CDD: 121

THE RELATION BETWEEN THE GENERAL MAXIM OF CAUSALITY AND THE PRINCIPLE OF UNIFORMITY IN HUME'S THEORY OF KNOWLEDGE

JOSÉ OSCAR DE ALMEIDA MARQUES

Departamento de Filosofia
Instituto de Filosofia e Ciências Humanas – IFCH
Universidade Estadual de Campinas
Cidade Universitária Zeferino Vaz
13083-970 CAMPINAS - SP
BRASIL

jmarques@unicamp.br

Received: 08.12.2010; Accepted: 06.01.2011

Abstract: When Hume, in the *Treatise on Human Nature*, began his examination of the relation of cause and effect, in particular, of the idea of necessary connection which is its essential constituent, he identified two preliminary questions that should guide his research: (1) *For what reason we pronounce it necessary that every thing whose existence has a beginning should also have a cause* and (2) *Why we conclude that such particular causes must necessarily have such particular effects?* (1.3.2, 14-15) Hume observes that our belief in these principles can result neither from an intuitive grasp of their truth nor from a reasoning that could establish them by demonstrative means. In particular, with respect to the first, Hume examines and rejects some arguments with which Locke, Hobbes and Clarke tried to demonstrate it, and suggests, by exclusion, that the belief that we place on it can only come from experience. Somewhat surprisingly, however, Hume does not proceed to show how that derivation of experience could be made, but proposes instead to move directly to an examination of the second principle, saying that, "perhaps, be found in the end, that the same answer will serve for both questions" (1.3.3, 9). Hume's answer to the second question is well known, but the first question is never answered in the rest of the *Treatise*, and it is even doubtful that it could be, which would explain why Hume has simply chosen to remove any mention of it when he recompiled his theses on causation in the *Enquiry concerning Human Understanding*.

Given this situation, an interesting question that naturally arises is to investigate the relations of logical or conceptual implication between these two principles. Hume seems to have thought that an answer to (2) would also be sufficient to provide an answer to (1). Henry Allison, in his turn, argued (in *Custom and Reason in Hume*, p. 94-97) that the two questions are logically independent. My proposal here is to try to show that there is indeed a logical dependency between them, but the implication is, rather, from (1) to (2). If accepted, this result may be particularly interesting for an interpretation of the scope of the so-called "Kant's reply to Hume" in the Second Analogy of Experience, which is structured as a proof of the *a priori* character of (1), but whose implications for (2) remain controversial.

Keywords: Hume. Causality. Regularity. Principle of uniformity.

I Hume's two causal principles and their logical relations

At the beginning of Part 3 of Book 1 of the *Treatise,* Hume proposes to investigate the origin of the idea of causation, i.e., to identify the impression that it would copy. He identifies three components of the idea of cause and effect: spatiotemporal contiguity between cause and effect, temporal priority of the cause over the effect, and necessary connection between cause and effect (1.3.2, 7-11).

Of these, the last one is the essential component. Since an examination of the objects themselves (events) as to their qualities and their relationship did not reveal any impression that could give rise to the idea of necessary connection, Hume proposes to address the problem indirectly from an examination of two questions:

(1) *For what reason we pronounce it necessary that every thing whose existence has a beginning should also have a cause* and (2) *Why we conclude that such particular causes must necessarily have such particular effects?* (1.3.2, 14-15)

That is, Hume asks for the reasons of our belief in two maxims or principles:

P1) Everything that begins to exist must have a cause of its existence. (1.3.3, 1)

P2) Particular causes must necessarily have such particular effects. (1.3.3, 9)

or, more fully:

P2') The same cause always produces the same effect, and the same effect never arises but from the same cause. (Rule 4 of the "Rules by which to judge of causes and effects" (1.3.15, 6))

Hume observes that our belief in these principles can result neither from an intuitive grasp of its truth nor from a reasoning that could establish them by demonstrative means. In particular, with respect to the first, Hume examines and rejects some arguments with which Locke, Hobbes and Clarke tried to demonstrate it, and suggests, by exclusion, that the belief that we place on it can only come from experience. Somewhat surprisingly, however, Hume does not go on to show how this derivation of P1 from experience could be done, but proposes to move directly to the second proposition, saying that "maybe in the end, the same answer will serve to both questions" (1.3.3, 9). Hume's answer to the second question is well known, but the first question is never answered in the rest of the text of the *Treatise,* and it is even doubtful that it could be answered, a fact that would explain why Hume chose simply to suppress any mention of it when he recompiled his theses on causation in the *Enquiry on Human Understanding.*

Given this situation, a question that naturally presents itself is to investigate the relations of logical or conceptual implication between these two propositions. Hume's procedure suggests that he might have considered that there would be a relation of implication between P2 and P1, so that, by providing an experimental justification for P2, he would *ipso facto* have justified P1, and this could very well be the reason why he never returned to examine this last principle. This interpretation was first proposed by Fred Wilson[1] and his argument will be discussed in detail in the next part of my presentation.

Henry Allison, in turn, argues that the two issues are logically independent[2]. He does not provide, however, a detailed argument for this claim, and merely says that "one can consistently maintain that every beginning of existence must have *some* cause while denying that any *particular* cause must have a *particular* effect, and vice versa.' (p.94).

---

[1] Wilson (1982, pp. 1-18).
[2] Allison (2008, pp. 94-97).

That is, he proposed that P2 is consistent with ~P1 (with which I agree) and that P1 is consistent with ~P2 (with which I disagree, *cf.* the end of my presentation).

The question that interests me, then, in this context is: Has Hume, as proposed by Wilson, established experimentally P1 albeit indirectly through the establishment of P2, or, as Allison suggests, there is no logical implication from P2 to P1, and therefore P1 receives no justification in Hume's empiricist system? This question is interesting for Hume scholars, but is particularly important for the discussion of so-called "Kant's reply to Hume," as influentially formulated by Lewis Beck[3]. In fact, Beck's argument supposes:

1) that Hume did not establish P1 (neither demonstratively nor by experience)

2) that gaps in the series of impressions result in a violation of P2.

3) that Hume employs P1 to "save" P2, although P1 is as much affected by the gaps as P2.

4) that, therefore, Hume's use of P1 is not legitimated within his system, which shows that he must have treated it as *a priori* valid (and this would be the "Prussian" element in Hume).

Thus, authors who agree that the Second Analogy of Experience provides an answer to Hume, such as Allison and Beck, are understandably interested in denying that P1 is established within Hume's system, and therefore cannot accept that P2 (which Hume *does* establish) somehow implies P1. On the other hand, authors like Wilson, who does not accept that Hume has committed the inconsistency of using an *a priori* principle in his explanation of causality, may find in the alleged implication of P2 to P1 a way to neutralize a crucial step in Beck's argument.

---

[3] Beck (1978, p. 111).

It is not my intention here to investigate whether the Second Analogy, as interpreted by Beck, is an effective response to Hume[4]. Nor do I intend to decide whether P1 is, after all, experimentally established within Hume's system. My goal is simply to examine the logical relations between P1 and P2; in particular, I wish to show, against Wilson, that P1 does not follow from P2, and, against Allison, that the reverse is true, i.e., that P2 is conceptually implied by P1. This latter result, if accepted, could be particularly interesting for an interpretation of the scope of Kant's Second Analogy of Experience, which is articulated as a proof of the *a priori* character of P1, but whose implications as regards P2 remain controversial.

II The implication P2 → P1

Before turning to the arguments for the thesis I presented above, I will spend some time examining in detail (though I find it misleading) the argument presented by Fred Wilson for his proposal that P2 implies P1. Although his argument ultimately fails, this examination will allow me to introduce some important elements for the subsequent discussion.

Wilson deals with P2 in its stronger version, which appears in Rule 4 of section 15 quoted above: "The same cause always produces the same effect, and the same effect never arises but from the same cause." This can best be presented with the aid of following diagram:

[4] To this end, a way that seems to me more productive than Wilson's has been proposed in Falkenstein (1998, pp. 331-360).

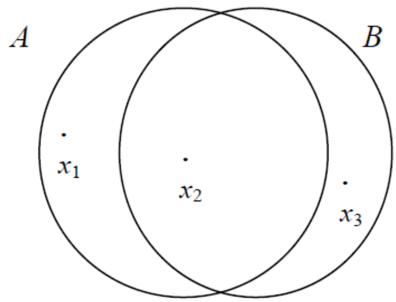

Figure 1

The circles A and B here represent *types* or *classes* of events, and the points x_i represent particular instances of their occurrence[5]. As these circles have a good portion of their surface in common, the situation is such that most occurrences of A coexist with occurrences of B, i.e., there is an approximate regularity in the conjunction of these two events. But by principle P2, we cannot say that in this situation A is the cause of B, since x_1 represents a case in which A occurs and the supposed effect B does not occur, and x_3 a case in which B occurs without the presence of the supposed cause A. In order that one could say that A is the cause of B, it would be necessary, according to principle P2, that there were an exact match between the two circles. In formal terms:

(1) $(x)(Ax \equiv Bx)$

In his treatment of the problem, Wilson considers a case in which we have several hypotheses as to the supposed cause of a certain effect, and we must identify which of them is its real cause. For this, he

---

[5] More precisely, the x's should represent strictly delimited regions in space and time, so that events contained in one of these regions are spatially and temporally contiguous. Thus, region x_2 contains an event of type A and another event of type B contiguous to the first.

draws upon another Humean maxim expressed in Rule 5 of the section on *Rules by which to judge of causes and effects*. Hume says:

> There is another principle which hangs upon this [Rule 4], *viz.* that where several different objects produce the same effect, it must be by means of some quality which we discover to be common amongst them. For as like effects imply like causes, we must always ascribe the causation to the circumstance wherein we discover the resemblance. (1.3.15, 7)

Wilson's argument is complex and involves aspects that can be disregarded for the purposes of this presentation. I will provide, then, a simplified version that preserves, as far as I understand, the core of his argument.

Suppose we have an effect B and two events A_1 and A_2 which may *prima facie* be considered its causes. B is represented by the whole circle and A_1 and A_2 are the two semicircles determined by the vertical line.

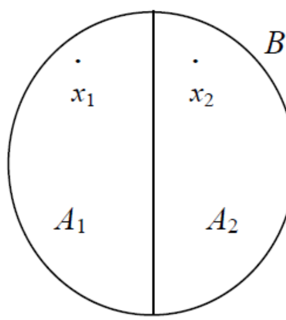

Figure 2

In this case we have:

(2) $(x)(A_1 x \supset Bx)$

and also:

(3) $(x)(A_2 x \supset Bx)$

That is, every occurrence of either A_1 or A_2 are followed by occurrences of *B,* thus both A_1 and A_2 satisfy the first part of Rule 4 (same causes produce same effects), but the second part of the rule is not satisfied, since the same effect *B* is conjugated to two different apparent causes, which, therefore, can neither be the real one. In this case, by Rule 5, we can anticipate that there will be another event *A*, which subsumes the common aspect of A_1 and A_2 and will be revealed as the true cause of *B*. Formally:

(4) $(\exists A)(x)(Ax \equiv Bx)$

Or, in other words, for any strictly limited region of space and time *x*, there exists a determinate type of event *A* such that, if *x* contains a particular event of type *B*, then *x* also contains a particular event of type *A*, and vice-versa.

As an intuitive model of the situation, consider a bulb that is lit *(B)* by the operation of any one of two parallel switches *(*A_1 and A_2*)*. Then one cannot say that the pressing of any of the switches is, as such, *the cause* of the lightening of the lamp. The "true" cause, in the case, could be identified to the application of a voltage to the bulb poles.

Formulas (2), (3) and (4) together affirm that, given a certain effect *B* for which a number of causes present themselves as candidates, there exists an event[6] that will be its real cause, namely, whose occurrence constitutes a sufficient and *necessary* condition for the occurrence of that effect. In Wilson's words, formula (4) (or, rather, the corresponding formula in his exposition) asserts that "for this *effect B* there is always a cause" (Wilson, p. 3-4, my italics).

---

[6] Strictly, what (4) affirms is that there is at least one [type of] event invariably associated to *B*, but as all those events will be extensionally equivalent under the aspect of their causal relation to *B*, they may be considered as one and the same [type of] event given under different descriptions.

Wilson then considers the result of a reversal in the position of the quantifiers in (4). This provides the formula:

(5) $(x)(\exists A)(Ax \equiv Bx)$

For Wilson, this formula also says that for *B* there is always a cause, but, unlike (4), does not say that this cause is the same in all cases. In his words, (5) states that "the effect always has *a* cause, but allows it to be a different cause on different occasions" (p. 4). Thus, in the example in Figure 2, *B* [the lightening of the lamp] always has a cause, but this is sometimes A_1 [the pressing of the first switch], sometimes A_2 [the pressing of the second switch].

Wilson's final step is to suggest that (4) provides the logical form[7] of P2 and, correspondingly, that (5) provides the logical form of P1. As it is a quite elementary result in the predicate calculus that (4) logically implies (5), or that (5) follows from (4), Wilson believes to have shown that P1 follows logically from P2 and that Hume, when he established the latter, *ipso facto* established P1, and was therefore legitimately exempt from returning to the issue in the remainder of Part 3 of Book 1 of the *Treatise.*

Wilson's argument has, however, a fundamental flaw, which may have already become apparent during my presentation. I noticed that (4) is a plausible formulation of principle P2, but *only* in conjunction with formulas like (2) and (3) that allows us to characterize *B* as an *effect,* i.e., as an event that already appears in regular conjunctions with other events that are presented, in preliminary form,

[7] Wilson notices that formulas (4) e (5) are still too specific, since they refer to a certain effect *B*. But as *B* is left undetermined and is, in logical terms, a free variable in the formulas, these can be closed by means of the universal quantifier, producing sufficiently general formulas that now hold for any event and maintain with each other the same relation of logical implication as before (Wilson, p.5):

(4') $(B)(\exists A)(x)(Ax \equiv Bx)$

(5') $(B)(x)(\exists A)(Ax \equiv Bx)$

as its possible causes[8]. Rule 5 is very clear about this: it only applies when "several different objects produce the same effect." But when Wilson performs the inversion of quantifiers to obtain formula (5), which he takes as symbolizing principle P1, he falls into a deadlock:

(i) if he maintains in (5) the same requirement that *B* is interpreted as an event, he will actually have proved that (5) follows from (4), but all that (5) states, in this interpretation, is that every *effect* has a cause, which is undoubtedly true, but only trivially, and not at all a formulation of P1.[9]

(ii) If he expands the interpretation of *B* in (5) to cover any kind of event[10], then certainly (5) represents a correct version of P1, but it is no longer possible to infer it from (4).[11]

III The implication P1 → P2

It seems safe to say, therefore, that Wilson has not demonstrated that P1 follows from P2. But that alone does not mean that another demonstration could not be discovered; thus, to conclusively resolve the issue, it would be necessary to prove the consistency between P2 and the negation of P1. One way to prove, in logic, the consistency of two propositions is to provide a model in which both are satisfied. I think it is possible to provide such a model in this case.

---

8 Wilson himself twice uses the word "effect" to designate *B* when presenting these formulas.

9 Cf. *Treatise* 1.3.4, 8: "Every effect necessarily presupposes a cause; effect being a relative term, of which cause is the correlative. But this does not prove, that every being must be preceded by a cause; no more than it follows, because every husband must have a wife, that therefore every man must be married."

10 As he in fact does, at p.5.

11 This critique has already been made to Wilson by Allison, although without much detail. See Allison (2008, p. 356).

The model I propose is a (potentially) infinite sequence formed by the letters A, B, C, D, E, in which every A is followed by B, all B is preceded by A, and there is no other pair of letters that display the same regularity. This sequence can be constructed by the rules:

R1: $\varnothing \rightarrow$A

R2: A $\rightarrow$ B

R3: B $\rightarrow$\{A, C, D, E\}

R4: C $\rightarrow$\{A, B, D, E\}

R5: D $\rightarrow$\{A, B, C, E\}

R6: E $\rightarrow$\{A, B, C, D\}

Rule R3 must be understood as stating that the first occurrence of B in the sequence is followed by A, the second by C, and so on, returning to A on the fifth occurrence. The other rules work similarly. I present below the beginning of the sequence thus constructed:

ABABCABDABEABABCBDBEBABCDCECABDEDABEABABCB

Let us now interpret this sequence as representing a series of occurrences of events of types A, B, C, D, E, among which we will investigate the existence of causal connections. If we adopt the criterion that causation involves regular and invariable successions of events, only successions AB are here to be classified as connections of cause and effect. In this case, our sequence does satisfy P2, for in it the same cause has always the same effect, and vice versa. But it *does not* satisfy P1 (or, which is the same, it satisfies the negation of P1), for it contains events that are not caused, according to the criterion we adopted.

Here, however, an objection might be raised. There is indeed no regularity in the sequences involving the events C, D and E, but why should that mean that they are not causes neither effects of the events that precede and succeed them? After all, couldn't we conceive that there are causes even though no regularity is displayed? Why not suppose that, at the beginning of the sequence, B caused A, then caused C, then caused D? In this case, the sequence of events represented could as well be described as satisfying P1 and not satisfying P2[12]. Why should one prefer one interpretation to the other?

The answer is that we *can* adopt this interpretation, but at the expense of being forced to adopt as well a conception of causes as "powers," "strength" or "influences", i.e., as an intrinsic property of the events themselves, which they possess irrespective their relations with other events, and through which they would be able to "produce" their effects. This is a respectable and traditional conception of causality, characteristic of Cartesian and scholastic philosophy, but is also one that Hume, I believe, decisively rejected and devoted himself to refute in the *Treatise*.[13]

Thus, if we adopt Hume's conception of causality and necessary connection, whose establishment rests entirely on the regularity of the successions of events, it is not conceptually possible to provide an example where P1 holds but P2 does not. To see this, it is enough to consider the "first definition" of cause proposed by Hume in the *Treatise*: "an object precedent and contiguous to another, and where all the objects resembling the former are plac'd in like relations of precedency and contiguity to those objects that resemble the latter""[14]. If we introduce now xRy as a relationship that exists between particular instances of events x and y whenever x and y are spatiotemporally

12 In fact something like that must be the case for Allison to be correct in his proposal that "one might consistently hold that every beginning of existence must have *some* cause [P1], while denying that any *particular* cause must have a *particular* effect [P2]" (2008, p. 94).

13 Of course there are interpretations of Hume that suggest that he might allow the operation of causal powers behind the horizon of our perceptions, but these causal powers would then have the function of *explaining* the observable regularities, not of dispensing with them.

14 *Treatise*, 1.3.14.31.

contiguous and x precedes y, the proposition "A causes B" (A and B being undetermined *types* of events) can be formally *defined* as:

(6) A causes $B =_{\text{def}} (x)\,[Ax \supset (\exists y)\,(xRy \;\&\; By)]$

or, more fully, incorporating the second clause of Rule 4 that requires regularity also from effects to causes:

(7) A causes $B =_{\text{def}} (x)\,[Ax \supset (\exists y)\,(xRy \;\&\; By)] \;\&\; (x)\,[Bx \supset (\exists y)\,(yRx \;\&\; Ay)]$

What this shows is that, in a situation in which events of the same type were not always followed (in the sense stipulated by the relation R) by events of the same type, and, conversely, that events of the same type were not always preceded by events of the same type, *no causal relationship would exist, by definition.* Therefore, what follows is that, if causal relations were to hold generally between events, this would imply the truth of P2 (same causes, same effects and same effects, same causes). Now P1 says that every event (of any type) has a cause, thus P1 (assuming that there are events at all) asserts the general holding of causal relations, thus P1 logically implies P2, *qed.*

References:

ALLISON, H. E. *Custom and Reason in Hume*. Oxford: Clarendon Press, 2008.

BECK, L. W. "A Prussian Hume and a Scottish Kant". In: Beck (ed) (1978), pp. 111.

————. *Essays on Kant and Hume*. New Haven and London: Yale University Press, 1978.

FALKENSTEIN, L. "Hume's Answer to Kant". *Noûs* 32:3, 1998, pp. 331-360.

HUME, D. *A Treatise of Human Nature*. David Fate Norton e Mary J. Norton (eds.). Oxford: Oxford University Press, 2000.

WILSON, F. "Is there a Prussian Hume? Or How Far Is It from Königsberg to Edinburgh?" *Hume Studies,* vol. VIII, n.1, April 1982, pp. 1-18.