The final and definitive version of this paper is published in the *Thought: A Journal of Philosophy.*

Why Humean Causation is Extrinsic

§1 Humean causes, inside and out

According to the Humean, causal facts[1] supervene on patterns of worldly entities. The simplest form of Humeanism is the constant conjunction theory: some particular type-*F* thing causes some particular type-*G* thing iff (i) that type-*F* is conjoined with[2] that type-*G* thing and (ii) all *F*'s are conjoined with *G*'s. For example, suppose that the bonfire in my backyard causes a plume of smoke. Then (i) that particular bonfire is conjoined with that particular plume of smoke and (ii) all bonfires are conjoined with smoke.

If the constant conjunction theory is true, then the causal facts pertaining to my backyard depend partly upon what goes on outside my backyard. If the world *outside* my backyard contains a bonfire which is not conjoined with smoke, then the bonfire *inside* my backyard does not cause the smoke in my backyard. The same goes for all other causal facts pertaining to my backyard. They, too, depend partly upon what goes on outside my backyard. So, if the constant conjunction theory is true, then all causal facts pertaining to my backyard are *extrinsic* to my backyard. Indeed, something much stronger is true: for all possible regions and all causal facts pertaining to those regions, those causal facts pertain to those regions extrinsically. Call this general thesis extrinsicness. Actual Humeans don't accept the constant conjunction theory; they accept more sophisticated versions of Humeanism. But I argue that they, too, are committed to extrinsicness. Humeans should, on Humean grounds, take causation in general to be extrinsic.

I should pause here to note Hume's views on causation are the subject of controversy.[3] So it's far from obvious that the view I'm attributing to "Humeans" is attributable to Hume himself.[4] But, for better or worse, "Humeanism" has come to be associated with a particular view—namely, the view that (i) there are facts about what causes what and (ii) these causal facts supervene on patterns of worldly entities. (And, moreover, the entities in the patterns are not themselves primitively causal.) This is the view I'm attributing to "Humeans." And I argue that "Humeans," so-defined, must accept extrinsicness.

In arguing that Humeans must accept extrinsicness, I'll be using an exchange between John Hawthorne and Brian Weatherson as a springboard. Hawthorne argues—very briefly—that Humeans are committed to something like extrinsicness (Hawthorne2004, p. 351). Weatherson argues that Hawthorne is mistaken (Weatherson [2007]). I will start by raising a challenge for Weatherson's argument. Then I will show that the challenge generalizes, so Humeans in general must accept extrinsicness.

Before we dive into Weatherson's argument, we should get clear about a few key terms. I've said that causal facts *pertain to regions*, but this way of talking isn't terribly precise. In the discussion that follows, I'll follow Hawthorne and Weatherson in talking of *causal profiles* of regions. A causal profile is a property of regions—it's a property of containing an x and a y such that (for some properties F and G) x has F, y has G, and x causes y. For example, my backyard has a certain causal profile: it has the property of containing an x and a y such that x is fire, y is smoke, and x causes y.5 We should also settle upon clear criteria for extrinsincness and intrinsicness. I will adopt David Lewis' model. A region R has a property P intrinsically iff R has P, and every possible duplicate of R has P. And R has P *extrinsically* iff R has P, and it's not the case that every possible duplicate of R has P (Lewis 1983, pp. 356–357). With these clarified terms at hand, we can define extrinsicness in full exactness.

Extrinsicness: All possible regions have all of their causal profiles extrinsically.[6]

If my arguments are successful, then Humeans must bite any bullets which come along with accepting extrinsicness. For example, it is plausible that mental properties essentially implicate causal facts. Necessarily, if I feel pain, then my pain has some sort of aversive effect. But if mental properties essentially implicate causal facts, and if all causal profiles are extrinsic, then mental properties are themselves extrinsic. But it's implausible that all mental properties are extrinsic. It's implausible that when I feel pain, it's extrinsic to me that I feel pain. So, arguably, extrinsicness leads to implausible results.[7] And if I'm right, Humeans can't avoid those results by rejecting extrinsicness.

§2 Local patterns, local Laws

Here's how Weatherson argues that Humeans need not accept extrinsicness. First, Weatherson introduces a theory of natural law in the spirit of David Lewis. According to this theory, the natural laws of a given world are whichever propositions describe that world in a maximally simple and informative way (Lewis [1973], [1979]). This *best-systems theory* is a perfectly Humean theory: it's a theory on which the natural laws supervene on patterns of worldly entities. And, Weatherson argues, the best-systems theory yields the result that some regions have their causal profiles intrinsically.

Weatherson's key claim is that the best-systems theory is more sensitive to *local* patterns than it is to *global* patterns. So if the best-systems theory is correct, then the natural laws of a given world will reflect local patterns more than global patterns. And as a consequence, some regions will have their causal profiles intrinsically. Weatherson illustrates his point with the following case. There are two worlds *W1* and *W2* which have different sets of propositions as laws: *L1* and *L2*, respectively. And there is a third world which "patches together" *W1* and *W2*.8 One part of this patchwork world is a duplicate of the largest region in *W1*; another part is a duplicate of the largest region in *W2*. Call those regions *r1* and

r2, respectively. Weatherson tells us what a Lewis-style Humean ought to think about the natural laws of this patchwork world:

> If the parts are large and isolated enough, it would be foolish to say that within those parts nothing is law-governed, or that within those regions there is no counterfactual dependence, or no causation. Much better to say that patterns obtaining within such a region are sufficiently simple and informative to count as laws. In our patchwork world, the laws might simply say *in r1, L1 and in r2, L2* (4).

At the *global* scale, the patchwork world seems disunified. There are no simple and accurate propositions which describe all the goings-on of that world. But at the *local* scale, within each of *r1* and *r2*, there are unified patterns. The best way to describe the patchwork world—the maximally simple and accurate way—is to describe the patterns within each of *r1* and *r2*. So, on the best-systems theory of laws, the laws of the patchwork world are descriptions of those patterns.

For each proposition in *L1* and *L2*, a corresponding *region-restricted* proposition is a law of the patchwork world. If *L1* includes the proposition that all bonfires are conjoined with smoke, then a corresponding region-restricted proposition is a law of the patchwork world—namely, the proposition that all bonfires in *r1* are conjoined with smoke—and so on for all other propositions included in *L1* and *L2*. In this way, the laws of the patchwork world describe the local patterns pertaining to *r1* and *r2*. Weatherson suggests that these laws preserve the causal profiles of *r1* and *r2*. So in *r1*, as in *W1*, bonfires cause smoke.

Having established this central idea, Weatherson turns our attention to the focus of Hawthorne's (2004) discussion: the smallest region containing Hawthorne himself. Hawthorne argues that on a Humean theory of causation, the smallest region containing his body does not have its causal profiles

intrinsically. Weatherson argues for the opposite conclusion. He argues that, for the same reasons *r1* has the same causal profiles as its duplicate in *W1*, all duplicates of the smallest region containing John Hawthorne have the same causal profiles as each other. Weatherson has us imagine a duplicate of John Hawthorne which is embedded in a world which hosts very different patterns than our world. He tells us that:

> …in such worlds, laws like *In R, L*, where *R* picks out the region Hawthorne's body occupies, and *L* picks out a real-world law, will be true, simple and informative. […] In other words, even if we embed a Hawthorne duplicate in a world with very different patterns, Humeans will still have good reason to say that the laws, and hence the facts about counterfactual dependence and causation, inside that duplicate are not changed. (5–6).

As it is with *r1* and *r2*, so it is with the smallest region containing John Hawthorne: the laws of worlds containing a Hawthorne-duplicate will reflect the local patterns pertaining to the Hawthorne-duplicate.

The upshot is that on Weatherson's proposal, at least some regions are such that all of their duplicates have the same causal profiles. Consider the actual Dan-region: the smallest region containing me. For any Dan-region, that Dan-region is either embedded in a world which shares our laws, or it is embedded in a world which does not share our laws. If it is embedded in a world which shares our laws, then it has the same causal profiles as the actual Dan-region. (Weatherson tells us "…the laws, and hence the facts about counterfactual dependence and causation, inside that duplicate are not changed" Weatherson [2007], p. 6.) If the Dan-region is embedded in a world which does not share our laws, then that world has region-restricted laws. Those laws have the form: "*c*'s in *R* are conjoined with *e*'s," where "*R*" refers to Dan-regions (or regions containing Dan-regions) and "*c*'s are conjoined with *e*'s" stands in for an

actual law of nature. In virtue of those region-restricted laws, Dan-regions in that world have the same causal profiles as the actual Dan-region. So, whether or not a Dan-region is embedded in a world which shares our laws, it has the same causal profiles as the actual Dan-region. So the actual Dan-region has its causal profiles intrinsically. So extrinsicness is false.

It is helpful to think of Weatherson as making two proposals: a general proposal and a specific proposal. The *specific proposal* is that every world containing a Dan-region either shares our laws of nature, or has region-restricted laws. The *general proposal* is that each world containing a Dan-region has *profile-preserving* laws. That is, they have laws such that Dan-regions in those worlds have the same causal profiles as the actual Dan-region. It is a consequence of the general proposal that Dan-regions have their causal profiles intrinsically.

In the next two sections, I'll show that both proposals fail. I'll start by showing that many Dan-regions have laws wildly unlike the actual laws. Then I'll argue those Dan-regions do not have the same causal profiles as the actual Dan-region. So the actual Dan-region doesn't have its causal profiles intrinsically. These results generalize; the upshot is that Humeans must accept extrinsicness.

§3 Edge cases and global defeaters

To start to see the difficulties facing Weatherson's specific proposal, we can begin by considering "edge cases." In the actual world, some *c*'s at the edges of the Dan-region are conjoined with *e*'s just outside the Dan-region. In many non-actual worlds, those *c*'s are not conjoined with *e*'s. So, in those non-actual worlds, it's simply false that *c*'s in Dan-regions are constantly conjoined with *e*'s. A false proposition cannot be a law, so Weatherson's proposed laws are not laws. So Weatherson's specific proposal is false.

To make this point vivid, consider *Solitude World*: a world containing nothing but a lone Dan-region floating in the void. Various particles are constantly moving into and out of the actual Dan-region. So, at the edges of the Dan-region in Solitude World, particles pop into and out of existence.[9] These phenomena constitute violations of actual laws of nature. So they constitute violations of the region-restricted "laws" proposed by Weatherson.

It might be tempting to patch up Weatherson's specific proposal by weakening his proposed laws. After all, one might think, there are more particles *within* a given Dan-region than there are at the edges of that region. So while it's not a law that *c*'s in Dan-regions are conjoined with *e*'s, it might be a law that *most c*'s in Dan-regions are conjoined with *e*'s.

This strategy might seem tempting when it comes to certain laws of nature, but it clearly will not work for other actual laws. Consider, for example, the laws of conservation of mass and energy. It's a law of nature that mass and energy remain constant in closed systems. But mass and energy are not conserved in Solitude World. Mass and energy are not even *almost* conserved in Solitude World (whatever that may mean). The Dan-region shrinks and grows; it loses and gains energy. Nothing like the conservation of mass and energy is true in Solitude World, and a proposition which suggests otherwise obviously has no claim to lawhood.

The lesson here is that many Dan-regions do not have laws of the kind Weatherson describes. Many Dan-regions are in worlds which do not share our laws, and many of those Dan-regions are not accurately described by Weatherson's proposed laws. So Weatherson's proposed laws are not laws of all Dan-regions,[10] and Weatherson's specific proposal comes out false.

Solitude World is not the most problematic counterexample to Weatherson's specific proposal. Other worlds are significantly more problematic for Weatherson, because they have laws which are

wildly unlike the laws of the actual world. I will consider two such worlds: Malebranchean World and Plentitude World.

Malebranchean World contains a menagerie of duplicates of planets from other worlds. One of the planets in the menagerie is a duplicate of the actual Earth, but the menagerie also contains many billions of other planets, each of which is very different than Earth. The menagerie of planets is arranged in a perfect circle, and at the center of the circle there is an Author. The Author spends her days dreaming up incredibly detailed stories, which she writes down in a notebook. Whenever the Author describes an event in her notebook, the event described occurs in one of the planets in the menagerie. Call this proposition notebook.

> Notebook: Whenever the Author describes an event in her notebook, the event described takes place a second later.

Notebook is an extremely simple and informative description of the goings-on of Malebranchean World. It is far simpler than a description which lists all the most general patterns in each of the billions of planets in the menagerie. So we ought to conclude that notebook figures in the simplest and most informative description of Malebranchean World. And so we ought to conclude that if the best-systems theory is true, then notebook is a law.

It's worth pausing for a moment to think about how notebook relates to Weatherson's distinction between local and global patterns. Unless we take a global perspective on Malebranchean World, it's not obvious that notebook is a law. If we looked at each planet individually, we'd see only the patterns pertaining to each individual planet. We might conclude that the best description of Malebranchean World is the set of descriptions of the planet-specific patterns. If so, we'd conclude that Malebranchean World has planet-specific laws. But we'd be wrong. When we consider the planets from a global scale, it's

clear that the planet-specific patterns are not the proper basis of the laws. They are *defeated* by a much simpler global pattern, described by notebook. So notebook is a law, whereas the planet-specific descriptions are not laws.

This relationship between laws and patterns is exhibited by a second world. *Plenitude World* is like the actual world, in that the two worlds share the same set of fundamental properties and relations. But Plenitude World, unlike the actual world, is plenitudinous. Plenitude World is a world at which plenitude is true:

> Plenitude: For every abstract pattern of fundamental property and relation instantiations, there is some set of objects which realizes that pattern.

It takes a bit of work to make this idea completely precise, but the basic idea is simple enough. For every way in which the actual fundamental properties and relations might be assigned to a set of objects, there is a set of objects in Plenitude world such that those properties and relations are assigned to those objects in that way (see Russell and Hawthorne [2019]).[11] I am composed of a set of objects instantiating a certain pattern of fundamental properties and relations. That pattern is realized infinitely many times in Plenitude World, so Plenitude World contains infinitely many Dan-regions.

Plenitude is an extremely simple and informative description of the goings-on of Plenitude World. It's far simpler than a description which lists the patterns pertaining to various regions within Plenitude World. So, on the best-systems theory, we get a familiar result: the global pattern, described by plenitude, defeats the local patterns. plenitude turns out to be a law.

In principle, a Humean could ignore the global patterns when formulating the laws. And in this way, the Humean could guarantee that global patterns never defeat local patterns, and guarantee that neither notebook nor plenitude are laws. But Humeans shouldn't ignore the global patterns. The

Humean's core idea is that natural laws are perspicuous descriptions of patterns. This isn't merely a commitment of the Lewisian best-systems theory—it's shared by Humean theories of laws more generally. And if we're looking to perspicuously describe the patterns of a world, the last thing we should do is refuse to consider global patterns in that world. So Humeans should allow themselves to consider global patterns. As a consequence, they should allow that global patterns may defeat local patterns, thereby yielding laws like notebook and plenitude.

§4 Strange laws and extrinsic causation

Now let's consider Weatherson's general proposal: all worlds containing Dan-regions have *profile-preserving* laws. As we've seen, some Dan-regions have laws which are wildly unlike the actual laws—notebook and plenitude, for example. It's intuitively implausible that Dan-regions with those laws have the same causal profiles as the actual Dan-region.

We can test this intuition using a specific theory about the relationship between causation and natural lawhood. On David Lewis's theory, we can use counterfactuals to test if notebook and plenitude are profile-preserving. In both Malebranchean World and Plenitude World, there is a duplicate of my backyard, complete with a bonfire *F* and smoke *S*. Here's the counterfactual test: if *F* were absent, would *S* also be absent? In Malebranchean World, there's no obvious reason to think that if *F* were absent, *S* would be absent. The Author would still describe *S* in her notebook. So we ought to think that, one second later, *S* would spring into being. Similarly, in Plenitude World, there's no obvious reason to think that if *F* were absent, *S* would be absent. It's not a violation of plenitude for *S* to appear without *F*. Whether or not *S* is conjoined with *F*, infinitely many duplicates of *F* and *S* are conjoined with one another. So plentitude is preserved either way. Moreover, a world containing *S* is closer to Plenitude World than a world which does not contain *S*—at least in the respect that it contains a counterpart of *S*.

On balance, then, it seems that a world containing *S* but not *F* is closer than a world containing neither *S* nor *F*. So if *F* were absent, *S* would *not* be absent. So in both Malebranchean World and Plenitude World, fire does not cause smoke. And if this simple causal relation does not obtain, there's no hope at all for the proposal that Dan-regions in those worlds have the same causal profiles as the actual Dan-region. Neither notebook nor plenitude is profile-preserving.

You don't have to be a strict Lewisian to think that neither notebook nor plenitude is profile-preserving. Everyone—Lewisian or otherwise—should think that natural laws have something importantly to do with causation. So if two regions differ radically with respect to their natural laws, we ought to think that those regions have different causal profiles. So Humeans in general ought to deny that either notebook or plenitude is profile-preserving.

But suppose a Humean is dead-set on rejecting extrinsicness, and insists upon the result that notebook and plenitude are profile-preserving. Then he faces a dilemma. On the one hand, he might grant that natural laws have something importantly to do with causation. In that case, he must argue that *all* sets of laws pertaining to any Dan-region are profile preserving. To call this an uphill battle would be an understatement. It's already deeply implausible that notebook and plenitude are profile-preserving. No doubt that many other strange and exotic laws pertain to other possible Dan-regions. The claim that *all* such laws are profile-preserving simply beggars belief. So this is an extremely unpromising strategy. On the other hand, suppose our Humean rejects the idea that natural laws have anything importantly to do with causation. So all possible Dan-regions have the same causal profiles despite varying wildly with respect to their natural laws. If the Humean is willing to say this, then he can reject extrinsicness. But he does so at the cost of accepting an absurd theory of natural laws.

The lesson here is that there's no remotely plausible way for a Humean to claim that all Dan-regions have profile-preserving laws, given that some of those regions have laws like notebook and plenitude. The laws are simply too different. So Humeans should reject Weatherson's general proposal; they should accept that Dan-regions have their causal profiles extrinsically. And of course, there's nothing so special about Dan-regions. *Every* possible region has duplicates such that wildly different laws pertain to those duplicates. So Humeans ought to conclude, on Humean grounds, that no regions have their causal profiles intrinsically. In other words, Humeans ought to accept extrinsicness.[12]

Notes

1 Some philosophers take there to be an important difference between *facts* and *truths*: the former carry ontological commitment in a way that the latter do not. In this discussion, I make no distinction between facts and truths. As I use "causal fact," I mean nothing more than *causal truth*.

2 Here and elsewhere, the phrase "conjoined with" stands in for a description of some relation involving temporal succession and spatial contiguity (or near-contiguity). When I say "all *c*'s are conjoined with *e*'s," I mean something like the following: for every *c*, there is an *e* which occurs after *c* and which is spatially contiguous (or near-contiguous) with that *c*. The details do not matter much for my purposes.

3 For a useful overview of the debate, see Beebee (2012).

4 I thank an anonymous reviewer for raising this point.

5 Our notion of a causal profile can be expanded to capture more complex causal relationships. For example, we might allow causal profiles of the form: being a region which contains an *x*, a *y*, and a *z* such that (i) *x* has *F*, (ii) *y* has *G*, (iii) *z* has *H*, and (vi) *x* and *y* cause *z*. For ease of readability I will continue to discuss the simplest case, but the discussion can be straightforwardly expanded to apply to more complex causal relationships.

6 There is an important point worth noting here. extrinsicness does *not* imply that causation is an *extrinsic relation*. There are Humean views according to which causation is an intrinsic relation, *and* extrinsicness is true. In fact, Weatherson considers one such view in his 2007 discussion. He tells us: "If one analysed causation as *that intrinsic relation that actually most tightly correlates with the constant conjunction relation*, then one would have guaranteed that causation was an intrinsic relation. Moreover, one would have a perfectly Humean theory of causation. (A perfectly awful theory, to be sure, but still a Humean one)" (Weatherson 2007, pp. 529–530). On this theory, "causation" involves a nonrigid definite description. In every world, it refers to an intrinsic relation. But it refers to *different* intrinsic relations relative to different worlds. So no pair of objects is intrinsically such those objects stand in the causation relation. So no regions have their causal profiles intrinsically—extrinsicness is true. As Weatherson notes, Peter Menzies has developed a sophisticated theory along these lines. On Menzies' view, as on Weatherson's proposed view, "causation" involves a nonrigid definite description which refers to intrinsic relations (Menzies 1996, 1999). So Menzies' theory also implies *both* that causation is an intrinsic relation, *and* extrinsicness.

7 This is a much-simplified version of an argument from John Hawthorne (2004).

8 Here I assume, with Weatherson, that the two worlds can be patched together. This isn't an entirely trivial matter—according to Lewis, there are constraints on possible patchings (Lewis 1986, pp. 88–90)).

9 I assume that the periphery of a Dan-region is shaped like a sheet of cellophane wrapped around the whole of my body. On this conception of the periphery of the Dan-region, the Dan-region contains both the stuff which composes me and the empty space between that stuff. On a competing conception of the periphery of the Dan-region, the cellophane is wrapped around each particular bit of stuff which composes me. So it does not include any of the empty space. Weatherson's phrasing is somewhat

suggestive of the first conception, but it's worth noting that the problems can't be avoided by embracing the second conception. It would still turn out that some *c*'s contained by Dan-regions are not conjoined with *e*'s, and for the same reasons.

10 It's worth noting that Weatherson's specific proposal faces another potential difficulty. Weatherson seems to assume that if two regions are intrinsic duplicates and share the same laws of nature, then they have the same causal profiles. So he grants that if a Dan-region has the same laws as the actual Dan-region, then it has the same causal profiles as the actual Dan-region. But this isn't obviously correct. Jonathan Schaffer argues that the Lewisian theory allows for cases of *causal trumping*—and if Schaffer is right, then Weatherson's assumption is false. If there is causal trumping, then two regions can be perfect duplicates and share the same laws, but differ in causal profiles nevertheless. There can be two regions each containing *x* and *y*, but in only one of the regions does *x* cause *y*. In the other region, the would-be causal relation is "trumped" by an external factor. If such trumping is possible, such cases constitute further counterexamples to Weatherson's proposal. For a thorough treatment of the trumping cases, see Lewis (2000) and Schaffer (2000). And see Bernstein (2015) for dissent—Bernstein argues that all alleged cases of causal trumping are instead cases of overdetermination or early preemption. My objections to Weatherson do not turn on debates concerning causal trumping.

11 You might worry that Plenitude World is impossible. Indeed, if we adopt broadly Lewisian constraints on combinatorial possibilities, we'll say that it is impossible. But Russell and Hawthorne argue, *pace* Lewis, that worlds like Plenitude World are indeed genuine possibilities (2019, 24–31). The arguments are complex, but there's no need to get into the weeds here. If you're worried that Plenitude World is impossible, we can appeal to a more modestly plenitudinous world to make the same point. We can appeal to a world at which an instance of the following schema is true: *N*-Sized plenitude: For every

abstract pattern of fundamental property and relation instantiations *among N or fewer objects*, there is some set of objects which realizes that pattern. For a sufficiently high *N*, *N*-sized plenitude guarantees that the world contains a Dan-region. So *N*-sized plenitude is all I need for my purposes. Thanks to an anonymous reviewer for raising this point.

12 I would like to thank John Hawthorne, Jeffrey Russell, and Mark Schroeder for their extremely helpful feedback on this paper. Thanks also to Jennifer Foster and Kenneth Silver for many fruitful conversations on the nature of causation.