HUME’S THEORY OF CAUSATION:

INFERENCE, JUDGMENT, AND THE CAUSAL SENSE

DON GARRETT

Book 1 Part 3 of A Treatise of Human Nature, entitled “Of Knowledge and Probability,” is the longest of the ten Parts that constitute the Treatise, and Hume devotes nearly all of it to a sustained and multi-faceted endeavor to “explain fully” the relation of causation (T 1.3.2.3, SBN 74). The discussion of the causal relation that he provides there—together with the corresponding “recasting” of it in Sections 4–7 of his later An Enquiry concerning Human Understanding—constitutes one of his best-known and most important contributions to philosophy. Yet the wealth of arguments and claims it contains have led to competing interpretations of his theory of the nature of causation itself—particularly in recent years.¹ He holds that a “necessary connexion” is required as an “essential… part” of the relation of cause and effect (T 1.3.6.3, SBN 87) and goes on to argue that what we take to be such a “necessity and power [lying] in the objects” is in fact merely an internal feeling of “the determination of the mind, to pass from the idea of an object to that of its usual attendant,” a feeling that the mind erroneously treats as a quality of the objects observed (T 1.3.14.25, SBN 167), even though the ideas derived from this feeling of determination “represent not any thing, that does or can belong to the objects” (T 1.3.14.19, SBN 164). Thus it seems that he denies that there are any such things as real causal relations in nature, admitting instead only fictitious projections of internal
sentiments onto objects that cannot genuinely be qualified by them. At the same time, however, he also offers what he calls a “precise definition” of cause as “[a]n object precedent and contiguous to another, and where all objects resembling the former are plac’d in like relations of precedency and contiguity to those objects, that resemble the latter” (T 1.3.14.30–31, SBN 169–70); and he goes on to cite and to employ this definition in the remainder of the Treatise. Thus it seems that by a method of semantic analysis, he reduces the relation of cause and effect to nothing more than what he calls “constant conjunction” between pairs of event-types. Yet he goes on to allow that this same definition may “be esteem’d defective, because drawn from objects foreign to the cause” (T 1.3.14.31, SBN 170); and he alludes variously to “the power, by which one object produces another” (T 1.3.1.1, SBN 69), the “internal structure or operating principle of objects” (T 1.3.14.29, SBN 169), and “the ultimate connexion of... objects,” concluding that “we can never penetrate so far into the essence and construction of bodies, as to perceive the principle, on which their mutual influence depends” (T 2.3.1.4, SBN 400) and that “[w]e cannot penetrate into the reason of the conjunction” (T 1.3.6.15, SBN 93). Thus it seems that he affirms that there are real (even if epistemically inaccessible) causal powers and relations that go beyond both the projection of internal sentiments and mere constant conjunction. Each of these seemingly incompatible elements of the Treatise is repeated in An Enquiry concerning Human Understanding—often with only minor variation, but sometimes (as in the case of references to “powers” and “ultimate principles”) with even greater force or frequency. Accordingly, Hume has seemed to some readers to hold a projectivist theory of causation, to some a reductionist
theory, and to some a *realist* theory; and many have concluded that he either equivocates among them or simply holds an inconsistent mixture of all three.²

Although the outcome of his discussion may seem equivocal, Hume makes it clear why he devotes such a preponderance of Book I Part 3 to an explanation of the causal relation. All reasoning, Hume claims, is a “discovery of… relations” (T 1.3.2.2, SBN 73). “Demonstrative” reasoning, as he characterizes it, is based on unchanging intrinsic “relations of ideas” and so can provide certain knowledge; but its employment, at least beyond obvious truths, is largely limited to mathematics. All other reasoning, according to his Lockean scheme of classification, qualifies as “probable” reasoning, the kind that predominates both in ordinary life and in philosophical enquiry; but all probable reasoning, he argues, “discovers” and depends on the relation of cause and effect, proceeding in one way or another from some immediately perceived or remembered “objects” (in a sense of the term broad enough to include states and events) to other “objects” treated as their causes or effects.

Having explained its importance, Hume proceeds to investigate this crucial relation by “examination of the idea” of it; and in order to do so as “perfectly” as possible, he invokes his prime methodological principle—grounded in the doctrine that ideas (or their simpler component ideas) are always copied from resembling impressions—that one should seek to clarify the content of an idea by examining the impression or impressions from which it is derived (T 1.3.2.4, SBN 74–75). He quickly determines that causes precede and are, at least typically, contiguous with their effects. However, he then finds himself stymied by the nature of a third element or aspect that is of “much greater importance”: the “necessary connexion” of cause and effect (T 1.3.2.11,
SBN 77). Since the impression from which *this* idea is copied is not obvious in the observation of either the cause or the effect, he proposes to approach his quarry obliquely, by examining two questions—which he calls “neighbouring fields”—about the necessity of causes (T 1.3.2.13, SBN 78):

*First*, For what reason we pronounce it *necessary*, that every thing whose existence has a beginning, shou’d also have a cause?

*Secondly*, Why we conclude, that such particular causes must *necessarily* have such particular effects; and what is the nature of that *inference* we draw from the one to the other, and of the belief we repose in it?

(T 1.3.2.14–15, SBN 78)

The investigation of the first question initially yields only a negative answer: the belief is not the result of either demonstration or immediate intuition (T 1.3.3). Hume then moves on to the second question, hoping that its investigation will also yield the positive answer to the first. Most of *Treatise* Book 1 Part 3 is thus an investigation of the nature of causal inferences—i.e., probable reasonings, which are based on the relation of cause and effect—that is intended to lead to the discovery of the impression of necessary connection and, with that discovery, to an understanding of the causal relation itself.

In what follows, I try to explain Hume’s own explanation of causation, and of our “discovery” of it, in such a way as to resolve the appearance of equivocation among incompatible projectivist, reductionist, and realist theories of causation. I begin by explaining his theory of the nature of *causal reasoning*—that is, of what occurs when one reaches a conclusion about a matter of fact through probable reasoning, the kind of
reasoning that “discovers” and depends on the relation of cause and effect. Next, I draw on that explanation, together with other aspects of his cognitive psychology (including his theories of relations, abstract ideas, and the correction and location of sensed qualities) to explain his theory of causal judgment—that is, of what occurs when one judges explicitly that two things stand in the relation of cause and effect. In doing so, I argue that Hume in effect recognizes a “causal sense,” analogous in some ways to the color sense (and other senses of secondary qualities), the moral sense, and the sense of beauty. Finally, I draw on that theory of causal judgment to argue that Hume has a coherent theory of causation that can concede something of importance to projectivism, to reductionism, and to realism, without collapsing into a simple version of any of them.

**1. Hume’s Theory of Causal Reasoning**

Hume identifies three elements or “component parts” of causal reasoning: first, a present “impression” (i.e., a sensation or internal feeling) or memory; second, an inference from the present impression or memory to a believed idea; and third, the believed idea itself (T 1.3.4). The reasoning is “causal” because it effectively treats the object of the impression or memory and the object of the believed idea either as cause and effect or as effect and cause.

**Three Elements of Causal Inference**

The starting point for belief-generating causal reasoning, Hume argues, must have two characteristics. First, it must have the high degree of “vivacity” or “liveliness” that he identifies as characteristic of an impression or memory—otherwise, there may still occur
an associative transition and even “hypothetical reasoning,” but no genuine belief will
result (T 1.3.5). Second, objects like that represented by the impression or memory, on
the one hand, and objects like that represented by the resulting believed idea, on the other,
must have been constantly conjoined, in a consistent order of temporal priority, in the
reasoner’s experience. For the representation of a cause and the representation of its
effect are always entirely distinct “perceptions”—‘perception’ being Hume’s most
general term for contents of the mind, including impressions, memories, and other
ideas—and nothing can be discovered in either perception that would indicate what (if
anything) must follow or be followed by its object. This is confirmed by the mind’s
evident inability to make any causal inferences about objects that are entirely unlike those
that the mind has perceived before. Since a causal inference cannot occur prior to
experience of a constant conjunction between two types of objects or events, the question
then naturally arises of how this inference is made even after experience of constant
conjunction. The negative first stage of Hume’s answer to this question is the famous, if
often misunderstood, conclusion that the inference is not “determin’d by reason”
(T 1.3.6.4, SBN 88)—a conclusion often described as Hume’s “inductive scepticism.”

It is important to understand clearly the question that Hume is asking, to which
“not reason” constitutes a negative part of his answer. Reason, for Hume, is the faculty of
reasoning or inference, the exercises of which are demonstrative and probable reasonings
or inferences. Thus, Hume is not asking whether probable or causal inferences are
themselves examples or instances of reasoning, exercises of the faculty of reason; they
obviously are, and he continues to characterize them as reasonings throughout his
writings, including his argument denying their “determination” by reason. Nor is he
asking, at this point, either whether one is justified in making such inferences or whether
their conclusions will be true. Rather, he is asking whether, within an episode of causal
reasoning, the transition that occurs from the impression or memory to the believed idea
following exposure to a constant conjunction of objects of the two kinds in question is
itself mediated by an inference or piece of reasoning. The outline of his argument is well
known. In order for the mind to move, after some prior observation of a constant
conjunction of two kinds of objects, from an impression or memory of one to a believed
idea of the other, the mind must, in some way, make a “presumption” to the effect that
“instances, of which we have had no experience, must resemble those, of which we have
had experience, and that the course of nature continues always uniformly the same”
(T 1.3.6.4, SBN 89), at least with respect to the conjunction in question, in application to
the present case. If that presumption of uniformity were itself produced by a piece of
reasoning, the presumption would take the form of a conclusion affirming such
uniformity that was produced by either demonstrative or probable reasoning. Such an
affirmed conclusion could not be produced by demonstrative reasoning, however, for
there is nothing in the intrinsic relations among the ideas that requires nature to remain
uniform in an as-yet-unobserved case—as we can see simply from the conceivability of a
change in the course of nature. Yet the presumption of the uniformity of nature could not
be the result of probable reasoning either; for all probable reasoning, Hume has argued,
depends without exception on the relation of cause and effect precisely by effectively
presuming that nature will be uniform in respect of the relevant constant conjunction. The
presumption of uniformity in question cannot be caused by probable reasoning because
the presumption must already have occurred in order to make possible any probable
reasoning about the subject. Hence, Hume concludes, the presumption of the uniformity of nature that facilitates the crucial step within causal reasoning is not itself caused by any reasoning at all. Instead it must be the result of some associative mechanism.

Having reached this largely negative conclusion about the cause of the inference, Hume examines the third element of causal reasoning: the idea believed. He concludes that the belief consists in the vivacity which the idea acquires from the mind’s enlivenment by the initial impression or memory with which the idea has become associated (as a result of the experienced constant conjunction). In turn, this conclusion allows him to offer a more specific positive answer to the question of how the inference occurs by identifying the mechanism behind the implicit presumption of the uniformity of nature: it is not reasoning but “custom” or “habit,” the mechanism by means of which something that has been often repeated in the mind is renewed again “without any new reasoning or conclusion” (T 1.3.8.10, SBN 102).

Further Influences on Belief

Hume next offers evidence to confirm his theory that the belief resulting from causal inference consists of an idea that has been enlivened by its association with a present impression or memory. He does so by noting that the three relations already identified in Treatise 1.1 as associative relations—resemblance, contiguity, and causation itself (whatever exactly it may turn out to be)—all have some enlivening influence on ideas when the objects of those ideas are taken to stand in those relations to the objects of current impressions or memories, even in the absence of the circumstances required for inference (T 1.3.8), and by noting that the associative relations of resemblance and
contiguity can further strengthen beliefs resulting from causal inference when the circumstances for inference are present (T 1.3.9). After a consideration of the influence of beliefs on action (T 1.3.10), he goes on to analyze—in sections devoted to species of “philosophical” and “unphilosophical” probability—various influences on the strength or weakness of belief in cases in which the transition to a particular believed idea does not convey the fullest degree of vivacity (T 1.3.11–13).

The Impression of Necessary Connection

On the basis of this account of the components of causal inference and the variety of influences on belief, and now armed with an understanding of the association-generating role played in causal inference by experience of constant conjunction, Hume returns to the question of the origin of the idea of the necessary connection that he has declared to be such an important aspect of causation. Because experienced constant conjunction is required before one can make a causal inference, one can never perceive a necessary connection after a first experience of one kind of object being followed by another; instead, the impression of necessary connection arises only after experience of their constant conjunction. Yet the mere repeated perception of the same types of objects cannot reveal anything new in the objects themselves. Hence the impression of necessary connection cannot be the sensory perception of something external to the mind and located in the cause and effect; instead it must be something new that arises in the mind itself as the result of the repeated conjunction. Hume concludes that the impression of necessary connection is an “internal impression” of the mind caused by other perceptions—an “impression of reflection” rather than an “impression of sensation,” in
his terminology—and is identical with the felt “determination of the thought to pass from causes to effects and from effects to causes, according to their experienc’d union” (T 1.3.14.22, SBN 166; cf. T 1.3.14.1, SBN 155).

Of course, in calling the impression of necessary connection a felt “determination” of the mind, Hume cannot mean that the mere occurrence of this impression in the mind constitutes immediate awareness, independent of experience, of a causal relation of mental determination holding between the impression or memory and the believed idea that results from the customary transition of thought that he has described. Since all discovery of causal relations depends on experience of constant conjunction, according to Hume, the discovery of the causal roles of one’s own perceptions does as well. The impression of necessary connection is nevertheless properly characterized as an impression “of” the mind’s “determination,” for one in fact feels this impression whenever the mind makes, or is about to make, a custom- or habit-based inference. Indeed, given its own constant conjunction with the occurrence of such inferences, the impression of necessary connection may well be a state of mind that itself constitutes an essential contributing cause to the completion of the inference from the impression or memory to the believed idea; and it may therefore even be characterized as “the determination of the mind” itself, and not merely as an impression “of” the presence of such a determination.

Hume concludes that the “necessity” of causes is constituted by this determination, or “propensity,” of the mind (T 1.3.14.22, SBN 165). In doing so, he strongly implies that the impression of determination that he has identified as that of “necessary connexion” is indeed properly characterized as an impression of necessity.
This implication is legitimate, because he holds that, in general, the application of the term ‘necessity’ reflects the mind’s determination to conceive things in a certain way—i.e., its inability to conceive and affirm otherwise—whether the necessity in question is that of “relations of ideas” (which are always either self-evident or demonstrable) or that of causes. Thus he writes:

As the necessity, which makes two times two equal to four, or three angles of a triangle equal to two right ones, lies only in the act of the understanding, by which we consider and compare these ideas; in like manner the necessity or power, which unites causes and effects, lies in the determination of the mind to pass from the one to the other.

(T 1.3.14.23, SBN 166)

In the latter case, “[t]he objects seem so inseparable, that we interpose not a moment’s delay in passing from the one to the other” (T 1.3.8.13, SBN 104).

Yet although necessary “relations of ideas” and necessary causal relations both involve an inability to think otherwise, there remains a crucial difference between the two cases. In the case of relations of ideas, the inability to conceive otherwise is grounded in the intrinsic character of the ideas themselves, which (at least when they are “adequate representations” [cf. T 1.2.2.1, SBN 29]) represent the intrinsic characters of their objects; whereas in the case of causal necessity, the inability is more properly a psychological difficulty in separating two perceptions (and a psychological inability to believe their objects to be separated), a difficulty resulting not from the intrinsic
characters of the objects as we conceive them but from the habitual association between them established by constant conjunction. Thus Hume writes:

’Tis natural for men, in their common and careless way of thinking, to imagine they perceive a connexion betwixt such objects as they have constantly found united together; and because custom has render’d it difficult to separate the ideas, they are apt to fancy such a separation to be in itself impossible and absurd.

(T 1.4.3.9, SBN 223)

Nevertheless, we may say that much as one “discovers” the necessity of twice two making four by being determined by intrinsic content to conceive of four in the act of conceiving twice two and by being unable to conceive twice two making any other quantity, so one “discovers” the necessity of causes by being psychologically determined (by the observation of genuine constant conjunction) to infer the existence of one object from that of another and by finding it difficult even to think of the one object without thinking of the other. It is precisely in making such inferences that one “discovers” the necessity of causes and, in effect, represents the two objects involved as being related by cause and effect.

2. HUME’S THEORY OF CAUSAL JUDGMENT

On Hume’s account, causal reasoning consists in an inference from an impression or memory to a lively idea— an inference that occurs after a previously experienced constant conjunction of objects like the former with objects like the latter and is accompanied by an impression of necessary connection that is a felt determination of the mind to make the
inference. Such reasoning, as he describes it, presupposes no uniquely human capacities.

Indeed, he argues in the final section of *Treatise* Book 1 Part 3 (T 1.3.16, “Of the reason of animals”) that this fact constitutes a very important point in favor of his account, since animals as well as humans show by their behavior that they perform causal reasoning. But while animals, like humans, effectively presume the uniformity of nature in the course of their causal reasoning, they do not formulate the principle of the uniformity of nature as an explicit doctrine to which they give assent, as at least some reflective humans do. Similarly, although animal reasoning implicitly discovers and depends on causal relations in order to infer one thing from another, animals do not make explicit judgments of the form: “Object A and Object B stand in the relation of cause and effect.” Human beings, however, do make explicit causal judgments of this form; and in order to understand Hume’s theory of what they are judging when they do so, we must understand his theory of how such causal judgments occur. In order to do this, in turn, we must understand something of his general theory of judgments of relations, of which his theory of causal judgment is an application.

*Relations and Abstract Ideas*

For Hume, perceptions may have, and may represent their objects as having, many different qualities; for example, they may represent their objects as being round and a particular shade of red. Similarly, perceptions may stand in, and may represent their objects as standing in, many different relations, either “natural” or “philosophical”; for example, a pair of perceptions may represent one object as similar in shape but double in size to another while also representing the objects as existing simultaneously but at a
distance equal to the height of the smaller. In order for perceptions to represent objects as having particular qualities and relations—and also in order for the perceptions themselves to have qualities and relations that determine the nature of further cognitive operations—it is not always necessary for the mind to have what Hume calls “general or abstract ideas” (what we would call “concepts”) of the qualities or relations in question. In order to think with generality about qualities and relations, however, or about the classes of things that have them it is necessary to form abstract ideas of them.

No idea, Hume argues, is general or indeterminate in its own nature. Therefore, in order to have thoughts with generality, the mind must employ what he calls the “imperfect” device of abstract ideas (T 1.1.7.2, SBN 17). Such ideas arise in the mind, on his view, in the following way. When perceptions resemble one another in some respect, it is natural to apply the same term to each, notwithstanding their difference. Later uses of the term come to elicit a particular and determinate idea (which we may call the “exemplar”) together with a disposition to revive and “survey,” as needed for reasoning or other purposes, the other ideas (which we may call the “revival set”) whose objects resemble one another in the operative respect (T 1.1.7.7–8, SBN 20–21). Thus the abstract idea of the quality red is a determinate idea of a particular thing having a particular shade of red but associated with the word ‘red’ in such a way that the mind is disposed to revive and survey any of a set of other ideas of red things for use as needed. If, for example, one’s abstract idea exemplar of red is an idea of a dark red circle, and the claim is made that all red things are dark red circles, the ideas of red things of other shades or other shapes will immediately come to mind, allowing one to reject the claim proposed. Similarly, we may infer, the abstract idea of a relation consists, for Hume, of a
determinate idea of a pair (or triple, etc.) of particular things taken to stand in that relation and associated with a word in such a way that the mind is disposed to revive and survey for use as needed any of a set of ideas of other pairs whose objects are taken to be similarly related. His application of this theory to the relation of causation is confirmed by his remark that “[w]e must not here be content with saying, that the idea of cause and effect arises from objects constantly united; but must affirm, that ’tis the very same with the idea of these objects” (T 2.3.1.16, SBN 405; emphasis added).

Judgment and Correction

To judge that a particular object has a specified quality, then, is to include a lively idea of that object—an idea constituting belief in its existence—within the revival set of the abstract idea of that quality; and to judge explicitly that two particular objects stand in a specified relation is to include a lively idea of that pair of objects in the revival set of the abstract idea of that relation. To judge that two objects A and B stand in the relation of cause and effect, therefore, is to include a lively idea of A and B in the revival set of the abstract idea of the causal relation—a revival set that has resulted from the effect on the mind of the resemblance that holds among the various object-pairs that, following experience of constant conjunction of like pairs, sustain causal inference with its characteristic impression of determination or necessary connection. Hume aims to capture this revival set with “two definitions” of ‘cause’ that—because “the nature of the relation depends so much on that of the inference” (T 1.3.14.30, SBN 169)—he offers only after his account of the impression of necessary connection:
We may define a *cause* to be “An object precedent and contiguous to another, and where all the objects resembling the former are plac’d in like relations of precendency and contiguity to those objects, that resemble the latter.”

“A cause is an object precedent and contiguous to another, and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other.”

(T 1.3.14.31, SBN 170)

These two definitions express what we might call “productive” and “responsive” conditions—constant conjunction and association-plus-inference, respectively—that lead to the impression of necessary connection and the inclusion of an object-pair in the revival set of the abstract idea of cause and effect.

In Book 3 of the *Treatise*, Hume discusses distinctive sentiments of moral approbation and disapprobation. These sentiments are, like the impression of necessary connection, impressions of reflection; and he characterizes our ability to feel them as a “moral sense” that enables us to discriminate vice and virtue (T 3.1.2). In the *Treatise* and even more clearly in *An Enquiry concerning the Principles of Morals*, he goes on to offer two definitions of ‘virtue,’ one of which appeals to features *external* to the moral spectator (“the possession of mental qualities, useful or agreeable to the person himself or to others” [EPM 9.1, SBN 268]) and one of which appeals to features *internal* to the moral spectator (“whatever mental action or quality gives to a spectator the pleasing sentiment of approbation” [EPM App. 1.10, SBN 289]). This analogy suggests that the
mental operations by which constant conjunction leads to association, inference, and the
impression of necessary connection may also be viewed as a kind of “causal sense,” one
that allows the mind to distinguish those pairs of objects that are causally related from
those that are not.

Of course, senses need not be, and generally are not, infallible; not all things that
appear to resemble each other in a given respect do resemble each other in that respect.
Hence, not every object that we might initially include in the revival set of an abstract
idea is properly so classified. The sizes, shapes, and relative positions of bodies as they
are initially sensed must often be corrected, Hume allows, by consideration of the
position of the observer (T 3.3.3.2, SBN 602–3); and the apparent colors, sounds, tastes,
and smells of objects must also sometimes be corrected, he notes, for features of the
circumstances of observation, including the health of the sense organs. He emphasizes
how the immediate deliverances of the moral sense must likewise be “correct[ed]” by
taking into account differences of perspective on the individuals judged so as to reduce
the “contradictions” in felt response that the same character would otherwise produce
among different observers and even within the same observer at different times (ibid.).
The initial deliverances of the sense of beauty must often be corrected as well, both by a
consideration of the physical circumstances of observation (T 3.3.1.15, SBN 582) and by
reflectively developed rules of criticism (T 2.2.8.18, SBN 379). In parallel fashion, Hume
offers, in the penultimate section of Part 3 of Book 1 of the Treatise, a set of eight “rules
by which to judge of causes and effects” (T 1.3.15), rules that serve to guide the
refinement of the revival set of one’s abstract idea of the relation of cause and effect.
These rules “are form’d on the nature of our understanding, and on our experience of its
operations in the judgments we form concerning objects”—i.e., by reflection on the mechanism of causal reasoning in light of the past successes and failures of causal inferences of various kinds—and by means of them “we learn to distinguish the accidental circumstances from the efficacious causes” (T 1.3.13.11, SBN 149). Of course, the greatest problem of perspective or situation in discerning causal relations lies in the limitation of our experience to only a small part of what actually occurs in the world, with the resulting danger of insufficient or unrepresentative samples. Enquiry into causes involves the development and use of both experiments and rules for judging causes that mitigate this insufficiency as much as possible. Just as the correct or true revival set for the abstract idea of a sensible, moral, or aesthetic quality is that which would arise in an idealized human observer judging in accordance with proper rules of correction, so the correct or true revival set for the abstract idea of cause and effect is the set that would arise in an ideal observer having the human causal sense, possessed of enough observations to constitute a sufficient and representative sample for any causal judgment, and employing the proper rules for judging of causes and effects.

**Judgment and Mislocation**

Hume’s distinction between “impressions of sensation” (such as those of color, sound, heat, and hunger) and “impressions of reflection” (such as passions, moral and aesthetic sentiments, and the impression of necessary connection) is itself a purely causal one: impressions of sensation are caused directly and externally, without the mediation of other perceptions, while impressions of reflection arise in the mind through the causal operations of other perceptions. Given Hume’s theory about the prerequisites for causal
inference, it follows that this distinction between impressions of sensation and
impressions of reflection can itself be drawn only after repeated experience. Within the
class of impressions of sensation, our drawing of the further distinction between those
impressions that resemble continuing qualities of the objects that cause them (such as
shape and motion) and those that do not (such as the pain produced by a sharp knife)
arises as the result of a complex cognitive process depending on aspects of “constancy”
and “coherence” in the occurrence of impressions of sensation (T 1.4.2.2–19, SBN 187–
95). Accordingly, it is very possible to make mistakes both about whether an impression
is one of sensation or of reflection, and about whether an impression of sensation does or
does not resemble a particular quality in its cause. Furthermore, once an impression is
taken either to be or to represent through resemblance a quality of an external cause, the
association of ideas also leads the mind, erroneously, to attribute literal spatial location in
the objects to such non-spatial sensible qualities as sounds and smells (T 1.4.5.12–13,
SBN 237–38; cf. T 1.4.3.5, SBN 221).

In An Enquiry concerning the Principles of Morals, Hume appears to suggest that
the mind has a tendency to locate both moral and aesthetic sentiments—all of which are
in fact impressions of reflection that do not resemble qualities in objects—in the objects
of evaluation themselves, thereby “gilding or staining… natural objects with the colours,
borrowed from internal sentiment” (EPM App. 1.25, SBN 294). In Treatise 1.4.4, he
argues that (what he calls) “the modern philosophy,” while subject to many objections,
has a “satisfactory” argument for the conclusion that there is nothing resembling the
impressions of so-called “secondary” sensible qualities (such as color, sound, taste, smell,
and heat and cold) in external bodies themselves—contrary to our usual way of
conceiving bodies (T 1.4.3–4, SBN 226–27). Hume claims that a similar process of attribution and location occurs erroneously in the case of necessary connection:

[T]he mind has a great propensity to spread itself on external objects, and to conjoin with them any internal impressions, which they occasion, and which always make their appearance at the same time that these objects discover themselves to the senses. Thus as certain sounds and smells are always found to attend certain visible objects, we naturally imagine a conjunction, even in place, betwixt the objects and qualities, tho’ the qualities be of such a nature as to admit of no such conjunction, and really exist no where…. Mean while ’tis sufficient to observe, that the same propensity is the reason, why we suppose necessity and power to lie in the objects we consider, not in our mind, that considers them; notwithstanding it is not possible for us to form the most distant idea of that quality, when it is not taken for the determination of the mind, to pass from the idea of an object to that of its usual attendant.

As a result of this misattribution and mislocation, our ideas of cause-and-effect pairs themselves may include an idea of necessary connection as a part. Because we thereby treat the necessity of causes as if it were grounded in the intrinsic character of the idea of particular cause-and-effect pairs, rather than in the custom-based psychological difficulty of their separation that is the true basis of causal necessity, we erroneously suppose,
Hume thinks, that we can observe causal necessity or power as a quality or relation in the objects themselves; and, moreover, we fail to distinguish the necessity of causes from the kind of necessity that pertains to demonstrations of “relations of ideas,” a necessity that is grounded in the intrinsic character of ideas.

3. CAUSAL PROJECTIVISM

Does Hume’s explanation of the causal relation, expressed in his theories of causal inference and causal judgment, entail that causality is merely projected and not ultimately real? Four reasons might be given for thinking that Hume must deny the reality of causal relations. These concern the topics of truth, determinacy, mind-independence, and explanatory role, respectively.

Projection and Truth

Hume acknowledges the importance of “necessary connexion” to the causal relation; and yet he identifies the impression of necessary connection with a mere feeling of determination that is often projected, through misattribution and mislocation, onto the pairs of objects taken to be causes and effect—even though this feeling cannot represent “any thing, that does or can belong to the objects” (T 1.3.14.19, SBN 164). Moreover, he defines “truth or falshood” as “an agreement or disagreement either to the real relations of ideas, or to real existence and matter of fact” (T 3.1.1.9, SBN 458; cf. T 2.3.3.5, SBN 415). But if truth requires agreement with reality, and we represent the necessity of supposed cause-and-effect pairs by attributing to them a quality that resembles nothing in them, then it seems to follow that Hume cannot regard ascriptions of causality as literally
true, and must instead regard them either as false, as in an “error theory,” or else as just more or less sophisticated expressions of that feeling of determination or of a readiness to make inferences, as in expressivism. While “quasi-realist” expressivist interpreters such as Simon Blackburn (2000) allow Hume to characterize causal judgments as “true” in virtue of meeting norms governing the expression of inferential attitudes, this is not a sense of ‘true’ that Hume own definition of that term would sanction.

Yet Hume constantly implies that many causal judgments are literally true in his own sense of ‘truth’. He writes, for example, of “prov[ing]” and “know[ing]” the causal “dependence” of ideas on preceding impressions (T 1.1.1.8, SBN 4–5), of “the true and real cause of [an] idea, and of the belief which attends it” (T 1.3.8.8, SBN 102); of the sufficiency of constant conjunction to render something a “real cause” (T 1.3.14.32, SBN 171); of having established “the truth of my hypothesis” about the causes of causal inference itself (T 1.4.1.8, SBN 183–84); and of having “prov’d” the causes of pride and humility (T 2.1.12.1, SBN 324–25). Indeed, Hume declares himself in the Introduction to the Treatise to be pursuing nothing less than the “truth” (T Intro. 3, SBN xiv) about the “principles of human nature” (T Intro. 6, SBN xvi), a project that requires forming a “notion of [the] powers and qualities” of the mind and “explaining all effects from the simplest and fewest causes” (T Intro. 8, SBN xvii). He clearly aims at least to contribute to the development of a true account of the causal relations governing the human mind.

Hume can retain this aim for two reasons. First, a Humean idea need not always precisely resemble its object in every respect in order for it to “agree with” it and so represent it truly. For example, a true idea of New York and Boston as being 190 miles apart need not include ideas that are themselves 190 miles apart; some conventional way
of representing large distances (which may include proportions among ideas, ideas of standard units, and ideas of numbers or numerals⁷) is sufficient. Nor, obviously, need the ideas of New York and Boston resemble the actual cities in every particular. Hume’s definition of truth and falsity as “agreement or disagreement” of ideas echoes Locke’s definition of “truth”; and in the Lockean tradition, the ideas provided by a sense can provide a true representation without perfect resemblance. Thus, Locke holds, for example, that nothing in external bodies resembles our sensory ideas of colors or other secondary qualities; but he insists that the representation of a body using these ideas can nevertheless constitute “real Knowledge” of color or other secondary qualities, for the sensory ideas are reliable indicators of the presence of the particular, though unknown, arrangements of corpuscles that produce them (Locke 4.4.4 [1700] 1975: 563–64). And Hume himself continues to characterize objects as having colors despite his allowing that there is a “satisfactory” argument for the conclusion that no quality in bodies themselves resembles our ideas of colors; indeed, he appeals to the “reality” of colors to explain that of beauty:

If, in the sound state of the organ, there be an entire or a considerable uniformity of sentiment among men, we may thence derive an idea of the perfect beauty; in like manner as the appearance of objects in day-light, to the eye of a man in health, is denominated their true and real colour, even while colour is allowed to be merely a phantasm of the senses.

(E-ST 234)
Second, the representation of causal pairs as having a “necessary connexion” as an intrinsic part—a representation which results from the “spreading” or projection of the impression—is not essential to the abstract idea of cause and effect itself. To overcome the illusion—as Hume implies can be done with care and attention (T 1.3.14.31, 1.4.3.9–10, 1.4.7.5–6, SBN 170, 222–24, 266–67)—is not to change the pairs of objects whose ideas constitute the revival set of the idea of cause and effect, but only to correct the way in which they are represented. To include a lively idea of a pair of objects in that revival set when the pair does indeed “agree” with the abstract idea by resembling its other members in the manner picked out by the causal sense is, prima facie, to believe truly that the objects are related by cause and effect.

Projection and Determinacy

Even if causal judgments are sometimes properly considered true, however, it may still be objected that relations of cause and effect nevertheless cannot be real because the nature of the causal sense guarantees that it will sometimes be indeterminate whether two objects are related as cause and effect, whereas every real relation is bivalent—i.e., such that every pair of objects determinately does or determinately does not stand in it (even if it cannot always be known which is the case). Hume clearly allows that there is not always a determinate answer to every question of aesthetic or moral merit—especially when the questions concern comparative merit or the precise boundaries of aesthetic or moral qualities—for the precise delineation of moral and aesthetic qualities depends on there being a uniquely correct reconciliation among conflicting evaluative sentiments involving psychologically varying observers in a wide variety of circumstances.8
Similarly, the objection may run, the causal sense cannot yield a determinate answer to every causal question because the precise delineation of causal relations depends on there being a single correct reconciliation of potentially divergent impressions of necessary connection and mental associations (which may be of varying strengths) among different observers (who may differ in psychological constitution) through exposure to constant conjunctions (which may be of varying constancy and numbers of instances) involving classes of objects whose members may be of varying degrees of resemblance. Thus, it may seem inevitable that some judgments of the form “A is the cause of B” will be of indeterminate truth-value, undecidable even by an ideally informed and ideally corrected causal sense.

First, however, it may be questioned whether the requirement of complete bivalence in the application of causal terms is too stringent, particularly in the context of Hume’s philosophy. The revival sets of some Humean abstract ideas—such as those of single shades of a color or of mathematical relations like “being one greater than”—may admit of no vagueness or indeterminacy. The vast majority of Humean abstract ideas, however, will admit of indeterminacy—for the signification of an abstract idea depends on the revival set of resembling ideas that it does or would elicit; and where precise degrees of resemblance are not easily measured or discerned, it will be common for some potential members of a revival set to be revivable only occasionally, variably, imperfectly, or hesitatingly, without any completely definite standard for correction. The general terms associated with such abstract ideas will, accordingly, be subject to what Hume calls irresolvable “verbal disputes.” It is not only abstract ideas derived from the moral, aesthetic, and causal senses that are potentially subject to such indeterminacy but
also ideas derived from the senses that pick out such qualities as colors, sounds, and
smells. There are many different kinds of resemblance, both simple and complex, many
of which are subject to variations of degree, some easily measurable and some not, and
there are many different classes of resembling objects (and pairs of objects)
corresponding to them. Hume can invoke his theory of abstract ideas to explain why
many predicates fail to delineate sharply a single such class as opposed to other closely
related ones; and given its pervasiveness, a modest amount of such semantic
indeterminacy may well seem insufficient to reject the reality of the underlying relations
themselves.

Second, however, and perhaps more importantly, while the class of pairs of
objects picked out by the causal sense is \textit{potentially subject} to vagueness and
indeterminacy, it does not follow that it \textit{is} vague or indeterminate. Hume believes, on the
basis of inductive evidence derived from the past inductive successes of natural science,
that the world is deterministic—that is, it is such that a finite set of generalizations is
sufficient to determine without exception all of the qualities and relations of each state of
the universe from those of the preceding state. (Indeed, it is this evidence for determinism
that provides the ultimate positive answer to the question of his first “neighbouring field”
mentioned earlier, concerning the cause of our belief that “a cause is necessary for every
beginning of existence.”) Hume expresses the Newtonian hope, and even the expectation,
that the generalizations in question, while perhaps difficult of discovery, are both
relatively simple and relatively few—providing what we might call a “unified theory of
nature.” In a relatively chaotic and indeterministic world, the set of pairs of objects
picked out by an idealized human causal sense might be highly indeterminate—indeed,
perhaps as indeterminate as the set of objects picked out by an idealized human aesthetic sense in the actual world. The set of objects picked out as “red” by an idealized human color sense—a set whose membership is slightly vague and indeterminate in the actual world—might be much more indeterminate in such a conceivable world as well. But in a deterministic world entirely subject to a simple unified theory of nature, the idealized output of the causal sense may be entirely determinate—ultimately yielding precisely that theory. In such a world, causal relations may be very “real”—at least, as far as bivalence is concerned.

Projection and Mind-Independence

Even if many causal judgments are true and all are determinately true or false, however, it may still be maintained that causal relations are nevertheless too “mind-dependent” to qualify as real. For doesn’t the very existence of causal relations depend on there being a human causal sense? And isn’t the causal relation ultimately just whatever that human causal sense happens to pick out, so that what causation is depends equally on the operations of the mind?

Hume considers explicitly the objection that his theory of causation “reverse[s] the order of nature” by making causes depend on thought, rather than vice versa. In doing so, he imagines an opponent who exclaims,

What! the efficacy of causes lie in the determination of the mind!

As if causes did not operate entirely independent of the mind, and wou’d not continue their operation, even tho’ there was no mind existent to contemplate them, or reason concerning them.
Hume’s response is as follows:

As to what may be said, *that the operations of nature are independent of our thought and reasoning, I allow it*; and accordingly have observ’d, that objects bear to each other the relations of contiguity and succession; that like objects may be observ’d in several instances to have like relations; and that all this is independent of, and antecedent to the operations of the understanding. But if we go any farther, and ascribe a power or necessary connexion to these objects; that is what we can never observe in them, but must draw the idea of it from what we feel internally in contemplating them.

The kinds of resemblance among pairs of objects that the actual human causal sense happens to pick out would continue to hold whether there were human beings or not, and events in nature would continue to exemplify the true generalizations of the unified theory of nature. Furthermore, although the “necessity” of causal relations would no longer be felt, the *basis* of that necessity in constant conjunction would continue to exist, in much the same way that, according to the “modern philosophy,” the *basis* of sensory “redness” would continue to exist in the corpuscular structure of bodies even if the impression of red (which does not resemble its basis in the bodies) were never to occur—and also in much the same way that the necessity of mathematical relations would never be felt in the absence of minds, even though the *basis* of that felt necessity would
continue to exist in the intrinsic natures of related things. For this reason, Hume is willing to say that “the constant conjunction of objects constitutes the very essence of cause and effect” (T 1.4.5.33, SBN 250) and that “‘tis from the constant union the necessity arises” (T 2.3.1.4, SBN 400). The causal relations that are in fact detected by a feeling of necessity—like the basis of that felt necessity in the constant conjunction of objects—are not dependent for their existence on the existence of the human mind or a causal sense that detects them.

Nor need we say that causal relations are whatever the causal sense happens to pick out, if this is taken to mean that a change in what produced association, inference, and the distinctive impression of necessary connection would entail a change in what causation is. Consider an analogy. Suppose that the basis of redness consists in a particular complex (and perhaps highly disjunctive) structural surface reflectance quality. Suppose, too, that, had the human sensory apparatus been different, a different quality of bodies would have produced the impression of red under standard circumstances. Then although it may be true that the quality of redness is what the visual sense of red actually discerns—and that this quality turns out to be a particular complex structural surface reflectance property—it will not follow that redness would have been a different property if the human sensory apparatus had been different. For the referent of ‘quality of red’ may be fixed rigidly by the actual present human sensory apparatus—i.e., in such a way that redness is that quality that produces the impression of red through the actual human sensory apparatus under standard circumstances in the actual world now, not whatever would produce the impression of red in another apparatus or under other circumstances. The quality of redness itself need not vary with the nature of the mind, even if a change in
the nature of the human mind would lead to something other than *redness* playing the role (including producing the qualitatively distinctive “impression of red”) that redness plays for us. Similarly, the referent of ‘causal relation’ may be fixed rigidly by the causal sense—so that the causal relation is, in each possible world, that relation that actually produces association, inference, and the impression of necessary connection in normal human beings in the *actual* world, not whatever relation *would* produce it in that other possible world. So understood, the relation of causation itself need not vary with the nature of the mind, even if a change in the nature of the human mind would lead to a relation other than causation playing something of the role (including the production of association, inference, and an impression of necessary connection) that the relation of causation plays for humans in the actual world.

Alternatively, it may be proposed that causal relations must be mind-dependent, on Hume’s account, because the causal pairs constituting the revival set of the abstract idea of causation will have no distinctive resemblance to one another except their tendency to produce association, inference, and the impression of necessary connection. However, the fifth of his “rules by which to judge of causes and effects” states 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.

(T 1.3.15.7, SBN 174)
For Hume, at least, the ability of all causal pairs to produce the same effect on the mind indicates that there is some mind-independent respect—presumably, the causal relation itself—in which all such pairs do resemble one another.

Projection and Explanatory Role

Blackburn has recently proposed that projectivists can eschew what he calls “[t]heoretical or ontological… [or] upper-case Realism” about a kind of entity by rejecting the claim that “we cannot understand what is going on [in discourse about the entity] except in terms of our responding to a world whose entities and properties and relations are the ones ostensibly referred to” in the discourse (2000: 110). However, it is at least debatable whether any full-fledged explanation of causal discourse could avoid making reference to causal relations as part of the explanation—by invoking, for example, our responding to exposure to constant conjunction and the role of such reactions in producing discourse. Without such references, it might be argued, one would have at most a narrative about, rather than an explanation of, causal discourse.

Thus it may be granted that Hume’s account of the operation of the causal sense involves the projection onto cause and effect pairs of an element of “necessity” that does not resemble anything in those pairs, and which derives its origin instead from the constant conjunction of types of objects. It does not follow from this, however, that Hume cannot regard causal judgments as true and determinate, and causal relations themselves as mind-independent. He may even regard reference to causal relations as essential to any explanation of our discourse about them. He need not, therefore, regard causal judgments as mere projections rather than true descriptions of reality.
Hume follows his account of the impression of necessary connection by declaring that he will “collect all the different parts of this reasoning, and by joining them together form an exact definition of the relation of cause and effect, which makes the subject of the present enquiry” (T 1.3.14.30, SBN 169). He proceeds, as we have seen, to define ‘cause’ in terms of constant conjunction—that is, as “[a]n 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” (T 1.3.14.31, SBN 170). He then cites this definition in seeking to justify a variety of further claims, including four “corollaries” concerning the causal relation (T 1.3.14.32–36, SBN 170–72) and his doctrine that human actions exhibit causal necessity (T 2.3.1). Thus, it seems that Hume offers a semantic analysis of the term ‘cause’ in order to show that there is nothing more to causation than constant conjunction.

Reduction and Semantic Analysis

There are, however, several reasons to think that Hume’s definition is not intended to provide a semantic analysis of the term ‘cause’—at least if this is understood to mean a synonymous term, one that can be substituted for it without change of meaning. First, as we have already seen, he offers not just one but two “definitions” of ‘cause,’ the second of which defines ‘cause’ quite differently, as “an object precedent and contiguous to another, and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other”
These two definitions of ‘cause’ are obviously not synonymous with each other, and hence at most one of them could be intended to be synonymous with ‘cause.’ Yet he treats the two definitions entirely on a par, citing both together, for example, in his defense of his four “corollaries” and in his defense of his doctrine of the necessity of human actions. Thus, it seems evident that he must intend neither of them to be synonymous with ‘cause.’ Indeed, he rejects several other proposed definitions of ‘cause’ and ‘necessary connexion’ precisely on the grounds that they are synonymous with their proposed definienda (T 1.3.2.10, SBN 77).

Second, Hume’s account of the “rules by which to judge of causes and effects” is incompatible with regarding the first definition as a synonym for ‘cause.’ His first three rules are:

1. The cause and effect must be contiguous in space and time.
2. The cause must be prior to the effect.
3. There must be a constant union betwixt the cause and effect.

As already noted, Hume characterizes all eight of the rules as “form’d on the nature of our understanding, and on our experience of its operations in the judgments we form concerning objects” (T 1.3.13.11, SBN 149). In addition, he gives an argument from the nature of time for the second rule (T 1.3.2.7, SBN 75–76), and he relaxes the first rule’s requirement of spatial contiguity when dealing with unlocated entities (referring forward to T 1.4.5: T 1.3.2.6, SBN 75). Yet these first three “rules” constitute the primary elements in the first definition of ‘cause.’ Accordingly, if that definition were intended to be synonymous with ‘cause,’ these three rules would be trivial analytic truths for Hume,
and could not be discoveries grounded on experience—let alone properly subject to either further argumentation or later relaxation.

Third, the first definition is much too schematic to specify effectively the set of cause-and-effect pairs—at least for someone lacking the guidance of a causal sense. For a causal sense is required to discern what kinds of resemblance among objects are most significant—and what kinds of difference are insignificant—for determining that an object “of one kind” has been regularly followed by an object of “another kind.” In fact, the second definition is also fundamentally unusable in the absence of a causal sense, for the question of whether an idea or an impression “determines” the mind to form another idea is itself a causal question, answerable ultimately only with a causal sense. Of course, in ordinary causal judgment, Hume surely thinks, one employs neither definition but simply utilizes (and corrects as appropriate) one’s own causal sense.

If the two definitions are not intended to be synonymous with ‘cause,’ then what is their purpose? Hume aims to explain the meaning of the term ‘cause’ by indicating the set of ideas that constitute the revival set of the abstract idea of cause and effect. Both definitions serve this purpose—one by indicating the initial feature of cause-and-effect pairs that leads them to produce the impression of necessary connection and be discriminated by the causal sense, and the other by indicating the associative-plus-inferential effects on the mind of cause-and-effect pairs that lead to the same result. The term ‘cause’ and its two definitions, taken together, indicate this revival set in three semantically distinct ways; hence, their co-extensivity is an a posteriori truth, potentially requiring considerable investigation on Hume’s part—or ours—to discover. This three-way co-extensivity is not only an a posteriori truth but also presumably a necessary one.
(i.e., a truth in every possible world)—just as the co-extensivity of ‘red’ with a particular
description of a complex structural surface reflectance property (or, analogously, ‘virtue’
with a particular description of features of character) might well be both a posteriori and
necessary.

Reduction and the Limitation of Causation

Because Hume’s first definition is not intended as a synonym for ‘cause,’ there is no
reason to think that he regards causation as “nothing but” constant conjunction. On the
contrary, since his two definitions already specify two different features that all cause and
effect pairs have in common, he has no reason to deny that there may also be other
features that all cause-and-effect pairs have in common as well. Such further additional
commonalities, too, might have something to do with the way in which cause-and-effect
pairs are picked out by the causal sense.

Thus, it may be granted that Hume’s first definition of ‘cause’ is intended to
establish that constant conjunction, of the sort discriminated by the causal sense, is
necessary and sufficient for causation; for this conclusion is required for the applications
he makes of the definition in his four “corollaries” and the argument for the causal
necessity of human actions. His definition establishes this conclusion not through
semantic analysis, however, but through an investigation of the basis of the
discriminations made by the causal sense, analogous to an investigation of the underlying
basis in bodies of the perception of redness; and it constitutes a “reduction” of causation
to constant conjunction only in the sense in which one might aim to “reduce” redness to a
structural surface reflectance property of red bodies. It does not positively exclude the
possibility that cause-and-effect pairs will have other important and relevant
commonalities as well, and hence it does not entail that there is “nothing more” to
causation that just constant conjunction.

5. CAUSAL REALISM

Does Hume hold that causal relations positively do consist of more than a particular kind
of constant conjunction that gives rise to association, inference, and an internal
impression of determination, an impression that is then mislocated in the objects
themselves? His apparent references to “powers” by which objects “produce” one
another, to “ultimate” and “operating principles,” and to an impenetrable “reason of the
conjunction” of causes and effects seem to imply that he does—even if he also thinks that
the nature of the “something more” is unknown and incomprehensible.

Realism and ‘Power’

Hume can allow attributions of “powers” to objects without implying that causation is
more than constant conjunction, for he takes the term ‘power’ to signify “that very
circumstance in the cause, by which it is enabled to produce the effect” (EHU 7.17,
SBN 67–68; cf. T 1.3.14.13, SBN 161). He frequently writes of natural philosophy as
discovering the powers of bodies, and of his own “science of man” as discovering powers
of the mind. This usage is appropriate, on his view, because the discovery of the precise
cause of some effect by means of the causal sense is, on his view, ipso facto the
discovery of the power that produced it. If, for example, the motion of one billiard ball is
the cause of the motion of another, then the motion of the first is itself the “circumstance
in the cause, by which it is enabled to produce to effect." Similarly, the discovery of the effects that a given object or state produces is a discovery of the “powers” of that object or state to produce those effects. Many “powers” of things, in this sense, may be hidden from us in the unobserved microscopic or insensible structure of things in a way that does not require causation to be more than constant conjunction. However, because the mind tends to project the impression of necessary connection onto the objects themselves, it is natural to suppose that the idea of necessary connection represents a further feature intrinsic to the precise or particular cause itself that makes necessary its connection with its effect (i.e., makes the separation of the effect from it unthinkable) and therefore constitutes the perception of an inner “power,” “energy,” or “efficacy” inherent in the particular cause itself. Accordingly, Hume holds, the term ‘power’ is often used with the intention of designating some intrinsic feature of the particular cause itself in which such a necessity of its being followed by its effect consists. Yet because such a further feature is never observed at all, it cannot be observed to be constantly conjoined with causes and so cannot be inferred to exist by means of ordinary probable inference. Nor can we infer that such a feature must exist on the grounds that the regularity of nature would otherwise be improbable, since for Hume the probability that nature will be regular is simply a direct function of the extent to which it has been regular in the past. Even more fundamentally, however, since “the necessity of causes” is in fact a psychological inseparability derived from constant conjunction via habitual association, it seems that no intrinsic causal necessity can be successfully conceived: the conditions that the necessity be both intrinsic and causal...
appear to be incompatible, so that they cannot be conceived to be jointly satisfied. Thus, he writes:

upon the whole we may infer, that when we talk of any being, whether of a superior or inferior nature, as endow’d with a power or force, proportion’d to any effect; when we speak of a necessary connexion betwixt objects, and suppose, that this connexion depends upon an efficacy or energy, with which any of these objects are endow’d; in all these expressions, so apply’d, we have really no distinct meaning, and make use only of common words, without any clear and determinate ideas. But as ’tis more probable, that these expressions do here lose their true meaning by being wrong apply’d, than that they never have any meaning; ’twill be proper to bestow another consideration on this subject, to see if possibly we can discover the nature and origin of those ideas, we annex to them.

(T 1.3.14.14, SBN 162)

Particular allusions on Hume’s part to secret or unknown powers may thus use the term ‘power’ in either of these two ways—that is, as alluding simply to unobserved elements of constant conjunctions or as alluding to something that is both intrinsic to causes and yet causally necessitating—or they may be intentionally ambiguous between them. But even allusions to unknown “powers” and “principles” that were unambiguously intended in the second way would not entail that Hume acknowledges that something does or can meet both of the conditions. For just as Berkeley in the Three
Dialogues first denies that we can know of objects that both exist and are outside the mind, and only later denies that any objects can meet both of those conditions, so Hume may deny that we know of powers that are both intrinsic to causes and causally necessitating, and also deny that anything can meet both of those conditions. Thus, in An Enquiry concerning Human Understanding, Hume alludes several times in his discussion of causal inference to our “ignorance” of the power located in causes themselves, an “ignorance” that is shown, he argues, by our inability to predict effects from causes prior to experience of constant conjunction. But, as Winkler (2000) emphasizes, these seeming references to “unknown powers” in the Enquiry occur prior to his explanation of the impression of necessary connection and its projective mislocation; and it is noteworthy that Hume remarks in a footnote to a typical such reference that: “The word, power, is here used in a loose and popular sense. The more accurate explication of it would give additional evidence to this argument. See Sect. 7 [i.e., “Of the Idea of Necessary Connexion”]” (EHU 4.16n7, SBN 33n). The “accurate explication” of ‘power’ gives even greater force to the argument that we do not make causal inferences through knowledge of powers located in causes themselves by showing that such powers are not merely unknown but inconceivable.

Realism and ‘Ultimate Principles’

Like the term ‘power,’ the term ‘ultimate principle’ may be taken in more than one way for Hume. In the first, an “ultimate principle” is the topic of one of the final causal generalizations that can be reached in a true unified theory of nature—“ultimate” because it serves to explain more specific causal regularities by subsuming them as special cases
while not itself being subsumable by any yet-more-inclusive generalization. Thus he writes of “[e]lasticity, gravity, cohesion of parts, communication of motion by impulse” that “these are probably the ultimate causes and principles which we shall ever discover in nature” (EHU 4.12, SBN 30). Because of the projective mislocation of the impression of necessary connection, however, it is natural, he thinks, to suppose erroneously that we can discover “ultimate principles” involving a necessity of such generalizations’ being true that lies in the nature of the causes and effects themselves. Such “ultimate principles,” one may then suppose, would help to explain regularities in nature by showing their intrinsic necessity. However, such supposed principles are just as inconceivable as powers located in objects:

We wouldn’t willingly stop before we are acquainted with that energy in the cause, by which it operates on its effect; that tie, which connects them together; and that efficacious quality on which the tie depends. This is our aim in all our studies and reflections: And how must we be disappointed, when we learn, that this connexion, tie, or energy lies merely in ourselves, and is nothing but that determination of the mind, which is acquired by custom, and causes us to make a transition from an object to its usual attendant, and from the impression of one to the lively idea of the other? Such a discovery not only cuts off all hope of ever attaining satisfaction, but even prevents our very wishes; since it appears, that when we say we desire to know the ultimate and
operating principle, as something, which resides in the external object, we either contradict ourselves, or talk without a meaning.

(T 1.4.7.5, SBN 266–67)

Hume’s claims that we cannot “penetrate” to ultimate principles parallel his claims in *An Enquiry concerning Human Understanding* about our ignorance of powers in bodies; and for the same reason, they do not commit Hume to the view that there are such principles. His claim that we cannot know of things meeting the characterization at issue are compatible with his later and stronger claims that we cannot even conceive of such things. Indeed, his practice with respect to “powers” and “ultimate principles” may be compared not only with Philonous’s discussion of unperceived existence in the *Three Dialogues* but also with Cleanthes’s discussion of “necessary existence” in Hume’s own *Dialogues Concerning Natural Religion*:

It is pretended that the Deity is a necessarily existent Being; and this necessity of his existence is attempted to be explained by asserting, that, if we knew his whole essence or nature, we should perceive it to be as impossible for him not to exist as for twice two not to be four. But it is evident, that this can never happen, while our faculties remain the same as at present. It will still be possible for us, at any time, to conceive the non-existence of what we formerly conceived to exist; nor can the mind ever lie under a necessity of supposing any object to remain always in being; in the same manner as we lie under a necessity of always conceiving twice two to be four. The words, therefore, *necessary existence,*
have no meaning; or, which is the same thing, none that is consistent.

But farther; why may not the material universe be the necessarily existent Being, according to this pretended explication of necessity? We dare not affirm that we know all the qualities of matter; and for aught we can determine, it may contain some qualities, which, were they known, would make its non-existence appear as great a contradiction as that twice two is five.... It must be some unknown, inconceivable qualities, which can make [God’s] non-existence appear impossible, or his attributes unalterable: And no reason can be assigned, why these qualities may not belong to matter. As they are altogether unknown and inconceivable, they can never be proved incompatible with it.

(DNR 9, 189–90)

Cleanthes’s reference to our ignorance of any necessary existence on the part of the material universe in the second paragraph does not commit him to the claim that there actually is a quality of necessary existence, a claim that he has already rejected as either meaningless or inconsistent in the first paragraph.

Realism and Relative Ideas

Hume does not deny that objects may have unknown qualities that are unobserved and inconceivable by us, but he does question whether we have any basis for calling any of them “powers”:
If we have really no idea of a power or efficacy in any object, or of any real connexion betwixt causes and effects, 'twill be to little purpose to prove, that an efficacy is necessary in all operations. We do not understand our own meaning in talking so, but ignorantly confound ideas, which are entirely distinct from each other. I am, indeed, ready to allow, that there may be several qualities both in material and immaterial objects, with which we are utterly unacquainted; and if we please to call these power or efficacy, 'twill be of little consequence to the world. But when, instead of meaning these unknown qualities, we make the terms of power and efficacy signify something, of which we have a clear idea, and which is incompatible with those objects, to which we apply it, obscurity and error begin then to take place, and we are led astray by a false philosophy. This is the case, when we transfer the determination of the thought to external objects, and suppose any real intelligible connexion betwixt them; that being a quality, which can only belong to the mind that considers them.

Like Locke, Hume recognizes the existence of “relative ideas,” which allow the mind to conceive of something simply as whatever it is that stands in a particular relation to something else. In order to have a relative idea of something, however—for Hume as for Locke—it is necessary to have an idea of one of the relata and an abstract idea of the relation. Thus, any attempt to regard an “unknown quality” as the intrinsic necessitating
power of a cause faces a dilemma. If we merely conceive the unknown quality as accompanying the cause, we have not yet conceived it as an intrinsic necessitating power, but merely as another quality that is constantly conjoined with the cause and effect. If, on the other hand, we try to conceive of it as causally necessitating the effect, it appears that we can conceive of this necessitation only as an associative psychological connection grounded in constant conjunction—for that is what causal necessitation has proven to be—and hence not as a necessity intrinsic to the individual causes and effects themselves.

Yet this dilemma is not quite the end of the matter. To conceive of powers or ultimate principles in causes themselves would require the conception of a kind of necessity other than the causal necessity that lies in an associative psychological determination resulting from habit. It would also require, presumably, the conception of a kind of necessity other than that which characterizes self-evident or demonstrable relations of ideas such as “twice two is four.” For although it might be proposed that truly adequate ideas of causes and effects would show the occurrence of each effect to be a self-evident or demonstrable matter of relations of ideas given only the occurrence of its cause, this proposal would require that our present ideas of things be erroneous even in representing causes and effects as distinct (in Hume’s sense of that term, which requires that distinct things be conceivable as existing without one another). That is, it would require that something like what Hume calls Spinoza’s “hideous hypothesis” of “the simplicity [i.e., lack of distinct parts] of the universe” (T 1.4.5.18–19, SBN 240–41)—a doctrine to which Hume shows no evident attraction. These two kinds of necessity, causal and demonstrative, exhaust the relations of necessity that we can positively conceive; and we cannot, therefore form a suitable relative idea of power by using an idea of the
relation of causal necessity or of the relation of demonstrative necessity. Nevertheless, we may be able to use the idea of a different relation—namely, the relation of resemblance—to form a relative idea of the needed relation. For it seems impossible to rule out positively the following hypothesis: that if we had experience of qualities of objects beyond those available to us now, we would then discern something characterizing cause-and-effect pairs themselves that we would find to resemble the two known kinds of necessity in some way that concerned (something resembling) determination-of-thought or inability-to-conceive-otherwise. Such a relation might even be assimilated to demonstrative necessity, if we form a relative idea of something resembling, but yet fundamentally different from, our present human ideas to stand in such a relation. Indeed, Hume’s qualifying phrase “while our faculties remain the same as at present” in the midst of Cleanthes’s discussion of the meaninglessness of ‘necessary existence’ may be understood as alluding to just such a hypothesis. By means of this hypothesis, then, one might form a doubly-relative idea—namely, the idea of a quality in or relation between objects such that, were we to acquire acquaintance of it through a radical alteration in our faculties, we would then find it somehow to resemble the kinds of necessity with which we are now familiar in such a way that we would see it as constituting a new and now-positively-unconceived kind of necessity between causes and effects.

Thus, Hume can allow that there may be something more to a causal relation than constant conjunction giving rise to association-plus-inference and a projected feeling of determination. Nor need he rule out the possibility of a quality or relation in causally-related pairs that would resemble known kinds of necessity enough that, if we could discern it, we could regard it as some third kind of “necessity,” one inherent in causes
and effects themselves. Hume neither clearly describes nor clearly endorses a belief in such a species of necessity. Still, he might well allow that such a belief would be psychologically attractive to someone convinced by his argument about the projective mislocation of the impression of necessary connection and yet still attracted—by the residual force of that projection—to the doctrine that causes and effects have some kind of intrinsic necessity. Indeed, his many references (particularly in the first Enquiry) to our “ignorance” of secret powers and ultimate principles may well be intended, at least in part, to accommodate such a belief. He would certainly deny, however, that awareness or even postulation of such a species of necessity is required for the discovery of causal relations. Such discovery requires only a causal sense.

6. CONCLUSION

What then, according to Hume, is the causal relation? It is, as he promised at the outset of his explanation, the relation that is “discovered” by the process of causal inference whereby objects of types that are constantly conjoined produce association, inference, and an impression of determination or necessary connexion. This process serves as the basis of a “causal sense.” It is by means of this sense and the judgments based upon it that we discriminate cause-and-effect pairs from other pairs, in something like the way that Hume’s visual sense discriminates colors, his moral sense discriminates vice and virtue, and his aesthetic sense discriminates beauty and deformity.

In the course of explaining this causal sense and the causal judgments to which it gives rise, Hume asserts that it involves the projection onto cause and effect pairs of an element of felt necessity that does not resemble anything in those pairs, deriving its origin
instead entirely from the experienced constant conjunction of types of objects. Yet at the same time, he regards many causal judgments as true and can regard all of them as determinately true or false; in addition, he can regard causal relations themselves as mind-independent and even as explanatory basic. He grants that constant conjunction of the kind ideally detected by the causal sense is sufficient for causation. Yet he denies that ascriptions of causation are synonymous with ascriptions of constant conjunction, and he does not assert that there is nothing more to causation than constant conjunction. He rejects the possibility of straightforwardly conceiving—and hence, too, of straightforwardly believing—that there are inherent necessitating powers or ultimate necessitating principles in causes and effects themselves. Yet he does not positively rule out the possibility that some by-us-inconceivable analogues of such powers or principles may characterize causes and effects. His causal sense theory allows him to do all of these things with complete consistency.

In light of these results, should we classify Hume’s theory of causation as projectivism, reductionism, or realism? If my interpretation of Hume is right, then he concedes something to the motivations of each of these, and he could with some justice be classified as holding any of them—or all of them, or none of them—depending on the details of the more specific definitions that might be proposed for them. This result is, I think, all to the good: any plausible theory of causation should concede something to each of these approaches—as, perhaps, should any theory of the objects of at least some of the other Humean senses.

Hume’s theory of causation is of historical interest not only because it has been an influential theory of causation but also because it sheds light on the many other aspects of
Hume’s philosophy to which it is related. Similarly, Hume’s theory of causation is of philosophical interest not only because it is a deep and original theory of causation that offers to reconcile the insights of causal projectivism, causal reductionism, and causal realism, but also because it offers the prospect of illuminating the dialectic among projectivism, reductionism, and realism on other topics as well.

This essay incorporates material from Garrett 2009 by permission of Oxford University Press.

1 Much of the recent debate about Hume’s theory of causation has been provoked by striking readings of Hume as a causal realist—readings that exemplify what is now (following Winkler 2000) commonly known as “The New Hume” interpretation. For a summary description of the debate and a collection of important contributions to it, see Read and Richman 2000.


3 For a fuller account of this argument, see Garrett 1997: ch. 4 and refinements in Garrett 1998 and 2001.

4 The term ‘relation,’ Hume claims, may be understood in either of two senses. The first, “natural” sense is restricted to “that quality, by which two ideas are connected together in the imagination, and the one naturally introduces the other”; this sense includes the three associative relations of resemblance, contiguity, and causation. The second, “philosophical” sense encompasses more broadly any “particular circumstance, in which, even upon the arbitrary union of two ideas in the fancy, we may think proper to compare them”; and there are seven different species of such relations, including relations of resemblance, space and time, and causation, but also identity, contrariety, degrees in quality, and proportions in quantity or number (T 1.1.5.1–2, SBN 13–14).

5 These rules include such principles as “[t]he cause and effect must be contiguous in space and time”; “[t]he same cause always produces the same effect, and the same effect never arises but from the same cause”; and “where several different objects produce the same effect, it must be by means of some quality, which we discover to be common amongst them” (T 1.3.15.2–10, SBN 173–74).

6 As Hume explains in Treatise 1.4.2, the “vulgar” typically attribute a “continu’d and distinct existence” to impressions of sensations themselves, not distinguishing them from objects; the more “philosophical” regard their impressions of sensation as caused by resembling qualities of objects.

7 Hume discusses the representation of large numbers in T 1.1.7.12, SBN 23.
Commentators have generally judged that the two definitions are not even co-extensive. However, I have argued (Garrett 1997: ch. 5) that the two definitions of ‘cause’ are each systematically ambiguous between a “subject-relative” reading and an “absolute” reading. On the subject-relative reading, the first definition concerns what has been observed to be constantly conjoined in a given subject’s experience, while the second definition concerns the subjects of association and inference in that subject’s mind. On the absolute reading, the first definition concerns what has been, is, and will be constantly conjoined through all time, while the second definition concerns the subject of association and inference in the mind of an idealized subject. So understood, the definitions are co-extensive on their subject-relative readings and are again co-extensive on their absolute readings; whereas the first reading specifies what a subject will include in his or her own revival set for the abstract idea of cause (and hence reflects what he or she takes to be causally related), the second reading specifies what a fully corrected revival set would include (and hence what is in fact causally related).

Since his definitions of ‘cause’ entail that a cause is always followed by its effect, one of the corollaries that he draws from them is that “[t]he distinction, which we often make betwixt power and the exercise of it, is… without foundation” (T 1.3.14.34, SBN 171). Nonetheless, he later allows, in treating of the passions, a less “philosophical” conception of power, in which a part of a complete cause may also be considered a “power” when it seems possible or probable that the effect will occur—especially if volition is thought to be the needed completion of the cause (T 2.1.10.4–12, SBN 311–16).

This point is made in Winkler 2000.

Presumably, when the same impression is spread onto the cause and effect together, rather than specifically onto the cause, it is taken to be an impression of a necessary connection or tie between the cause and effect, rather than a power in the cause alone.

This line of argument is suggested in Strawson 1989.