Causality*

Jessica Wilson†

October 2, 2005

Introduction

Arguably no concept is more fundamental to science than that of causality, for investigations into cases of existence, persistence, and change in the natural world are largely investigations into the causes of these phenomena. Yet the metaphysics and epistemology of causality remain unclear. For example, the ontological categories of the causal relata have been taken to be objects (Hume 1739), events (Davidson 1967), properties (Armstrong 1978), processes (Salmon 1984), variables (Hitchcock 1993), and facts (Mellor 1995). (For convenience, causes and effects will usually be understood as events in what follows.) Complicating matters, causal relations may be singular (Socrates’s drinking hemlock caused Socrates’s death) or general (Drinking hemlock causes death); hence the relata might be tokens (e.g., instances of properties) or types (e.g., types of events) of the category in question. Other questions up for grabs are: Are singular causes metaphysically and/or epistemologically prior to general causes or vice versa (or neither)? What grounds the intuitive asymmetry of the causal relation? Are macro-causal relations reducible to micro-causal relations? And perhaps most importantly: Are causal facts (e.g., the holding of causal relations) reducible to non-causal facts (e.g., the holding of certain spatiotemporal relations)?

---

*Forthcoming in The Philosophy of Science: An Encyclopedia
†Department of Philosophy, University of Toronto; jessica.m.wilson@utoronto.ca
Some Issues in Philosophy of Causality

The Varieties of Causation

Causes can apparently contribute to effects in a variety of ways: by being background or standing conditions, “triggering events”, omissions, factors which enhance or inhibit effects, factors that remove a common preventative of an effect, etc. Traditionally accounts of causation have focussed on triggering events, but contemporary accounts are increasingly expected to address a greater range of this diversity.

There may also be different notions of cause characteristic of the domains of different sciences (see Humphreys 1986 and Suppes 1986): the seemingly indeterministic phenomena of quantum physics may require treatment different from either the seemingly deterministic processes of certain natural sciences or the “quasi-deterministic” processes characteristic of the social sciences, which are often presumed to be objectively deterministic but subjectively uncertain. Also relevant here is the distinction between teleological (intentional, goal-oriented) and non-teleological causality: while the broadly physical sciences tend not to cite motives and purposes, the plant, animal, human and social sciences often explicitly do so. Contemporary treatments of teleological causality generally aim to avoid positing anything like entelechies or “vital forces” (of the sort associated with 19th-century accounts of biology), and also to avoid taking teleological goals to be causes that occur after their effects (see Salmon 1989, §3.8 for discussion). On Wright’s (1976) “consequence-etiology” account, teleological behaviors (e.g., stalking a prey) are not caused by future catchings (which, after all, might not occur), but rather by the fact that the behavior in question has been often enough successful in the past that it has been evolutionarily selected for; and for creatures capable of intentional representation, alternative explanations may be available. While teleological causes raise interesting questions for the causal underpinnings of behavior (especially concerning whether a naturalistically acceptable account of intentionality can be given), the focus in what follows will be on non-teleological causality, reflecting the primary concern of most contemporary philosophers of causation.
Singular vs. General Causation

Is all singular causation ultimately general? Different answers reflect different understandings of the notion of “production” at issue in the platitude “causes produce their effects”. On generalist (or “covering law”) accounts (see ‘Hume and Pearson: Correlation, not Causation’ and ‘Generalist Accounts’ below), causal production is a matter of law: roughly, event $c$ causes event $e$ just in case $c$ and $e$ are instances of terms in a law connecting events of $c$’s type with events of $e$’s type. The generalist interpretation is in part motivated by the need to ground inductive reasoning: unless causal relations are subsumed by causal laws, one will be unjustified in inferring that events of $c$’s type will, in the future, cause events of $e$’s type. Another motivation stems from thinking that identifying a sequence of events as causal requires identifying the sequence as falling under a (possibly unknown) law.

Alternatively, singularists (see ‘Singularist Accounts’ below) interpret causal production as involving a singular causal process (variously construed), that is metaphysically prior to laws. Singularists also argue for the epistemological priority of singular causes, maintaining that one can identify a sequence as causal without assuming that the sequence falls under a law, even when the sequence violates modal presuppositions (as in Fair’s 1979 case: intuitively, one could recognize a glass’s breaking as causal, even if one antecedently thought glasses of that type were unbreakable).

Counterfactual accounts (see ‘Counterfactual Accounts’ below) analyze singular causes in terms of counterfactual conditionals (as a first pass, event $c$ causes event $e$ just in case if $c$ had not occurred, then $e$ would not have occurred). Whether a counterfactual account should be considered singularist, however, depends on whether the truth of the counterfactuals is grounded in laws connecting types of events or in, e.g., propensities (objective single-case chances) understood as irreducible to laws. Yet another approach to the issue of singular vs. general causes is to deny that either is reducible to the other, and to rather give independent treatments of each type (as in Sober 1984).
Reduction vs. Non-reduction

There are at least three questions of reducibility at issue in philosophical accounts of causation, which largely cut across the generalist/singularist distinction. The first concerns whether causal facts (e.g., the holding of causal relations) are reducible to non-causal facts (e.g., the holding of certain spatiotemporal relations). Hume’s generalist reduction of causality (see ‘Hume and Pearson: Correlation, not Causation’ below) has a projectivist or anti-realist flavor: according to Hume, the seeming “necessary connexion” between cause and effect is a projection of a psychological habit of association between ideas, which habit is formed by regular experience of events of the cause type being spatially contiguous and temporally prior to events of the effect type. Contemporary neo-Humeans (see ‘Hempel: Explanation, not Causation’ and ‘Probabilistic Relevance Accounts’ below) dispense with Hume’s psychologism, focusing instead on the possibility of reducing causal relations and laws to objectively and non-causally characterized associations between events. (Whether such accounts are appropriately deemed “anti-realist” is a matter of dispute, one philosopher’s reductive elimination being another’s reductive introduction.) Non-reductive generalists (often called “realists”—see ‘Causal Powers, Capacities, Universals, Forces’ below), by way of contrast, take modally robust causal connection between event types to be an irreducible feature of reality (See REALISM). Singularists also come in reductive or realist varieties (see ‘Singularist Accounts’ below).

A second question of reducibility concerns whether a given account of causation aims to provide a conceptual analysis of the concept (hence to account for causation in bizarre worlds, containing magic, causal action at a distance, etc.), or instead to account for the causal relation in the actual world, in terms of physically or metaphysically more fundamental entities or processes. These different aims make a difference to what sort of cases and counterexamples philosophers of causation take to heart when developing or assessing theories. A common intermediate methodology focuses on central cases, leaving the verdict on far-fetched cases as “spoils for the victor”.

A third question of reducibility concerns whether macro-causal relations (holding between
entities, or expressed by laws, in the special sciences) are reducible to micro-causal relations (holding between entities, or expressed by laws, in fundamental physics). This question arises from a general desire to understand the ontological and causal underpinnings of the structured hierarchy of the sciences, and from a need to address, as a special case, the “problem of mental causation”: the problem of whether and how mental events (e.g., a feeling of pain) can be causally efficacious vis-à-vis certain effects (e.g., a grimacing) which appear also to be caused by the brain (and ultimately, fundamental physical) events upon which the mental events depend.

Causal reductionists (Davidson 1970, Kim 1984) suggest that mental events (more generally, macro-level events) are efficacious in virtue of supervening on (or being identical with) efficacious physical events. Many worry, however, that these approaches render macro-level events causally irrelevant (or “epiphenomenal”). Non-reductive approaches to macro-level causation come in both physicalist and non-physicalist varieties (see Wilson 1999 for an overview; and see PHYSICALISM). Some physicalists posit a relation (e.g., the determinable/determination relation or proper parthood) between macro- and micro-level events which entails that the set of causal powers of a given macro-level event $m$ (roughly, the set of causal interactions that the event, in appropriate circumstances, could enter into) are a proper subset of those of the micro-level event $p$ upon which $m$ depends. On this “proper subset” strategy the fact that the sets of causal powers are different provides some grounds for claiming that $m$ is efficacious in its own right, but since each individual causal power of $m$ is identical with a causal power of $p$, the two events are not in causal competition. On another non-reductive strategy—emergentism—the causal efficacy of at least some macro-level events (notably, mental events) is due to their having genuinely new causal powers—powers not possessed by the physical events upon which the mental events depend (see EMERGENCE). When the effect in question is physical, such powers violate the causal closure of the physical (the claim that every physical effect has a fully sufficient physical cause); but such a violation arguably isn’t at odds with any cherished scientific principles, such as conversation laws (see McLaughlin 1992).
Features of Causality: Asymmetry, Temporal Direction, Transitivity

Intuitively, causality is asymmetric: if event $c$ causes event $e$, then $e$ does not cause $c$. Causality also generally proceeds from the past to the future. How to account for these data remains unclear. The problem of explaining asymmetry is particularly pressing for accounts that reductively analyze causality in terms of laws of association, for it is easy to construct cases where the laws are reversible, but where the causation is not (as in Sylvain Bromberger’s case in which the height $h$ of a flagpole is correlated with the length $l$ of the shadow it casts, and vice versa, and intuitively $h$ causes $l$, but $l$ does not cause $h$). Both the asymmetry and temporal direction of causality can be accommodated (as in Hume) by stipulatively identifying causal with temporal asymmetry: causes differ from their effects in being prior to their effects. This approach correctly rules out $l$’s causing $h$ in the case above. But it also rules out simultaneous and backwards causation, which are generally taken to be live (or at least not too distant) possibilities.

Accounts on which the general temporal direction of causation is determined by physical or psychological processes may avoid the latter difficulties. On Reichenbach’s (1956) account, the temporal direction of causation reflects the direction of “conjunctive forks”: processes where a common cause produces joint effects, and where, in accordance with what Reichenbach called “the principle of the common cause”, the probabilistic dependence of the effects on each other is “screened off”—goes away—when the common cause is taken into account. Such forks are, he claimed, always (or nearly always) open to the future and closed to the past. Some have suggested that the direction of causation is fixed by the direction of increasing entropy, or (more speculatively) by the direction of quantum collapse of the wave packet. Alternatively, Price (1992) suggests that human experience of manipulating causes provides a basis for the (projected) belief that causality is forward-directed in time (see ‘Counterfactuals and Manipulability’ below). These accounts explain the usual temporal direction of causal processes, while allowing the occasional exception.

It remains the case, however, that accommodating the asymmetry of causation by appeal to the direction of causation rules out reducing the direction of time to the (general) direction
of causation, which some (e.g., Reichenbach) have wanted to do. More importantly, neither stipulative nor non-stipulative appeals to temporal direction seem to explain the asymmetry of causation, which intuitively has more to do with causes producing their effects (in some robust sense of ‘production’) than with causes being prior to their effects. Nonreductive accounts on which causality involves manifestations of powers or transfers of energy (or other conserved quantities) may be better situated to provide the required explanation, if such manifestations or transfers can be understood as directed (which remains controversial).

Another feature commonly associated with causality is transitivity: if \( c \) causes \( d \), and \( d \) causes \( e \), then \( c \) causes \( e \). This assumption has come in for question of late, largely due to the following sort of case (see Kvart 1991): A man’s finger is severed in a factory accident; a surgeon reattaches the finger, which afterwards becomes perfectly functional. The accident caused the surgery, and the surgery caused the finger’s functionality; but it seems odd to say that the accident caused the finger’s functionality (see Hall 2000 for further discussion).

**Challenges to Causality**

**Galileo, Newton, and Maxwell: How, not Why**

From the ancient through modern periods, accounts of natural phenomena proceeded by citing the powers and capacities of agents, bodies and mechanisms to bring about effects (see Hankinson 1998 and Clatterbaugh 1999). Galileo’s account of the physics of falling bodies initiated a different approach to scientific understanding, on which this was a matter of determining how certain measurable quantities are functionally correlated (the “how” of things, or the kinematics), as opposed to determining the causal mechanisms responsible for these correlations (the “why” of things, or the dynamics). This descriptive approach enabled scientific theories to be formulated with comparatively high precision, which in turn facilitated predictive and retrodictive success; by way of contrast, explanations in terms of (often unobservable) causal mechanisms came to be seen as explanatorily otiose, at best, and unscientific, at worst.

Newton’s famous claim in the *Principia* (1687/1999) that “Hypotheses non fingo” (“I
frame no hypotheses”) regarding gravitation’s “physical causes and seats” is often taken as
evidence that he advocated a descriptivist approach (though he speculated at length as re-
gards the causes of gravitational forces in the *Optics*). And while Maxwell drew heavily
upon Faraday’s qualitative account of causally efficacious electromagnetic fields (and associ-
ated lines of force) in the course of developing his theories of electricity and magnetism, he
later saw such appeals to underlying causes as heuristic aids which could be dropped from the
final quantitative theory. Such descriptivist tendencies have been encouraged by perennial
worries about the metaphysical and epistemological presuppositions of explicitly causal ex-
planations (see ‘Causal Powers, Capacities, Universals, Forces’ below), and the concomitant
seeming availability of eliminativist or reductivist treatments of causal notions in scientific
laws. For example, Russell (1912) influentially argued that since the equations of physics
do not contain any terms explicitly referring to causes or causal relations, and moreover (in
conflict with the presumed asymmetry of causality) appear to be functionally symmetric (one
can write \( a = f/m \) as well as \( f = ma \)), causality should be eliminated as “a relic of a by-
gone age”. Jammer (1957) endorsed a view on which force-based dynamics is a sophisticated
form of kinematics, with force terms being mere “methodological intermediaries” enabling
the convenient calculation of quantities (e.g., accelerations) entering into descriptions. And
more recently, van Fraassen (1980) has suggested that, while explanations going beyond de-
scriptions may serve various pragmatic purposes, these have no ontological or causal weight
beyond their ability to “save the phenomena”.

Whether science really does, or should, focus on the (non-causally) descriptive is, how-
ever, deeply controversial. Galileo himself sought for explanatory principles going beyond
description (see Jammer 1957 for discussion), and notwithstanding Newton’s professed neu-
trality about their physical seats, he took forces to be the “causal principle[s] of motion and
rest”. More generally, notwithstanding the availability of interpretations of scientific the-
ories as purely descriptive, there are compelling reasons (say, the need to avoid a suspect
action at a distance) for taking the causally explanatory posits of scientific theories (e.g.,
fields and forces) ontologically seriously. The deeper questions here, of course, concern how
to assess the ontological and causal commitments of scientific theories; and at present there
is no philosophical consensus on these important matters. In any case it is not enough, in assessing whether causes are implicated by physical theories, to note that terms like “cause” don’t explicitly appear in the equations of the theory, insofar as the commitments of a given theory may transcend the referents of the terms appearing in the theory, and given that many terms that do so appear—force, charge, valence—are most naturally defined in causal terms (‘force’, for example, is usually defined as ‘that which causes acceleration’). It is also worth noting that the apparent symmetry of many equations, as well as the fact that cause terms do not explicitly appear in scientific equations, may be artifacts of scientists’ using the identity symbol as an all-purpose connective between functional quantities, that enables the quantities to be manipulated using mathematical techniques, but which is nonetheless implicitly understood as causally directed, as in $f = ma$.

Nor does scientific practice offer decisive illumination on whether scientific theorizing is or is not committed to causal notions: as with Maxwell, it remains common for scientists to draw upon apparently robustly causal notions when formulating or explaining a theory, even while maintaining that the theory expresses nothing beyond descriptive functional correlations of measurable quantities. Perhaps it is better to attend to what scientists do rather than what they say. That they rarely leave matters at the level of descriptive laws linking observables is some indication that they are not just concerned with the “how” question—though, to be sure, the tension between descriptive and causal/explanatory questions may recur at levels below the surface of observation.

**Hume and Pearson: Correlation, not Causation**

The Galilean view of scientific understanding as involving correlations among measurable quantities was philosophically mirrored in the empiricist view that all knowledge (and meaning) is ultimately grounded in sensory experience (see EMPIRICISM). The greatest philosophical challenge to causality came from Hume, who argued that there is no experience of causes being efficacious, productive, or powerful vis-à-vis their effects; hence “we only learn by experience the frequent CONJUNCTION of objects, without being ever able to comprehend anything like CONNEXION between them” (1748, p. 46). In place of realistically
interpreted “producing” theories of causation (see Strawson 1987 for a taxonomy) Hume offered the first regularity theory of causation, according to which event $c$ causes event $e$ just in case $c$ and $e$ occur, and events of $c$’s type have (in one’s experience) been universally followed by, and spatially contiguous to, events of $e$’s type. (As discussed, such constant conjunctions were the source of the psychological imprinting that was, for Hume, the ultimate locus of causal connection.) Hume’s requirement of contiguity may be straightforwardly extended to allow for causes to produce distant effects, via chains of spatially contiguous causes and effects.

Hume’s requirement of universal association is not sufficient for causation (night always follows day but night does not cause day). Nor is Hume’s requirement necessary, for even putting aside the requirement that one experience the association in question, there are many causal events that happen only once (e.g., the Big Bang). The immediate move of neo-Humeans (e.g. Hempel and Oppenheim 1948 and Mackie 1965) is to understand the generalist component of the account in terms of laws of nature, which express the lawful sufficiency of the cause for the effect, and where (reflecting Hume’s reductive approach) the sufficiency is not to be understood as grounded in robust causal production. One wonders, though, what is grounding the laws in question, if associations are neither necessary nor sufficient for their holding, and robust production is not allowed to play a role. If the laws are grounded in brute fact, it’s not clear that the reductive aim has been served (but see the discussion of Lewis’s account of laws, below).

In any case, neo-Humean accounts face several problems concerning events that are inappropriately deemed causes (“spurious causes”). One is the problem of joint effects, as when a virus causes first a fever, and then independently causes a rash: here the fever is lawfully sufficient for, hence incorrectly deemed a cause of, the rash. Another involves violations of causal asymmetry: where events of the cause’s type are lawfully necessary for events of the effect’s type, the effect will be lawfully sufficient for (hence inappropriately deemed a cause of) the cause. Cases of preemption also give rise to spurious causes: Suzy’s and Billy’s rock-throwings are each lawfully sufficient for breaking the bottle; but given that Suzy’s rock broke the bottle (thereby preempting Billy’s rock from doing so), how to rule out Billy’s
rock-throwing as a cause?

The above cases indicate that lawful sufficiency alone is not sufficient for causality. One response is to adopt an account of events on which these are finely individuated, so that, for example, the bottle-breaking resulting from Suzy’s rock-throwing turns out to be of a different event-type than a bottle-breaking resulting from Billy’s rock-throwing (in which case Billy’s rock-throwing does not instantiate a rock-throwing–bottle-breaking law, and so does not count as a cause). Another response incorporates a proviso stating that the lawful sufficiency at issue is sufficiency in the circumstances (as on Mackie’s ‘INUS’ condition account, on which a cause is an Insufficient but Necessary part of a condition that is, in the circumstances, Unnecessary but Sufficient for the effect).

Nor is lawful sufficiency alone necessary for causality, as the live possibility of irreducibly probabilistic causality indicates. This worry is usually sidestepped by a reconception of laws according to which these need express only some lawlike pattern of dependence; but this reconception makes it yet more difficult for regularity theorists to distinguish spurious from genuine causes and laws (see ‘Probabilistic Relevance Accounts’ below for developments). One neo-Humean response is to allow certain a priori constraints to enter into determining what laws there are at a world, as in the the “Best System” theory of laws of Lewis 1994, on which the laws at a world are those that systematize the phenomena with the best combination of (predictive) strength and (formal) simplicity, so as to accommodate probabilistic (and even uninstantiated) laws (see also ‘Hempel: Explanation, not Causation’ below).

The view that causation is nothing over and above (appropriately complex) correlations was widespread following the emergence of social statistics in the nineteenth century, and was advanced by Karl Pearson (The Grammar of Science, 1890), one of the founders of modern statistics. Pearson’s endorsement of this view was, like Hume’s, inspired by a rejection of causality as involving mysterious productive powers, and contributed to causes (as opposed to associations) being to large extent expunged from statistics and from the many sciences relying upon statistics. An intermediate position between these extremes, according to which causes are understood to go beyond correlations, but are not given any particular metaphysical interpretation (in particular, as involving productive powers), was advanced by the
evolutionary biologist Sewall Wright, the inventor of path analysis. Wright (1921) claimed not only that path analysis enabled previously known causal relations to be appropriately weighted, but moreover enabled the testing of causal hypotheses in cases where the causal relations were (as yet) unknown. Developed descendants and variants of Wright’s approach have found increasing favor of late (see ‘Bayesian Networks and Causal Models’ below), contributing to some rehabilitation of the notion of causality in the statistical sciences.

**Hempel: Explanation, not Causation**

Logical empiricists were suspicious of causation understood as a metaphysical connection in nature; instead they located causality in language, interpreting causal talk as talk of explanation (see Salmon 1989 for discussion). On Hempel and Oppenheim’s (1948) influential D–N (Deductive–Nomological) model of scientific explanation, event $c$ explains event $e$ just in case a statement expressing the occurrence of $e$ is the conclusion of an argument with premises, one of which expresses the holding of a universal generalization to the effect that events of $c$’s type are associated with events of $e$’s type, and another of which expresses the fact that $c$ occurred (see EXPLANATION). Imposing certain requirements on universal generalizations (e.g., projectibility) enabled the D–N account to avoid cases of spurious causation due to accidental regularities ($s$ being a screw in Smith’s car does not explain why $s$ is rusty, even if all the screws in Smith’s car are rusty). And requiring that explanations track temporal dependency relations (as a variation on identifying causal with temporal asymmetry) can prevent, for example, the length of the shadow from explaining the height of the flagpole. It is less clear, however, how to deal with explanatory preemption, as when Jones, immediately after ingesting a pound of arsenic, is run over by a bus and dies. Hempel’s account allows us to explain Jones’s death by citing the law that anyone who ingests a pound of arsenic dies within 24 hours, along with the fact that Jones ate a pound of arsenic; but such an explanation will be spurious. One way to respond is to require that the premises in explanatory arguments only cite laws and facts that are causally relevant to the event being explained; but this is in obvious tension with the empiricist’s goal of characterizing causation in terms of explanation.

To accommodate the possibility of irreducibly probabilistic associations, and explana-
tions (characteristic of the social sciences) proceeding under conditions of partial uncertainty, Hempel (1965) proposed an inductive-statistical (I–S) model, in which event $c$ explains event $e$ if $c$ occurs and it is an inductively grounded law that the probability of an event of type $E$ given an event of type $C$ is high (see EXPLANATION; INDUCTIVE LOGIC). This account is subject to counterexamples in which a cause produces an effect but with a low probability, as in Scriven’s case (discussed by him prior to Hempel’s extension, and developed in Scriven 1975) where the probability of paresis given syphilis is low, but when paresis occurs, syphilis is the reason. A similar point applies to many quantum processes. Such cases gave rise to two different approaches to handling probabilistic explanation (or causation, by those inclined to accept this notion). One approach (see Railton 1978) locates probabilistic causality in propensities; the other (see ‘Probabilistic Relevance Accounts’ below) in more sophisticated probabilistic relations.

At this point the line between accounts that are reductive (in the sense of reducing causal to non-causal goings-on) and nonreductive, as well as the line between singularist and generalist accounts, begins to blur. For while a propensity-based account of causality initially looks nonreductive and singularist, some think that propensities can be accommodated on a sophisticated associationist account of laws; and while an account based on relations of probabilistic relevance initially looks reductive and generalist, whether it is so depends on how the probabilities are interpreted (e.g., as given by frequencies, irreducible propensities, etc.).

**Contemporary Generalist Accounts**

**Probabilistic Relevance Accounts**

A natural response to Scriven-type cases is to understand positive causal relevance in terms of probability-raising (Suppes 1970): event $c$ causes event $e$ just in case the probability of events of $e$’s type is higher given events of $c$’s type than without. (Other relevance relations, such as being a negative causal factor, can be defined accordingly.) A common objection to such accounts (see Rosen 1978) proceeds by constructing “doing it the hard way” cases in
which it seems that causes lower the probability of their effects (e.g., where a mis-hit golf ball
ricochets off a tree, resulting in a hole-in-one; or where a box contains a radioactive substance
$S$ which produces decay particles, but where the presence of $S$ excludes the the more effective
radioactive substance $S'$). Such cases can often be handled, however, by locating a neutral
context (where the golf ball is not hit at all, or where no radioactive substance is in the box)
relative to which events of the given type do raise the probability of events of the effect type.

A more serious problem for probability-raising accounts is indicated by Simpson’s paradox,
according to which any statistical relationship between two variables may be reversed by
including additional factors in the analysis. The “paradox” reflects the possibility that a
variable $C$ can be positively correlated with a variable $E$ in a population, and yet $C$ be
negatively correlated with $E$ in every partition of the population induced by a third variable
$X$. Just this occurred in the Berkeley sex discrimination case. Relative to the population of all
men and women applying to grad school at Berkeley, being male ($C$) was positively correlated
with being admitted ($E$). But relative to every partition of the population containing the
men and women applying to a particular department ($X$), this correlation was reversed. In
this case the difference between the general and specific population statistics reflected the
fact that, while in every department it was easier for women to be admitted than men,
women were more likely to apply to departments that were harder (for everyone) to get into.
The general population statistic was thus confounded: being a male, simpliciter, was not
in fact causally relevant to getting into grad school at Berkeley; rather, (assuming no other
confounding was at issue) applying to certain departments rather than others was what was
relevant.

To use probabilistic accounts as a basis for testing hypotheses and making predictions—
and especially in order to identify effective strategies (courses of action) in the social sciences
and indeed, in everyday life—statistical confounding needs to be avoided. Cartwright (1979)
suggested that avoiding confounding requires that the relevant probabilities be assessed rel-
ative to background contexts within which all other causal factors (besides the variable $C$
whose causal relevance is at issue) are held fixed. Opinions differ regarding whether events
of type $C$ must raise the probability of events of type $E$ in at least one such context, in a
majority of contexts, or in every such context. So one might take $C$ to be a positive causal factor for $E$ just in case $P(E|C \land X_i) \geq P(E|\bar{C} \land X_i)$ for all background contexts $X_i$, with strict inequality for at least one $X_i$. On this approach, smoking would be a positive causal factor for having lung cancer just in case smoking increases the chance of lung cancer in at least one background context, and does not lower it any background context.

Practically, Cartwright’s suggestion has the disadvantage that one is frequently not in a position to control for all alternative causal factors (though in some circumstances one can avoid having to do this; see ‘Bayesian Networks and Causal Models’ below). Philosophically, the requirement threatens reductive versions of probabilistic accounts with circularity. Attempts have been made to provide a non-circular means of specifying the relevant background contexts (e.g., Salmon 1984) but it is questionable whether these attempts succeed, and many are presently prepared to agree with Cartwright that “no causes in, no causes out”.

As mentioned, probabilistic accounts may or may not be reductive, depending on whether the probabilities at issue are understood as grounded in associations (as in Suppes 1970) or else in powers, capacities or propensities (as in Humphreys 1989 and Cartwright 1989). On the latter interpretation, further divisions are introduced: if the propensities are taken to be irreducible to laws (as on Cartwright’s account), then the associated probabilistic relevance account is more appropriately deemed singularist. Complicating the taxonomy here is the fact that most proponents of probabilistic accounts are not explicit as regards what analysis should be given of the probabilities at issue.

**Bayesian Networks and Causal Models**

Philosophical worries concerning whether statistical information adequately tracks causal influence are echoed in current debates over the interpretation of the statistical techniques used in the social sciences. As noted above (‘Hume and Pearson: Correlation, not Causation’) these techniques have frequently been interpreted as relating exclusively to correlations, but increasingly researchers in computer science, artificial intelligence and statistics (see Spirtes et al. 1993 and Pearl 2000) have developed interpretations of these approaches as encoding explicitly causal information.
On the causal modeling approach, one starts with a set of variables (representing properties) and a probability distribution over the variables. This probability distribution, partially interpreted with prior causal knowledge, is assumed to reflect a causal structure (a set of laws expressing the causal relations between the variables, which laws may be expressed either graphically or as a set of structurally related functional equations). Given that certain conditions (to be discussed shortly) hold between the probabilities and the causal structure, algorithmic techniques are used to generate the set of all causal structures consistent with the probabilities and the prior causal knowledge. Techniques also exist for extracting information regarding the results of interventions (corresponding to manipulations of variables). Such strategies appear to lead to improved hypothesis testing and prediction of effects under observation and intervention.

Another advantage claimed for such accounts is that they provide a means of avoiding confounding without imposing the often impracticable requirement that the relevant probabilities be assessed against background contexts taking into account all causal factors, it rather being sufficient to take into account all common causal factors. As a simple illustration, suppose that $A$ and $B$ are known to causally influence $C$, as in Figure 1.

\[
\begin{array}{ccc}
A & B \\
& \uparrow & \uparrow \\
D & C
\end{array}
\]

Figure 1: $A$ is a common causal factor of $D$ and $C$

To judge whether $D$ influences $C$, Cartwright generally recommends holding fixed both $A$ and $B$, while Spirtes et al. 1993 instead recommend holding fixed only $A$. Cartwright allows, however, that attention to just common causal factors is possible when the causal Markov condition holds. Since this is one of the conditions that must hold in order to implement the causal modeling approach; so the restriction to common causal factors is not really an advantage over Cartwright’s account.
While causal modeling approaches may lead to improved casual inference concerning complex systems (in which case they are of some epistemological interest), it is unclear what bearing they have on the metaphysics of causality. Spirtes et al. present their account not so much as an analysis of causality as a guide to causal inference. Pearl (2000), however, takes the appeal to prior causal intuitions to indicate that causal modeling approaches are nonreductive (and moreover based in facts about humans’ cognitive capacities to make effective causal inferences in simple cases). In any case, the potential of causal models to provide a basis for a general theory of causality is limited by the need for certain strong conditions to be in place in order for the algorithms to be correctly applied.

One of these is the aforementioned causal Markov condition (of which Reichenbach’s 1956 “principle of the common cause” was a special case), which says that once one conditions on the complete set $P_v$ of “causal parents” (direct causes) of a variable $V$, $V$ will be probabilistically independent of all other variables except $V$’s descendants: that is, for all variables $X$, where $X$ is not one of $V$’s descendants, $p(V|P_v \land X) = p(V|P_v)$. (In particular, where $V$ is a joint effect, conditioning on the causal parents of $V$ screens off the probabilistic influence of the other joint effects on $V$.) Here again there is the practical problem that in the social sciences, where the approaches are supposed to be applicable, one is often not in a position to specify states with sufficient precision to guarantee that the condition is met. A metaphysical problem is that (contrary to Reichenbach’s apparent assumption that the condition holds in all cases involving a common cause of joint effects) the causal Markov condition need not hold in cases of probabilistic causation: when a particle may probabilistically decay either by emitting a high-energy electron and falling into a low-energy state, or by emitting a low-energy electron and falling into a different energy state, the joint effects in either case will not be probabilistically independent of each other, even conditioning on the cause; and certain cases of macro-causation appear also to violate the condition.

A second assumption of the causal modeling technique is what Spirtes et al. 1993 call “faithfulness” (also known as “stability” in Pearl 2000), according to which probabilistic dependencies faithfully reveal causal connections. In particular, if $Y$ is probabilistically independent of $X$, given $X$’s parents, then $X$ is assumed not to cause $Y$. Again, this condition
cannot be assumed to hold in all cases, since some variables (properties) may sometimes prevent, sometimes produce, an effect (as when birth control pills are a cause of thrombosis, yet also prevent thrombosis, insofar as pregnancy causes thrombosis, and the pills prevent pregnancy). In circumstances where the positive and negative contributions of $X$ to $Y$ are equally effective, the probabilistic dependence of effect on cause may cancel out, and thus $X$ inappropriately be taken not to be causally relevant to $Y$.

**Causal Powers, Capacities, Universals, Forces**

As mentioned, some proponents of probabilistic relevance accounts endorse metaphysical interpretations of the probabilities at issue. Such positions fall under the broader category of non-reductive (“realist”) covering-law theories in which laws express (or are grounded in) more than mere associations. The job such accounts face is to provide an alternative basis for causal laws. Among other possibilities, these bases are taken to be relations of necessitation or “probabilification” among universals (Dretske 1977, Tooley 1977, Armstrong 1978), (law-based) capacities or powers associated with objects or properties (Shoemaker 1980, Martin 1993), or fundamental forces or interactions (Bohm 1957, Strawson 1987).

Such accounts sidestep many of the problems associated with reductive covering-law accounts. Since laws are not just a matter of association, a realist has the means to deny, in the virus–fever–rash case, that there is a law connecting fevers with rashes; similarly, in cases of preemption a realist may claim that, for example, Billy’s rock-throwing and the bottle’s breaking did not instance the law in question (even without endorsing a fine-grained account of event individuation). Of course much depends here on the details of the proposed account of laws. On the Dretske-Tooley-Armstrong account, causal laws are contingent, brute relations between universals; some find this less than satisfying from a realist point of view, insofar on it is compatible with, for example, the property of *having spin* 1 bestowing completely different causal powers (say, all those actually bestowed by *having spin* 1/2) on its possessing particulars. Other realists are more inclined to see the nature of properties and particulars as essentially dependent upon the causal laws that actually govern them, a view which is not implausible for scientific entities: “[C]ausal laws are not like externally imposed
legal restrictions that, so to speak, merely limit the course of events to certain prescribed paths […] the causal laws satisfied by a thing […] are inextricably bound up with the basic properties of the thing which helps to define what it is” (Bohm 1957, p. 14).

The primary problem facing realist accounts is that they require accepting entities and relations (universals, causal powers, forces) that many philosophers and scientists find metaphysically obscure and/or epistemologically inaccessible. How one assesses these assessments often depends on one’s other commitments. For example, many traditional arguments against realist accounts (e.g., Hume’s arguments) are aimed at showing that these do not satisfy a strict epistemological standard, according to which the warranted posit of a contingent entity requires that the entity be directly accessible to experience (or a construction from entities that are so accessible). But if inference to the existence of an unexperienced entity (as the best explanation of some phenomena) is at least sometimes an acceptable mode of inference, such a strict epistemological standard (and associated arguments) will be rejected; and indeed, positive arguments for contemporary realist accounts of causality generally proceed via such inferences to the best explanation—often, of the patterns of association appealed to by reductivist accounts.

**Contemporary Singularist Accounts**

Singularists reject the claim that causes follow laws in the order of explanation, but beyond this there is considerable variety in their accounts.

*Contra* Hume, Anscombe (1971) takes causation to be a (primitive) relation that may be observed in cuttings, pushings, fallings, etc. It’s worth noting that a primitivist approach to causality is compatible with even a strict empiricism (compare Hume’s primitivist account of the resemblance relation). The empiricist Ducasse 1926 also locates causation in singular observation, but non-primitively: a cause is the change event observed to be immediately prior and spatiotemporally contiguous to an effect event. While interesting in allowing for an non-associative, non-primitivist, empiricist causality, Ducasse’s account is unsatisfactory in allowing only the coarse-grained identification of causes (as some backwards temporal segment
of the entire observed change); hence it fails to account for most ordinary causal judgments. Note that singularists basing causation in observation need not assert that one’s knowledge of causality proceeds only via observations of the preferred sort; they rather generally maintain that such experiences are sufficient to account for one’s acquiring the concept of causation, then allow both that causation need not be observed, and that confirming singular causal claims may require attention to associations. Another singularist approach takes causation to be theoretically inferred, as that relation satisfying (something like) the Ramsey sentence consisting of the platitudes about causality involving asymmetry, transitivity, and so on (see Tooley 1987). One problem here is that, as may be clear by now, such platitudes do not seem to uniformly apply to all cases. Relatedly, one may wonder whether they are consistent; given the competing causal intuitions driving various accounts of causality, it would be surprising if they were.

Finally, a wide variety of singularist accounts understand causality in terms of singular processes. Such accounts are strongly motivated by the intuition that, in cases of preemption (e.g., the Suzy and Billy case), what distinguishes Suzy’s throw as a cause is that Suzy’s throw initiates a process ending in the shattering, while the process initiated by Billy’s throw never reaches completion (see Menzies 1996 for discussion). Commonly, process singularists attempt (like Ducasse) to provide a non-primitivist causality that is both broadly empiricist, in not appealing to any properly metaphysical elements, and non-associationist, in recognition of the difficulties associationist accounts have (both with preemption and with distinguishing genuine causality from accidental regularity). Hence they typically fill in the “process” intuition by identifying causality with fundamental physical processes, including transfers or interactions, as on Fair’s (1979) account of causation as identical with the transfer of energy-momentum, Salmon’s (1984) “mark-transmission” account, and Dowe’s (1992) account in which the transfer of any conserved quantity will suffice.

An objection to the claim that physical processes are sufficient for causality is illustrated by Cartwright’s (1979) case of a plant sprayed with herbicide, which improbably survives and goes on to flourish (compare also Kvart’s 1991 “finger-severing” case, discussed previously). While transfers and interactions of the requisite sort can be traced from spraying
to flourishing, intuitively the former did not cause the latter; however, accepting that the spraying did cause the flourishing may not be an overly high price to pay. A deeper worry concerns the epistemological question of how physical process accounts link causation, understood as involving theoretical relations or processes of fundamental physics, with causation as ordinarily experienced. Fair suggests that ordinary experience involves the experience of macro-processes, which are in turn reducible to the relevant physical processes; but even supposing such reductions are in place, ordinary causal judgments do not seem to presuppose them.

**Contemporary Counterfactual Accounts**

Counterfactual accounts of causality, which may also be traced back to Hume, take as their starting point the intuition that a singular cause makes an important difference to what happens. As a first pass, $c$ causes $e$ (where $c$ and $e$ are actually occurring events) only if, were $c$ not to occur, then $e$ would not occur. As a second pass, $c$ causes $e$ only if $c$ and $e$ are connected by a chain of such dependencies (see Lewis 1973), so as to ensure that causation is transitive (causation is thus the “ancestral”, or transitive closure, of counterfactual dependence). In addition to the requirement of counterfactual necessity of causes for effects, counterfactual accounts also commonly impose a requirement of counterfactual sufficiency of causes for effects: if $c$ were to occur, then $e$ would occur. Insofar as counterfactual accounts are standardly aimed at reducing causal to non-causal relations, and given plausible assumptions concerning evaluation of counterfactuals, the latter requirement is satisfied just by $c$ and $e$’s actually occurring (which occurrences, as above, are assumed); hence standard counterfactual accounts do not have a non-trivial notion of counterfactual sufficiency. A non-trivial notion of counterfactual sufficiency can be obtained by appeal to nested counterfactuals (see Vihvelin 1995): $c$ causes $e$ only if, if neither $c$ nor $e$ had occurred, then if $c$ had occurred, $e$ would have occurred.
Problems, Events, and Backtrackers

While counterfactual accounts are often motivated by a desire to give a reductive account of causality that avoids problems with reductive covering law accounts (especially those of joint effects and of preemption), it is unclear whether counterfactual accounts do any better by these problems. First, consider the problem of joint effects. Suppose a virus causes first a fever, and then a rash, and suppose moreover that the fever and rash could only have been caused by the virus. It seems correct to reason in the following “back-tracking” fashion: if the fever had not occurred, then the viral infection would not have occurred, in which case the rash would have not occurred either. But then the counterfactual ‘If the fever had not occurred, the rash would not have occurred’ turns out true, which here means that the fever causes the rash, which is incorrect. Proponents of counterfactual accounts have responses to these objections, which require accepting controversial accounts of the truth conditions for counterfactuals (see Lewis [1979]); even so the responses appear not to succeed (see Bennett 1984 for discussion).

Second, consider the problem of preemption. In the Suzy-Billy case, it seems correct to reason: if Suzy had not thrown her rock, then Billy’s rock would have gotten through and broken the bottle. Hence the counterfactual ‘If Suzy’s throw had not occurred, the bottle-breaking would not have occurred’ turns out false; so Suzy’s rock-throwing turns out not to be a cause, which is incorrect. In cases (as here) of so-called “early preemption”, where it makes sense to suppose that there was an intermediate event \( d \) between the effect and the cause upon which the effect depended, this result can be avoided: although the breaking does not counterfactually depend on Suzy’s throwing, there is a chain of counterfactual dependence linking the breaking to Suzy’s throwing (and no such chain linking the breaking to Billy’s throw), and so her throw does end up being a cause (and Billy’s does not). But the appeal to an intermediate event seems \textit{ad hoc}, and in any case cannot resolve cases of “late preemption”. Lewis (developing an idea due to Paul 1998) eventually responded to such cases by allowing that an event may be counted as a cause if it counterfactually influences the mode of occurrence of the effect (e.g., how or when it occurs), as well as if it counterfactually
influences the occurrence of the effect, *simpliciter*.

**Counterfactuals and Manipulability**

Where counterfactual accounts may be most useful is in providing a basis for understanding or formalizing the role manipulability plays in the concept of causation. One such approach sees counterfactuals as providing the basis for an epistemological, rather than a metaphysical, account of causation (see Pearl 2000 for discussion). The idea here is that counterfactuals nicely model the role manipulability (actual or imagined) plays in causal inference, for a natural way to determine whether a counterfactual is true is to manipulate conditions so as to actualize the antecedent. Another approach takes counterfactuals to provide a basis for a generalist account of causal explanation (see Woodward 1997), according to which such explanations track stable or invariant connections, and where the notion of invariance is understood non-epistemologically in terms of a connection’s continuing to hold through certain counterfactual (not necessarily human) “interventions”. Whether the notion of manipulability is itself a causal notion, and so bars the reduction of causal to non-causal facts, is still an open question.

**References**


23


