

# Kant's Proof of a Universal Principle of Causality: A Transcendental Idealist's Reply to Hume

REZA MAHMOODSHAHI

In his famous dictum, Lord Russell remarked: "The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm."<sup>1</sup> Russell took the principle of causality to be entirely incoherent, and it's no wonder: since Hume, philosophers have thought of 'causality' as a metaphysically dubious concept, one which purports a mysterious *necessary connection* between an event A and its respective effect B. Hume's momentous critique of the rationalist principle spawned a contemporary debate, one which undoubtedly motivated the entire Kantian enterprise, but one to which Kant also directly contributed in the Second Analogy of the Transcendental Analytic.

In the introduction to the *Prolegomena*, Kant summarized Hume's accomplishment: "he proved incontrovertibly that it is entirely impossible for reason to think such a combination *a priori* and from concepts, for such a combination contains necessity; but it absolutely cannot be conceived why, because something is, some else must also necessarily be, and thus how the concept of such a connection can be introduced *a priori*" (4:257). Hume demonstrated that the rationalist *a priori* principle of causality is groundless, for "when we look about us towards external objects, and consider the operation of causes, we are never able, in a single instance, to discover any power or necessary connexion; any quality, which binds the effect to the cause."<sup>2</sup> Causation, for Hume, has mere inductive status; as such, it is not determinate and only succeeds in establishing a contingent connection between two events. The occurrence of an event A immediately and regularly followed by an event B is not an instantiation of the rationalist notion of necessary connection; rather, the mistaken construal of A and B's close arrangement as a necessary one is a consequence of mere habit of mind brought about by the constant 'conjunction' of the two events in experience.<sup>3</sup> In short,

event A does *not* cause B, but merely precedes it in occasion.

Kant thought that the only way to vindicate any principle of causality was to abandon attempts to derive its necessity through experiential grounds; "it must either be grounded completely *a priori* in the understanding or be entirely abandoned as a mere chimera" (B123). As Hume demonstrated, the explanatory efficacy of experience is necessarily limited to the observation of customary occurrences through which at best I might be able to affirm that in all formerly observed instances of A, B subsequently follows. Such grounds fall short of what's needed, that is, some grounds through which to derive a necessary law to which all-future, hitherto unknown, experience must conform. In recognizing that the objective reality of an *a priori* principle of causality can only be established through *a priori* means, Kant's reply to Hume must therefore be understood as an attempt to positively establish the concept through an appeal to the understanding, where "the effect is not merely joined to the cause, but rather is posited *through* it and results *from* it" in accordance with a universal rule (B124).

In the Transcendental Analytic, Kant treats the understanding as the source of *a priori* concepts, which along with the forms of intuition, give rise to *a priori* cognition. Kant derives the pure concepts of the understanding, or the *categories*, from twelve logical functions or forms of judgment. These twelve logical functions are supposed to serve as 'clues' to the corresponding ways in which we form concepts of objects. On the supposition that the "understanding is completely exhausted and its capacity entirely measured by these [logical] functions" (B107), Kant derives his Table of Categories: twelve categories for conceiving of the quantity, quality, relation, and modality of objects (B106). Kant goes on to argue in the transcendental deduction that *all* twelve pure concepts of the understanding apply universally and necessarily to the objects of experience. His argument here relies on the "transcendental unity of apperception": a single unitary consciousness or continuous string of experiences is possible if and only if our intuitions, procured through the sensibility, are synthesized via thought *through* the categories so as to present us

with the objects of experience. The application of the categories to what we might call our 'sense-data' is a necessary condition for the representation of the objects of experience. In the second analogy, the category of interest—derived from the hypothetical form of judgment—purports to explain causal relations and dependencies (B106) amongst the objects of experience for "only thereby can I be justified in saying of the appearance itself, and not merely of my [own subjective] apprehension, that an [objective] sequence is to be encountered in it..." (B238).

### SUBJECTIVE-OBJECTIVE SUCCESSION

The "Analogies of Experience", of which the Second Analogy is a part, concerns the class of categories Kant calls relations. The relational category of causality, once applied to what's given to us in space and time, necessarily grounds "the real upon which, whenever it is posited, something else always follows" (B183). The argument for causality relies on a distinction between an objective and a subjective succession of representations, since Kant takes judgments concerning the objective alterations of the states of a substance to be justified if and only if every objective alteration behaves according to a necessary rule of succession, viz. causality.

The analogies of experience, broadly speaking, rely on two assumptions: (i) the unity of apperception and (ii) the application of schematized categories. Again, "the unity of apperception" requires the necessary connection of perceptions and the synthetic unity of appearances in a *single time*. This ensures one, and not many, temporal intervals. The second assumption arises out of the need to place events along a temporal interval despite an inability to perceive time in itself. Time understood in abstraction from its phenomenal content tells us that we must pass through  $T_1$  before we reach  $T_2$ . We cannot experience  $T_1$  after or at the same time as  $T_2$ . It is through this trivial precept of time-relations that we avoid the contradictory notion of  $T_1$  as both present and future. That is,  $T_1$  which is prior to  $T_2$ , cannot be both simultaneous and subsequent to  $T_2$ , for "successive periods of

time constitute a series in which no one period can bear the same relation to that which precedes and that which follows."<sup>4</sup> Accordingly, *objective* time-relations are of two sorts: successive and simultaneous (co-existent). The Second Analogy turns to the successive order of our *subjective* perceptions and asks whether these successive perceptions of the states of a substance could have been ordered differently. To put it more precisely, given that private perceptions of the objects of experience constitute a successive sequence, are there sequences of perceptions such that the temporal-order is irreversible?

Kant's thought is that if the temporal-order of a sequence of perceptions is irreversible (and certain other conditions hold), then our objective experience is possible only through the application of an *a priori* concept of the understanding. In other words, our experience of objective events presupposes the application of the causal category. Alternatively, if our apprehension of the manifold yields a sequence of perceptions such that the temporal-order *is* reversible, then in virtue of the reversibility of the subjective succession of representations, we know that no objective event has occurred. The absence of an objective event implies an indeterminate, wholly subjective temporal-order. An object that is not successive in itself is apprehended in some unique temporal order merely because our apprehension of the manifold of appearances is always successive (B234). In the absence of an objective event, we *know* that the states of the substance itself are co-existent; though our perceptions of it might occur in some other temporal order, such an order is contingent upon *our* assorted perceptual freedoms, e.g. scanning left-to-right, right-to-left, top-to-bottom, and not determined by succession in the object itself.

"Thus, e.g., the apprehension of the manifold in the appearance of a house that stands before me is successive. Now the question is whether the manifold of this house itself is also successive, which certainly no one will concede" (B235). Let's call our perception of the roof of a house  $A_R$  and our perception of the doorway  $B_R$ , and let's assume that  $A_R$  and  $B_R$  are independently perceptible. The house is meant to exemplify an object in which A and B do not succeed one another. Rather, they are co-

existent since it is possible to experience either  $A_R$  or  $B_R$  prior to the other.  $A_R$  and  $B_R$  possess what Strawson calls “order-indifference”<sup>5</sup>, in view of the fact that  $[A_R B_R]$ -irreversibility does not hold. To use Beck’s terminology<sup>6</sup>,  $[A_R B_R]$  does not imply the objective event  $[AB]$ , which symbolizes a state  $A$  in an object which precedes a state  $B$  in an object. Nothing has happened; no objective event has occurred; no state has come to be in a substance that formerly was not (B237).

Kant contrasts this sequence of successive perceptions of a house, which does not constitute an objective event (given that the manifold is not apprehended in a necessary order), with successive perceptions of a ship driven downstream. A moving ship is meant to serve as an obvious example of a sequence of successive perceptions that lacks “order-indifference”, and hence constitutes an objective event. “My perception of its position downstream follows the perception of its position upstream, and it is impossible that in the apprehension of this appearance the ship should first be perceived downstream and afterwards upstream” (B237). The subject’s various perceptual freedoms, e.g. scanning left-to-right, right-to-left, top-to-bottom, etc., have no bearing on the temporal-order of the successive perceptions—the order is objectively determined. Let’s call our perception of the ship upstream  $A_R$  and our perception of the ship downstream  $B_R$ . As a result of the successiveness of the object itself, it is not possible to view  $B_R$  prior to  $A_R$ , all subjects necessarily apprehend  $A_R$  prior to  $B_R$ , i.e.  $[A_R B_R]$ -irreversibility holds. Apprehension is “bound to” the order of the sequence of perceptions.

Causality figures into Kant’s objective-subjective distinction through the claim that a subject’s conception of an objective event, i.e.  $[AB]$ , necessitates or presupposes the application of a causal principle to the relevant objects of perception. In the absence of such a principle, we’d lack the ability to Comprehend a determinate, necessary temporal-ordering. The successive perceptions of an objective event are *necessarily connected* according to a rule (B238). For otherwise, “if one were to suppose that nothing preceded an occurrence that it must follow in accordance with a rule, then all sequence of perception would be determined

solely in apprehension, i.e., merely subjectively, but it would not thereby be objectively determined which of the perceptions must be the preceding one and which the succeeding one" (B239).

Again, Kant's argument relies on a crucial objective-subjective distinction, since an irreversible sequence of perceptions would require that one perception succeed another *in the object* of experience and not merely in the subject's apprehension of the manifold of appearances. Conceived in this manner, objectivity is effectively a form of inter-subjectivity: any subject must apprehend such an irreversible sequence of perceptions in a determinate order. The understanding, according to the universal law of cause and effect, imputes a temporal order to phenomena by attributing to each phenomenon a place in a temporal interval in relation to antecedent and subsequent phenomenon. In the Transcendental Deduction Kant established that we must employ concepts of objects in order to have objective experience. Here, in the Second Analogy, Kant affirms that "[we] render [our] subjective synthesis of apprehension objective only by reference to a rule in accordance with which the appearances in their succession, that is, as they happen, are determined by the preceding state" (B240).

### A NON SEQUITUR OF NUMBING GROSSNESS

In the classic, *The Bounds of Sense*, P.F. Strawson famously assessed the merits of Kant's argument: "the order of perceptions is characterized not only as *necessary*, but as a *determined* order, an order in which our apprehension is *bound down*, or which we are *compelled* to observe. These may all perhaps be admitted as legitimate ways of expressing the denial of order-indifference. But from this point the argument proceeds by a *non sequitur* of numbing grossness."<sup>7</sup> As Strawson recognized, [A<sub>R</sub>B<sub>R</sub>]-irreversibility does not imply [AB]-irreversibility, since this would require an A-type state of substance to *necessarily* give way to a B-type state of substance. No such necessity has been established. We cannot infer from the irreversibility of perceptions of the states of a substance, the irreversibility of the objects

themselves. Thus, what Lovejoy similarly deemed to be “one of the most spectacular examples of the *non sequitur*...to be found in the history of philosophy”<sup>8</sup> is as follows:

1.  $[A_R B_R]$ -irreversibility  $\rightarrow [AB]$
2.  $[AB] \rightarrow [AB]$ -irreversibility
3. Therefore,  $[A_R B_R]$ -irreversibility  $\rightarrow [AB]$ -irreversibility

Strawson’s charge denies the validity of (1) and *a fortiori* the validity of (2), which together amount to the implausible claim that  $[A_R B_R]$ -irreversibility  $\rightarrow [AB]$ -irreversibility. The *non sequitur* is rooted in Kant’s failure account for two conditions that must be satisfied if  $[A_R B_R]$ -irreversibility is to imply  $[AB]$ -irreversibility. The first of these must be satisfied in order to know simply whether an objective event has occurred. Recall Kant’s example of the house, where  $A_R$  (the roof) co-exists with  $B_R$  (the doorway). The principle of opposites or contraries, a metaphysical offshoot of the principle of non-contradiction, implies that incompatible conditions cannot co-exist. A static state of substance cannot logically suffer contrary things at the same time in the same part of itself.<sup>9</sup> A house’s roof and doorway certainly are not incompatible states of a substance, and as such, they are co-existent. Alternatively, Kant’s example of a boat being driven downstream satisfies the non-coexistence condition as it cannot be both upstream and downstream (at the same time) in relation to some point along the river. Hereinafter,  $[AB]$  symbolizes an objective event, i.e. an objective succession in the substance itself;  $[A_R B_R]$  symbolizes our subjective representations of the states of substance. This notation is borrowed from Lewis White Beck (see references). Therefore, at the very least, we know that the movement of the boat constitutes an objective event, but this does not tell us whether  $[AB]$  or  $[BA]$  occurs. If we suppose non-coexistence, premise (1) should be reformulated as:

- i.  $[A_R B_R]$ -irreversibility  $\rightarrow [AB]$  or  $[BA]$  (given non-coexistence)

To know that  $[AB]$  and not- $[BA]$  has occurred, we must know that perceptual isomorphism, i.e. “the condition that there be no relevant difference in the modes of causal dependence of  $A_R$  on A and  $B_R$  on B” holds. Perceptual isomorphism requires

that the causal process that connects A with its perceptual effect AR occur prior to the causal process that connects B with its perceptual effect BR. There are a number of ways perceptual isomorphism can fail to hold. "A cunning arrangement of mirrors, designed to reflect some of the light over large distances before it reached my eyes might ensure that I saw later events before the earlier."<sup>10</sup> Or, to give a more concrete example, given that light travels faster than sound, we might see Cornell University's McGraw clock tower strike midnight before we hear its bells chime despite the fact that McGraw strikes one and its bells begin to ring at exactly the same time, i.e. midnight. Nonetheless, if this condition holds, AR will necessarily precede BR, that is, the objective event [AB] will compel us to observe AR & BR in one and only one order, viz., [ARBR]. Therefore:

ii. [AB]→[ARBR]-irreversibility (given perceptual isomorphism)

In light of (i) and (ii), Kant's causal schema, i.e. [AB]-irreversibility, derived in premise (3) is valid if and only if we know that A and B are not co-existent *and* perceptual isomorphism holds. If we know non-coexistence, as we do in Kant's own boat example, the crux of Strawson objection has to do with the invalid move from the plausible objective *temporal* claim that B succeeds A in the object, i.e. [AB], to the objective *casual* schema, i.e. [AB]-irreversibility, which makes the stronger claim that A *never* succeeds B in the object, i.e. never-[BA]. To make the move from (1) to (2), we must know or have sufficient reason to believe that perceptual isomorphism holds, but to know this, we must know or have sufficient reason to believe [AB]-irreversibility. Alas, this is the very causal schema Kant is seeking!

### LEWIS WHITE BECK: SAVIOR OF THE SECOND ANALOGY?

Now, there are some who would like to save Kant from Strawson by claiming that a general causal law is the only thing that could ground objective succession—they try to avoid the *non sequitur*



by avoiding appeals to particular causal laws. They argue that Kant is setting out to establish a universal principle of causality, a principle he treats as distinct from any empirical instantiations that might ground a causal connection between a particular A and a particular B. Although Kant's proof aims to provide the *a priori* basis for the relationship between successive objective states of substance, he says nothing about any particular causal law or what might constitute a proper antecedent condition. As he puts it himself, "there are thus certain laws, in fact *a priori* laws, that first make a nature possible. Empirical laws can obtain, and be discovered, only by means of experience, and indeed in virtue of these original laws through which experience itself first becomes possible" (B263).

Though while a transcendental principle of the understanding is established entirely independent of all experience, particular casual laws are not solely derived through empirical means, since they are "grounded in or made possible by [a] transcendental principle of the understanding."<sup>11</sup> Hence, a defense against the classical charge of *non sequitur* that relies on a strong distinction between the transcendental principle and its empirical instantiations is unavailable. A particular causal law is necessarily subsumed under the transcendental principal, consequently, the separation is not sufficiently strong; the two are not logically exclusive, as they must be, if the proof is to survive Strawson's charge. It therefore seems that we must look elsewhere for an adequate defense of Kant.

If Strawson is correct to interpret Kant to infer as he takes him to infer, then the charge of *non sequitur* is fitting. There's no doubt that simply inferring [AB]-irreversibility from [A<sub>R</sub>B<sub>R</sub>]-irreversibility would qualify as a *non sequitur*, but as Lewis White Beck argues in his defense of Kant's proof, this is not Kant's inference. In a short essay, entitled *A Non Sequitur of Numbing Grossness?*, Beck interprets Kant as follows<sup>12</sup>:

1. [A<sub>R</sub>B<sub>R</sub>]-irreversibility  $\rightarrow$  [AB]-irreversibility if and only if (i) A and B are not coexistent and (ii) perceptual isomorphism does not fail.

2. To *know* [AB], given [A<sub>R</sub>B<sub>R</sub>]-irreversibility, requires:
  - (i) knowledge that A & B are opposite states of a substance; and (ii) knowledge of [AB]-irreversibility to ensure that perceptual isomorphism does not fail.
3. Knowledge of 3(i) is sufficient to know that A & B are not coexistent, i.e. there is an objective event, but knowledge of 3(i) is *not* sufficient to know whether [AB] or [BA] has occurred.
4. In virtue of Hume's conception of causality, I can *know* that [AB] occurs.
5. If I know [AB], then I know not-[BA], which implies knowledge of [AB]-irreversibility.
6. Given knowledge of [AB]-irreversibility, I know perceptual isomorphism does not fail in virtue of 3(ii).
7. [AB]-irreversibility is the schema of causation
8. Therefore, to know that [AB] occurs, I must know that A contains the casual condition of B.

Beck's interpretation, in contrast to Strawson's, differs in that premise (2) attributes to Kant's proof recognition of the two conditions outlined above. Beck takes Kant to acknowledge these conditions at B234: "The objective relation of appearances [that is, of A and B] that follow upon one another is not to be determined through mere perception [that is, from the sequential relation of A<sub>R</sub> & B<sub>R</sub>]." <sup>13</sup>

However, one might be skeptical of Beck's defense. Premise (4), in particular, seems problematic, since to know perceptual isomorphism holds, I must know [AB] irreversibility, but to know this, I must know that [AB] and not-[BA] has occurred. But, can I *know* [AB] has occurred from mere experience? It appears, at first glance, that Beck is begging the question, since all that experience can offer is knowledge of [A<sub>R</sub>B<sub>R</sub>]. And this would be a valid objection to Beck as Kant is not entitled to (4) insofar as his argument is a general proof of the universal principle of causality. But as Beck points out, insofar as the Second Analogy is meant to serve as a reply to Hume, Kant is entitled to claim he *knows* [AB] occurred, "for Hume knows [AB] but has skeptical

doubts about [AB]-necessarily.”<sup>14</sup> Meaning, if we’re to grant Kant knowledge of [AB], then perceptual isomorphism holds, thereby yielding sufficient grounds for inferring from [A<sub>R</sub>B<sub>R</sub>]-irreversibility that [AB]-irreversibly occurs. If we treat the Second Analogy as a direct reply to Hume, in which case Kant can make use of Hume’s own assumptions, Beck would have us believe that Strawson’s charge is misplaced.

But what did Hume mean when he said that we can *know* [AB]? In anticipation of Kant’s later analytic/synthetic distinction, Hume maintained a two-pronged conception of reason, where on the one hand, it served to discover the pure relations between ideas, while on the other, it served to discover matters of fact in sensory experience.<sup>15</sup> Hume’s epistemic criteria therefore says that statements of relations of ideas are ‘either intuitively or demonstratively certain’, where by ‘certain’ Hume means that we are justified ‘by the mere operation of thought’ in not questioning a statement’s truth.<sup>16</sup> Statement’s about matters of fact, on the other hand, depend on evidence gained through experience. Accordingly, Hume rejects ‘obscure and uncertain’<sup>17</sup> metaphysical concepts such as power, force, or necessary connection, since they don’t fall under either function of reason:

Let an object be presented to a man of ever so strong natural reason and abilities; if that object be entirely new to him, he will not be able, by the most accurate examination of its sensible qualities, to discover any of its causes or effects.<sup>18</sup>

So, then, how can we *know* or have sufficient reason to believe that [AB] occurs within the Humean architectonic? To believe, says Hume, is simply to judge a proposition to be true. Causal inference is simply the product of belief in cause-effect relations:

Belief is nothing but a more vivid, lively, forcible, firm, steady conception of an object, than what the imagination is ever able to attain. This variety of terms, which may seem so unphilosophical, is in-

tended only to express that act of mind, which renders realities, or what is taken for such, more present to us than fictions.<sup>19</sup>

And so, while our causal inferences are not justified, we nonetheless come to believe them, and some cases, even come to know them. Once Kant *knows* [AB] has occurred and *not* [BA], he can then justifiably know [AB]-irreversibility occurs. By falling back on his grand doctrine of transcendental idealism, it then follows that A must contain the causal conditions of B, for otherwise Kant would argue, there'd be no grounds in the understanding for experiencing the objective event [AB]-irreversibly. To know any such objective event necessitates the application of causal category, since the effect, i.e. B, is joined to the cause, i.e. A, in our understanding by thinking *through* the causal category.

But, one is now stricken by the utter dependence of the second analogy on Kant's underlying doctrine of transcendental idealism. If Kant intends for [AB]-irreversibility to serve as the causal schema, then it can only do so if we presuppose Kant's conception of the world of appearances as given to us in the sensibility and brought under the categories. For otherwise, we'd have yet another whopping *non sequitur*: it doesn't follow from Kant's premises in the second analogy that we *must* apply the pure concepts of the understanding to the manifold of appearances. We needn't, at least not in virtue of Kant's objective/subjective distinction, posit a causal law. We might very well posit some other doctrine to account for irreversible objective succession in objects. We might be sympathizers of the early Cartesian, Nicholas Malebranche, whereby we account for irreversible objective succession by attributing it to God's will. If we obtain ideas of external things by viewing them within God himself, then there's no need (nor are we justified) to treat [AB]-irreversibility as the causal schema. Again, if we abandon the force of transcendental idealism, Kant's reply to Hume fails. To establish the principle of causality, we must read the second analogy through a transcendental idealist's spectacles, since once removed a Malebranchian theory will do just as well. Kant, of

course, wouldn't have it any other way.

### Notes

- <sup>1</sup> Russell, 1.
- <sup>2</sup> Hume, 136.
- <sup>3</sup> Hume, 141.
- <sup>4</sup> Ewing, 75-76.
- <sup>5</sup> Strawson, 133.
- <sup>6</sup> Beck, "A Non Sequitur of Numbing Grossness?"
- <sup>7</sup> Strawson, 137.
- <sup>8</sup> Loveoy, 303.
- <sup>9</sup> See Plato, Republic Book IV
- <sup>10</sup> Wilkerson, 80.
- <sup>11</sup> Freidman, 172.
- <sup>12</sup> My reconstruction here discards Beck's first premise and fills in premise 5
- <sup>13</sup> Beck, "A Non Sequitur of Numbing Grossness?" quoted with Beck's addendums, 149.
- <sup>14</sup> Beck, "A Non Sequitur of Numbing Grossness?" , 150.
- <sup>15</sup> Hume, 14.
- <sup>16</sup> Hume, 24.
- <sup>17</sup> Hume, 135.
- <sup>18</sup> Hume, 4.6.
- <sup>19</sup> Hume, 5.12.

### Bibliography

- Beck, Lewis White. "A Non Sequitur of Numbing Grossness?" *Kant Studien* 67 (1976); reprinted in *Essays on Kant and Hume* (New Haven: Yale University Press, 1978)
- Beck, Lewis White. "Six Short Pieces on the Second Analogy of Experience," in *Essays on Kant and Hume* (New Haven: Yale University Press, 1978)
- Clatterbaugh, Kenneth. *The Causation Debate in Modern Philosophy* (New York: Routledge, 1999)
- Ewing, A.C. *Kant's Treatment of Causality* (London: 1924)

- Freidman, Michael. "Causal Laws and the Foundations of Natural Science" in *The Cambridge Companion to Kant*, ed. Paul Guyer (Cambridge University Press, 1992)
- Hume, David. "The Idea of Necessary Connexion", *An Enquiry Concerning Human Understanding*. (Oxford University Press; New Ed edition. 1999)
- Loveoy, On *Kant's Proof of the Causal Principle*, Archiv fur-Geschichte der Philosophie (1906), p.303
- Russell, Bertrand. "On the Notion of Cause", Proceedings of the Aristotelian Society, 13 1913), pp. 1-26.
- Strawson, P.F. *The Bounds of Sense* (London, 1966)
- Van Cleve, James. "Another Volley at Kant's Reply to Hume," in *Kant on Causality, Freedom, and Objectivity*, ed. William Harper and Ralf Meerbote (U. of Minnesota P., 1984)
- Wilkerson, T.E. *Kant's Critique of Pure Reason*, (London: Oxford University Press, 1976)

### Translation

- *The Cambridge Edition of the Works of Immanuel Kant* translated and edited by Paul Guyer and Allen W. Wood. (Cambridge University Press; New Ed edition, 1995)