REGULARITY AND COUNTERFACTUALITY IN HUME'S TREATMENT OF CAUSATION

José Oscar de Almeida Marques*
jmarques@unicamp.br

ABSTRACT *Of the several theories of causation current in our days, Hume is said to be the inspiration of two of the most influential and accepted: the regularity theory, first clearly formulated by Thomas Brown in 1822, and the counterfactual theory, proposed by David Lewis in 1973. After a brief outline of the comparative merits and difficulties of these two views, I proceed to examine whether Hume's own treatment of causation actually corresponds to any of them. I will show that his first definition of cause, coupled with his rules by which to judge about causes and effects, contains elements that, properly developed, allow us to address successfully some traditional difficulties of the regularity view of causation, without resorting to the conceptual resources employed in the counterfactual approach. Therefore, we can properly classify Hume as an advocate of the conception of causation as regularity, noting however that his primary goal in his research and definitions of the concept was to provide not so much an analysis of causation as such, but of causation as we apprehend it, in the form of our ability to make causal inferences and refine them to reach the more sophisticated causal reasonings that are required in the theoretical and practical issues of life.*

Keywords *Causality, Regularity theory of causation, Counterfactuality*

RESUMO *Das diversas teorias da causação existentes em nossos dias, Hume pode ser considerado o precursor de duas das mais influentes e aceitas: a teoria regularista, formulada claramente pela primeira vez por Thomas Brown, em 1822, e a teoria contrafatualista, proposta por David Lewis em 1973. Depois de um breve resumo dos méritos e dificuldades comparativos dessas duas perspectivas, passo a examinar se o tratamento de Hume da causação corresponde, na verdade a algum deles. Mostro que a sua primeira definição de causa, juntamente com suas regras para julgar sobre as causas e efeitos, contém elementos que, devidamente desenvolvidos, permitem-nos abordar com sucesso algumas dificuldades tradicionais da visão de regularista da causação sem recorrer aos recursos conceituais empregados na abordagem contrafatualista. Podemos, portanto, classificar corretamente Hume como um defensor da maneira de conceber a causação como regularidade, observando, porém, que o principal objetivo de sua pesquisa e definições do conceito foi o de fornecer, não tanto uma análise de causação, como tal, mas da causação tal como nós a apreendemos, como nossa capacidade de fazer inferências causais e refiná-las para alcançar os raciocínios causais mais sofisticados que são necessárias nas questões teóricas e práticas da vida.*

Palavras-chave *Causalidade,   Teoria   regularista   da   causação, Contrafatualidade*

In the Introduction to *The Oxford Handbook of Causation* we read:

> The regularity and counterfactual theories [of causation] described in the first two chapters may be said to have their origins in Hume's two definitions of causation (…) These theories take as their starting point some characteristic feature of causation – that causal relations instantiate regularities (…) or are marked by relations of counterfactual dependence. [1]

This brief mention, that is not explained anywhere in the two chapters mentioned, puzzled me very much and was actually the starting point of this investigation. The questions it rose in my mind, and that I intend to answer in this paper, were:

1) What is meant here by saying that Hume's two definitions of cause correspond to these two approaches in contemporary analysis of causation?

---

1   BEEBEE, H., HITCHCOCK, C., MENZIES, P. (Eds.) *The Oxford Handbook of Causation*, Oxford UP, 2009, p. 5.

2) More specifically, does Hume account of causation have anything to do with a counterfactual analysis of causation?

3) Can we say that, with his two definitions of cause, Hume intended to provide a theory of causation at all?

I came to find an answer to the first question in the opening paragraphs of David Lewis's 1973 paper on causation. Lewis says:

> Hume defined causation twice over. He wrote "we may define a 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." (…) Descendants of Hume's first definition still dominate the philosophy of causation: a causal succession is supposed to be a succession that instantiates a regularity. (…) Hume's "other words" – that if the cause had not been, the effect never had existed – are no mere restatement of his first definition. They propose something altogether different: a counterfactual analysis of causation.[2]

So, what Lewis (and possibly others) has in mind when he refers to Hume's "two definitions" of causation is actually the *two parts* of the first definition of cause as presented in the *Enquiry concerning Human Understanding*, section 7, §29.

Lewis's aim in this influential paper was to establish the counterfactual analysis of causation as a theory that is more suited than the regularity view to capture an intuitive aspect of our common understanding of what a cause is, and as able to solve some known difficulties of the regularity view.

In the first section of this paper, I will present briefly the general outlines of the regularity theory and of causation and some difficulties it encounters. In the next section, I will sketch briefly the counterfactual approach and show how Lewis's uses it to solve the above difficulties. I conclude by examining if Hume can be said to have a counterfactual theory of causation, or, indeed, any theory of causation of all.

1. The regularity view of causation and its difficulties

The first part of Hume's definition can very well be taken as a basis for a regularity view of causation. We have a regularity, in the form of events of type *A* being regularly followed by events of type *B*, and we have a particular event *a* of type *A* followed by an event *b* of type *B*, so we can say that *a causes b*. Causation is just the instantiation of a regularity.

---

2   LEWIS, David, "Causation". *The Journal of Philosophy* **70** (1973), p. 556, 557.

However, this does not seem to be enough. There are cases in which we want to speak of causes and effects although there is no perfect regularity. I press a switch and the light goes on. This occurs most of the time, but not always. Even so, we want to say, when it occurs, that the pressing of the switch *caused* the light to go on.

Because of this problem, modern formulations of the theory locate the strict regularity in *laws* that describe nomological regularities, and take into account the particular conditions that hold in the case. So, in the example, if we have some set of (true) laws of physics *L* and a set of true particular propositions *C* that describe the situation in all relevant aspects (there is power, the bulb isn't broken and is properly screwed, the wiring isn't damaged, the contacts in the switch are clean, etc), such that *L* and *C* together imply logically that if the switch is pressed the light goes on, then we can say that the first event is the cause of the second.

The regularity view of causation has, however, some problems. Suppose that *a* causes *b*, and that *b* does not cause *a*. Consider a set of laws *L* and a set of particular propositions *C* that imply logically that if the event *a* occurs then the event *b* also occurs. Let us suppose further that the situation is such that the effect *b* cannot occur except by being caused by the event *a*. In that case, from *L* and *C* it would logically follow that if *b* occurs then *a* also occurs. That is to say, the theory would allow us also to infer, wrongly, that *b* causes *a*. This is the problem of the *asymmetry* of the causal relation.

Another problem has to do with *collateral effects* of the same cause. Suppose *a* causes *b* and also causes *c*, and that *b* does not cause *c*. Consider a set of laws and conditions *L* and *C* that imply logically that if *a* occurs *b* occurs, and also that if *a* occurs *c* occurs. Then it would also follow logically that if *b* occurs *c* occurs. That is to say, the theory would allow us also to infer, wrongly, that *b* causes *c*.

With this in mind, let us proceed to the counterfactual analysis of causation as proposed by Lewis and see if it fares better.[3]

2. The counterfactual view of causation

While the regularity view aims to explain causation between two particular events not in terms of the events themselves but in reference to a regular conjunction of other similar events[4], the counterfactual approach tries to do

---

3   There is also another important problem, that of *pre-emption*, but I will not discuss it here.
4   Or, as Hume puts it, by drawing "from objects foreign to the cause" (*Treatise*, 1.3.14.31)

justice to the ordinary intuition that it is, after all, this particular event *a*, as such, that explains the occurrence of *b*, that it is *because* this particular event *a* occurred that this particular event *b* occurred, and not because other events occurred formerly or elsewhere. The idea is that the cause must, so to say, make a difference in the situation, that it controls the occurrence of the effect, that the occurrence of the effect *depends* on the occurrence of the cause.[5]

So, in our former example, when we say that my pressing of the switch was the cause of the light going on, what is meant, according to the counterfactual theory, is that *if I had not* pressed the switch on that particular occasion (which I did), the light would have stayed off. Alternatively, just to consider another example, if I merely put my finger on the switch without actually pressing it, we want to say that *if I had* pressed the switch on that particular moment (which I did not), the light would have gone on.

If we want to give meaning to such subjunctive counterfactuals, we cannot treat them as mere material implications (if *a* then *b*) as in the regularity theory, because they would be trivially true, since their antecedents are *ex hypothesi* false. Now, if I press (or refrain to press) a switch, this is something that happens in the world, and once done, cannot be undone. Thus, the only way to give any useful meaning to a counterfactual statement is to consider that it refers, not to the real world, but to a *possible world*, in which things happened differently than they did in the real world.

Of course, we cannot simply mean that in *any* possible world where I do not press the switch the light stays off, because there are countless possible worlds in which the light still goes on for a variety of reasons, such as the operation of other causes. We need to restrict the set of possible worlds in order to arrive at something useful for establishing the notion of causal dependence between events. Lewis believes he can solve this problem through a notion of *comparative similarity* of possible worlds.

In short, he believes that all we have to say is that, among all the possible worlds in which I *did not* press the switch, the one that more closely resembles the real world is one in which the light *did not* go on.[6] If everything remained pretty much the same, except for my not pressing the switch, this is the result to be expected, since the operation of other causes would introduce some major differences between the worlds.

---

5   This is not to be confused, however, with the ancient theory of causation as influence, according to which the cause "produces" the effects, or is "connected" to it by some sort of ontological necessity.

6   Of course, this crucial notion of comparative similarity is very tricky and much work has been done and still needs to be done to clarify it.

Of course, in order to reach such conclusions, we must rely in nomic regularities such as appear in the regularity theory, but the point of the counterfactualist isn't to dispense with regularities, but only to show that they are not sufficient to establish the notion of causal dependence involved in a proper analysis of causation. This becomes clear when we examine the way Lewis proposes to solve the problems we identified in the regularity approach to causation.

So, the problem of the asymmetry of cause and effect is solved by showing that even if both material conditionals "if *a* occurs then *b* occurs" and "if *b* occurs then *a* occurs" are logically implied by the same set of general laws and particular propositions, we can still identify which is the cause and which the effect because of the asymmetry of the notion of causal dependence. We want to say that the effect depends on the cause and not the cause on the effect. How this is done can be seen in the following example:

Suppose that the atmospheric pressure is dropping quickly at a certain place, and the needle of a barometer situated there is also dropping quickly. Suppose further that the laws of physics and the particular propositions that are true of the barometer and its immediate surroundings allow us to deduce logically that, if the atmospheric pressure is dropping at a certain time, then the barometer's needle is also dropping at that same time. By the regularity view, this allows us to say that the drop in pressure is the cause of the drop of the needle. However, since the barometer is operating properly, the reciprocal is also true: if the barometer's needle is dropping at a certain time then the atmospheric pressure is also dropping at the same time. According to the counterfactualist, however, we cannot say that the dropping of the needle *causes* the dropping of the pressure because we cannot say that the latter *depends* counterfactually on the former, that is, we cannot say that, in the precise circumstances, *if the needle stopped to drop the pressure would also stop to drop*. Among all the possible worlds in which the needle stops to drop there will be many in which the pressure also stopped to drop, but they are by far less similar to the real world (considering the immense changes in the atmosphere involved) than a possible world in which the atmospheric situation is much the same and the needle simply got stuck due to some small causal interference.

A parallel solution can be given to the problem of collateral effects. A sudden drop of pressure causes both a drop in the barometer's needle and the approach of a storm later on. By the regularity theory, we must agree that, in the circumstances and given the laws of physics, if the needle drops then a storm approaches, but this does not allows us to say that the drop of the

needle caused the approach of the storm for pretty much the same reasons we discussed in the former example.

3. Does Hume have a counterfactual theory of causation?

For all the intrinsic interest of the counterfactual analysis of causation, the question that interests me is whether Lewis (and possibly others) is right when he names Hume as a forerunner of the counterfactual approach. I will present briefly some reasons why I think this attribution is wrong.

Let us consider first the textual argument. Lewis is certainly right when he observes that the second part of the first definition in the *Enquiry* says something very different of what is said in the first part. In the first part, a cause is presented as a sufficient condition for the occurrence of the effect, and in the second part, it is presented as a necessary condition, so they really appear to be different definitions. Also the very words with which Hume formulates the second part – *if the first object had not been, the second never had existed* – sound remarkably in tune with the subjunctive conditionals employed in the counterfactual approach.

On the other hand, Hume does not seem to believe he is giving a different definition in the second part, or saying something different from what he said in the first part. He just seems to think it is another way of stating the same point "in other words", as he says. Certainly, some explanation is in order here.

I think we can understand better what is happening if we go back to the first definition of cause as given in the *Treatise* (1.3.14.31). There Hume says that a cause is "an object precedent and contiguous to another, and where all the objects resembling the former are plac'd in like relations of precedence and contiguity to those objects that constitute the latter." This definition, as we see, characterizes a cause just as a sufficient condition for the occurrence of the effect, and makes no mention of its status also as a necessary condition of the effect, as expressed in the counterfactual wording of the definition in the *Enquiry*.

But we know that Hume conceived causes in the *Treatise* also as necessary conditions of their effects; and although he did not include this clause in the definition at 1.3.14.31, he stated it explicitly in the *Rules by which to judge of causes and effects*, namely as the Fourth Rule: "The same cause always produces the same effect, and the same effect never arises but from the same cause". This formulation is clear-cut and refers obviously to actual successions of events taking place in the actual world; there is no need

to resort to possible worlds to make sense of what he says here. My proposal is that we should understand the second part of the definition in the *Enquiry* ("if the first object [the cause] had not been, the second [the effect] never had existed") as meaning exactly the same as "the same effect never arises but from the same cause", and with exactly the same ontological commitments, in spite of its (perhaps unfortunate) wording.

If the textual argument can be answered in this way, it is not clear that conceptual side of the question is so easy to handle. On the one hand, I do not think there is support in Hume's text for any notion of causal or counterfactual dependence between events of the kind Lewis needs for his theory. On the contrary, any "strong" connection between events can only be the (causal) result of the operation of the habit after being conditioned by the experience of regular successions of events, and there seems to be no place for reasonings such as are based in the relative similarity of possible worlds. One should note, moreover, that, for Hume, the only notion of possibility is logical possibility, which would make the notion of possible worlds so encompassing to the point of depriving it of any utility.

On the other hand, there is one interesting aspect in the counterfactual model that could very well be accepted by Hume as a means to implement and make operational his "Rules by which to judge of causes and effects". I mean the "controllability" or "manipulability" inherent to the notion of counterfactual dependence, by means of which we can make things happen, or, reciprocally, to prevent some occurrences, and use the results from these operations as a means of testing hypotheses about causal connections.

Even so, the chief interest of this procedure would be not so much to gain an intellectual insight into the nature of causality itself as to help us to refine and amplify our natural capacity to make causal inferences.

4. Does Hume have any theory of causation at all?

This brings us finally to the question of whether Hume is actually trying to propose a theory of causation in the sense of those two that we examined above, a theory that would provide us with necessary and sufficient conditions to decide, for any pair of particular events, if they are to be subsumed under the relation of cause and effect or not.

Sometimes it could appear to be so, since Hume actually gave definitions of the notion, as if he were intending to provide a philosophical analysis of causation. If we interpret Hume in this way, we would say that his theory is very close to the regularity view that we examined above, although in a more

primitive form. His theory would also include some elements that are not normally part of the regularity theory, as his demands of temporal succession and spatial contiguity between cause and effect. Such elements would perhaps help to solve (in some cases) some difficulties of the regularity theory, as the problem of asymmetry and of collateral effects. On the other hand, they would introduce some drastic modifications in the way the notion of cause and effect is normally understood: for example, the demand that causes be spatially contiguous with their effects would seem to forbid the transitivity of the causal relation.

If, however, we follow the line of his argument (which is presented much more clearly in the *Enquiry* than in the *Treatise*, it should become clear that Hume's chief problem, the one that takes a central place in his investigation, is how we manage to have beliefs (and very successful ones, at that) about unobserved matters of fact, that is, matters of fact that lie outside the field of our senses and memory. We manage this because we are able to make causal inferences, but these inferences are not based in some conceptual knowledge that we have about the nature of causation, but are made spontaneously as a result of our exposition to regular succession of events and a natural propensity to associate in our minds the kinds of events that we observed to follow regularly each other.

If we are to make correct causal inferences based on regularities, it does not matter much if we infer from causes to effects, or from effects to causes or even from an effect to another effect collateral with the first. We will arrive mostly at correct predictions, and this is what matters. We are busy, at first, with building a repertory of pairs of ideas that became associated in our minds and allow us to infer from one member the occurrence of the other. As our experience is limited and our powers of discrimination are not very keen, the class of pairs that we manage to assemble will be imperfect compared with the actual regularities that really hold in the world. So, another step must be taken, and it is here that the definition of cause has a role.

The class of pairs of ideas that we came to associate in this way corresponds, in Humean terms, to our abstract idea of cause and effect; and a definition of a term, for Hume, is no more than a way to present the class of particular ideas that constitute the abstract idea corresponding to that term. This is how we should understand the role of Hume's definition of cause. It serves to refine and correct our abstract idea of cause, allowing us to include and exclude some particular pairs from the class that constitutes that idea. Far from being a theoretical that would allow us to construct such a class *ab ovo*, the definition only enables us to refine a notion that we can originally acquire only

by completely different and independent means. Hume's definition of cause belongs properly among his "rules by which to judge of causes and effects", and the fact that the first three rules are a precise restatement of the original definition shows exactly what is its place and function in Hume's system.

Referências bibliográficas

BEEBEE, H., HITCHCOCK, C., MENZIES, P. (Eds.) *The Oxford Handbook of Causation*. Oxford: Oxford University Press, 2009
HUME, David. *A Treatise of Human Nature*. Oxford: Oxford University Press, 2000
HUME, David. *An Enquiry Concerning Human Understanding*. Oxford: Oxford University Press
LEWIS, David, "Causation". The Journal of Philosophy 70 (1973)