsicality. E.g. if causation is the conservation of some quantity, then it is, presumably, either a natural relation or a relation that supervenes on natural relations, hence duplicates of causal pairs will be causal. Similarly for transfer of energy, second-order universals, and primitive causal relations. This is only speculation, of course.

Martin, C. B. (1993). Power for realists. In: J. Bacon, K. Capbell, L. Reinhardt (eds): *Ontology, Causality and Mind*, pp. 175–86. Cambridge: Cambridge University Press.
Menzies, P. (1999). Intrinsic versus extrinsic conceptions of causation. In: H. Sankey (ed): *Causation and Laws of Nature*, pp. 313–29. Dordrecht: Kluwer.
Salmon, W. (1997). Causality and explanation. *Philosophy of Science*, 64: 461–77.
Schaffer, J. (2000). Trumping preemption. *Journal of Philosophy*, 97/4: 161–81.
Sider, T. (1993). *Naturalness, Intrinsicality, and Duplication*. PhD thesis, University of Massachusetts.
Tooley, M. (1990). The nature of causation: a singularist account. *Canadian Journal of Philosophy*, (supp. 16): 271–322.
—— (2003). Causation and supervenience. In: M. J. Loux and D. W. Zimmerman (eds.), *The Oxford Handbook of Metaphysics*, pp. 386–434. Oxford: Oxford University Press.