What is Causation?

Stathis Psillos

INTRODUCTION

When we philosophers think about causation we are primarily interested in what causation *is* – what exactly is the relation between cause and effect? Or, more or less equivalently, how and in virtue of what is the cause connected to the effect? But we are also interested in an epistemic issue, viz., the possibility of causal knowledge: how, if at all, can causal knowledge be obtained? The two issues are, of course, conceptually distinct – but to many thinkers, there is a connection between them. A metaphysical account of causation would be useless if it did not make, at least in principle, causal knowledge possible. Conversely, many philosophers, mostly of an empiricist persuasion, have taken the possibility of causal knowledge to act as a constraint on the metaphysics of causation: no feature that cannot in principle become the object of knowledge can be attributed to causation.

As we shall see, though philosophers have thought about causation for ages, there is hardly an agreement as to what causation is. Perhaps, this is reason enough to be sceptical about the whole prospect for a single and unified metaphysical account of causation. Perhaps, what we are trying to figure out – *causation* – is not one single condition with a determinate nature. But this crisis in the foundations of causation is no reason for despair. After all, we can know a lot about what causes what without knowing what causation is – or better, without being committed to the view that causation is one single thing in the world.

Causation is too precious to be given up. We search for causes in order to *explain* and *understand* how and why things around us are the way they are, or behave and change in certain ways. We also search for causes in order to *intervene* in the course of nature (or in the course of events, in general) and bring certain effects about or prevent others from occurring. We are interested in causation because we are *thinkers* and *agents*, because we are both theoretical and practical beings. From Aristotle on, commitment to causation has been constitutive of the ways we conceptualise the world, interpret it and change it. Even David Hume, who was critical of the whole idea of offering a metaphysics of causation, noted that causation (together with resemblance and contiguity – the three principles of association) is "really *to us* the cement of the universe, and all the operations of the mind must, in a great measure, depend on [it]."

In this paper, I will offer a review of the basic trends in our current thinking about causation, aiming also to highlight their historical origins and development.

TWO CONCEPTS OF CAUSATION

What are the data of a philosophical theory of causation? What kinds of constraint are imposed on it? It seems that, no matter what the details of the philosophical theory, there are a number of platitudes that our ordinary concept of causation satisfies – stemming mostly from how we ordinarily think about causation and what we want to do with it. Here are four of them:

- The difference platitude: causes make a difference, viz., things would be different if the causes of some effects were absent.
- The recipe platitude: causes are recipes for their effects, viz., causes are the means to produce (or prevent) certain ends (effects).
- The explanation platitude: causes explain their effects, but not vice versa.
- The evidence platitude: causes are evidence for their effect, viz., knowing that *c* causes *e*, and knowing that *c* occurred, gives us (some) reason to expect that *e* will occur.

It can be argued that any philosophical theory of causation should accommodate these platitudes, that is, show how each of them is brought out by whatever constitutes, according to the theory, the relation of cause and effect. But if we dig a bit deeper, we can see that – alongside these platitudes –there are two very firm pre-philosophical views about what causation is – what I might call *intuitions* about causation.

- The regularity intuition: whether or not a sequence of two distinct events *c* and *e* is causal depends on whether or not events like *c* are regularly followed by events like *e*.
- The intrinsic-relation intuition: whether or not a sequence of two distinct events *c* and *e* is causal depends wholly on the events *c* and *e* and their own properties and relations, that is, it depends wholly on the intrinsic and local features of the actual sequence of events.

The regularity intuition is mostly driven by folk epistemological considerations: how can causal relations be known or reliably manipulated *unless* they embody or instantiate regularities? Unless, that is, they are such that it is part of the causal connection that links two events that whenever the cause obtains so does the effect? This intuition is underpinned by the fact that we are unwilling to pronounce a sequence of events *c* and *e* causal unless there has been a regular association between events like *c* and events like *e*. If a causal relation was a one-off thing (*this* causing *that* here and now), causation would be of little usefulness and causal knowledge would require some kind of special non-inductive method. The intrinsic-relation intuition, on the other hand, is mostly driven by folk metaphysical considerations: causal relatedness is a matter of something *in* the cause bringing about the effect; it is a tie between cause and effect which is independent of things that happen at other places and other times. These intuitions are equally firm, but each of them is too controversial to be taken as a platitude of causation.

The problem, of course, is that no single theory of causation can accommodate both intuitions by giving equal footing to them. There can, of course, be compatibilist accounts of causation. There can be accounts of causation based on the intrinsic-relation intuition that can accommodate the thought that, as a matter of fact, causal relations give rise to stable dependencies (e.g. regularities: the world is essentially nomological). But such accounts put a premium on the folk metaphysical intuition of intrinsic relation. Conversely, there can be accounts of causation based on the regularity intuition that can accommodate the thought that, as a matter of fact, causal relations are exemplified by regularly operating mechanisms. But such accounts put a premium on the folk epistemological intuition of regularity. Both moves presuppose that one of the two intuitions is really more central to our folk theory of causation, while the other is derivative. An egalitarian view that gives, as it were, equal footing to both intuitions is thereby excluded.

A natural reaction at this stage is to start questioning the presumed fact that we are dealing with a single, unitary, concept of causation. The case for there being two concepts of CAUSATION has been made by

Ned Hall (2004). He has distinguished between causation-as-dependence and causation-as-production. Hall takes dependence to be simple counterfactual dependence, while he takes the concept of production (*c* produces *e*) as primitive. This distinction is plausible. In fact, it can be argued that there have been (historically and conceptually) two broad approaches to the metaphysical issue of causation.[1] On the dependence approach, to say that *c* causes *e* is to say that *e* suitably depends on *c*. On the production approach, to say that *c* causes *e* is to say that something in the cause produces (brings about) the effect or that there is something (e.g., a mechanism) that links the cause and the effect. There have been different ways to cash out the relation of dependence: nomological dependence (cause and effect fall under a law); counterfactual dependence (if the cause hadn't happened, the effect wouldn't have happened); probabilistic dependence (the cause raises the probability of the effect). Similarly, there have been different ways to cash out the concept of production, but the most prominent among them are cast in terms of something being transferred from the cause to the effect (e.g., a property, or some physical quantity – force, energy etc.). A key thought in the production approach is that cause and effect are connected by means of a local mechanism.

Why should we take seriously the two-concept view? One reason is that the two concepts align quite naturally with the distinct intuitions about causation: the production view aligns with the intrinsic-relation intuition, while the dependence approach aligns with the intuition that causation is an extrinsic relation between events (a species of which is a regularity). Another reason is that the two views set conceptually distinct constraints on causal relatedness. On the production view causal connectedness amounts to the presence of some tie between cause and effect, while no such tie is required by the dependence approach – just a robust dependence. Finally, dependence theories and production theories are extensionally distinct. There can be cases of causation licensed by dependence theories without being licensed by production theories and conversely.[2]

Still, one may wonder: in virtue of what is it the case that these two concepts are both concepts of CAUSATION? There are some interesting answers to this question. The two-concept view is genuinely egalitarian: since there is no way to privilege one basic intuition concerning CAUSATION over the other, and since there is no other way to tie causation uniquely with either production or dependence, egalitarianism dictates that both approaches are equally acceptable accounts of causation. Besides, both concepts of causation, suitably developed, can easily accommodate the platitudes associated with causation. The difficulty is that the two-concept view does not tell us when to apply the one concept and when the other: we somehow have to rely on our intuitions for this.

Let us now take a synoptic look at the major competitors for a theory of causation and their problems.

DEPENDENCE THEORIES

The Regularity View of Causation

On this view that goes back to Hume, causation reduces to a relation of spatiotemporal contiguity, succession and constant conjunction (regularity) between distinct events. That is, c causes e iff

[1] For more on this, see my (2004).

[2] Most typical cases, however, concern situations based on no clear-cut intuitions, such as causal overdetermination and causation by disconnection.

i. *c* is spatiotemporally contiguous to *e*;
ii. *e* succeeds *c* in time; and
iii. all events of type *C* (i.e., events that are like *c*) are regularly followed by (or are constantly conjoined with) events of type *E* (i.e., events like *e*).

A corollary of this view is that there is no necessity in causation: there is no necessary connection between the cause *c* and the effect *e* that goes beyond – or underpins – their regular association. This is a big break with a long philosophical tradition that goes back to Aristotle and Descartes. In the rich Aristotelian ontology, causes were essential properties of their subjects and necessitated their effects. Aristotle thought that the logical necessity by which the conclusion follows from the premises of an explanatory argument mirrors the physical necessity by which causes produce their effects. This thought (that there is necessity in nature and that this necessity is the same as the logical necessity of a demonstrative argument) was adhered to by many moderns until Hume subjected it to severe criticism.

Before Hume, it was Nicholas Malebranche's occasionalism that paved the way for the Regularity View of Causation. Occasionalism is the view that the only real cause of everything is God and that all causal talk which refers to finite substances is a sham. But Malebranche drew a distinction between real causes and natural causes (or occasions) and claimed that real causation involves a necessary connection between the cause and the effect. Since there is no perception of necessary connection in nature (but only in the case of the will of God and its effect), Malebranche concluded that there is no worldly causation: all there is in the world is regular sequences of events, which strictly speaking are not causal.

Hume (1739/ 1978) secularised the concept of causation by arguing that the appeals to the power of God to cause things to happen are worthless since such claims give us "no insight into the nature of this power or connection" (p. 249). When a sequence of events that is considered causal is observed, e.g., two billiard balls hitting each other and flying apart, there are impressions of the two balls, of their motions, of their collision and of their flying apart, but there is *no* impression of any alleged necessity by which the cause brings about the effect. For Hume, there is no distinction between "cause and occasion" (p. 171). Hence, causation is regular succession: one thing invariably following another. His famous *first* definition of causation runs as follows: "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 precedency and contiguity to those objects, that resemble the latter'" (p. 170).

Yet, Hume faced a puzzle. According to his empiricist theory of ideas, there are no ideas in the mind unless there were prior impressions (perceptions). He did, however, recognise that the concept of causation involved the idea of *necessary connection*. Where does this idea come from, if there is no perception of necessity in causal sequences? Hume argued that the *source* of this idea is the perception of "a new relation betwixt cause and effect": a "constant conjunction" such that "like objects have always been plac'd in like relations of contiguity and succession" (p. 88). The perception of this constant conjunction leads the mind to form a certain habit or custom: to make a "customary transition" from cause to effect. It is this felt determination of the mind which affords us the idea of necessity.

So instead of ascribing the idea of necessity to a feature of the natural world, Hume took it to arise from *within* the human mind, when the latter is conditioned by the observation of a regularity in nature to form an expectation of the effect, when the cause is present. Indeed, Hume offered a *second* definition of causation: "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" (p. 170). Hume thought that he had unpacked the "essence of necessity": it "is something that exists in the mind, not in the objects" (p. 165). He claimed that the supposed objective necessity in nature is *spread* by the mind onto the world. Hume can be seen as offering an objective theory of causation in the world (since causation amounts to regular succession), which was however accompanied by a mind-dependent view of necessity. This dual aspect of Hume's account of causation is reflected in his two definitions.

The Regularity View of Causation has been espoused by many eminent philosophers and has been the official *Humean* view. But it does face a number of difficulties. Hume might have left the metaphysics of causation behind, but he ended up with a *loose* notion of causation. On the one hand, it seems that there can be causation without regularity. This is the case of the so-called *singular* causation (the driving force behind the intrinsic-relation intuition noted above), where one event causes another to happen *without* this particular (singular) sequence of events falling under a regularity. On the other hand, there can be regularity without causation. There are cases in which events regularly follow each other (like the night always follows the day) without being the cause of each other. Indeed, there are too many regularities in nature and not all of them are, intuitively, causal. Hume might have prioritised the epistemological issues of causation, but, critics argue, it does not follow from this that, metaphysically speaking, causation consists in regular sequence. Note that according to the Regularity View of Causation, whether or not a particular sequence of events (*this* billiard ball moving *that* billiard ball after colliding with it) is causal depends on things that happen elsewhere and elsewhen in the universe, and in particular on whether or not this particular sequence instantiates a regularity. This is quite unusual, critics argue. To say that '*c* causes *e*' has the same structure as to say that '*x* loves *y*'; yet, it would be absurd to say that whether or not 'Mary loves John' is true depends on anything other than Mary, John and their properties and relations.

Based on thoughts like this, an opposite theory that became prominent in the twentieth century, due mostly to the work of Curt John Ducasse, is that what makes a particular sequence of events causal is something that happens there and then: a local tie between the cause and the effect, or an intrinsic feature of the particular sequence. According to Ducasse's (1969) *single-difference* account, an event *c* causes an event *e* if and only if *c* was the last – or, the only – difference in *e*'s environment before *e* occurred. This view takes causation to link individual events independently of any regular association that there may, or may not, be between events like the cause and events like the effect. Causation, non-Humeans argue, is essentially singular: a matter of *this* causing *that*. The epistemology, however, is not on the side of the non-Humeans. For if causation is a singular relation, how can it be known? Wouldn't its knowledge require some special non-inductive method? In any case, singularists seem unable to offer a good mark of causation, simply because the very notion of 'last' and 'only' change in the effect's environment before it happened are vague and obscure.[3]

Some Humeans, (most notably John Stuart Mill and John L Mackie), advanced more sophisticated versions of the Regularity View of Causation. For Mill, regular association is not, on its own, enough for causality. A regular association of events is causal only if it is "unconditional," that is only if its occurrence does not depend on the presence of further factors which are such that, given their presence, the effect would occur even if its putative cause was not present. A clear case in which unconditionality fails is when the

[3] For more on this see my (2002, chapter 2).

events that are invariably conjoined are effects of a common cause. Ultimately, Mill took to be causal those invariable successions that constitute laws of nature.

Among the more recent attempts to develop more defensible versions of the *Regularity View of Causation* Mackie's *inus*-conditions approach stands out. Mackie (1974) stressed that effects have, typically, a plurality of causes. That is, a certain effect can be brought about by a number of distinct clusters of factors. Each cluster is sufficient to bring about the effect, but none of them is necessary. So, he takes the regularities in nature to have a complex form (*A&B&C* or *D&E&F* or *G&H&I*) ? *E*, which should be read as: all (*A&B&C* or *D&E&F* or *G&H&I*) are followed by *E*, and all *E* are preceded by (*A&B&C* or *D&E&F* or *G&H&I*). How do we pick out the cause of an event in this setting? Each single factor of *A&B&C* (e.g., *A*) is related to the effect *E* in an important way. It is an *insufficient* but *non-redundant* part of an *unnecessary* but *sufficient* condition for *E*. Using the first letters of the italicised words, Mackie has called such a factor an *inus* condition. Causes, then, are *inus* conditions. So to say that short circuits cause house fires is to say that the short circuit is *inus* condition for house fires. It is an insufficient part because it cannot cause the fire on its own (other conditions such as oxygen, inflammable material etc. should be present). It is, nonetheless, a non-redundant part because, without it, the rest of the conditions are not sufficient for the fire. It is just a part, and not the whole, of a sufficient condition (which includes oxygen, the presence of inflammable material etc.), but this whole sufficient condition is not necessary, since some other cluster of conditions, e.g., an arsonist with petrol etc. can produce the fire. One attraction of Mackie's proposal is that we can see how we can engage in causal inference. Take a regularity of the form *AX* or *Y* ? *E*. Suppose that an instance of *E* has occurred. If we also happen to know that, in this particular instance, *Y* has not occurred, we can infer that *AX* has occurred.

Counterfactual Dependence

In *An Enquiry Concerning Human Understanding* (1748/ 1974) Hume stated briefly that an object is the cause of another when "if the first object had not been, the second never had existed". He never pursued this thought, since he wrongly took it to be a version of his first definition of causation. But it is not. It defines causation not in terms of actual regularities, but in terms of a counterfactual dependence of the effect on the cause: the cause is rendered counterfactually necessary for the effect. What have been called counterfactual conditionals (conditionals of the form: if *p* hadn't happened, then *q* wouldn't have happened, or if *p* had happened, then *q* would have happened) have caused many sleepless nights to philosophers because providing semantics for them has proved to be a very tough problem. The truth-conditions of such conditionals cannot be specified by means of truth-tables (and in particular by the truth-table of the material implication), because their antecedent is false (given that *p* did or did not actually happen), and hence the counterfactual conditional would end up being trivially true. Yet, there is some quite intimate connection between causation and counterfactual conditionals. We after all say things like: the short-circuit caused the fire since *if the short-circuit hadn't happened, the fire wouldn't have ensued*. The use of counterfactuals seems to capture the difference platitude: causes make a difference to their effects – if the cause had not happened, the effect would not have happened either – just as Hume suggested. Besides, it seems that this notion of the counterfactual dependence of the effect on the cause can capture the claim that effects depend robustly on their causes in such a way that a) there is no need to think of causation as a productive relation and b) we can think of it as a singular relation: a matter of *this* causing *that*.

In the twentieth century, the counterfactual approach was developed in great detail and elegance by David Lewis (1986). He defined causation by reference to a *causal chain* of counterfactually dependent events, where a sequence of events <*c*, *e*, *e'*, ...> is a chain of counterfactual dependence if and only if *e* counterfactually depends on *c*, *e'* counterfactually depends on *e* and so on. This move is meant to enforce that causation is a transitive relation among events (that is, if *c* causes *e* and *e* causes *e'*, then *c* causes *e'*). So one event is a cause of another if and only if there exists a causal chain – consisting, perhaps, of many discrete events – leading from the one to the other. Lewis articulated a rather complicated logic of counterfactual conditionals, which was based on the idea that, besides the actual world, there are also other possible worlds, that are more or less similar to the actual. A chief criterion for judging the similarity among worlds was taken to be whether the same laws of nature govern the worlds under comparison. One of course need not take this talk of other possible worlds literally. One might think of this talk as convenient *facon de parler*. But Lewis was a modal realist. He thought there are other possible worlds as real as the actual and are ranked in terms of a similarity metric. To many, Lewis's reification of possible worlds inflates our metaphysical commitments unnecessarily, and therefore, it is otiose. Even if it succeeds in offering semantics for counterfactual conditionals, it is nowhere near common sense and the practice of science.

Be that as it may, the promise of the counterfactual approach is to analyse causation as a singular relation among discrete events; hence, it downplays the role of regularities (or laws) in causation. But regularities do enter the counterfactual approach in a roundabout way: as means to capture the conditions under which counterfactual assertions are true. A chief difficulty for the counterfactual account comes from cases of causal overdetermination, where there are two factors each of which is sufficient to bring about the effect, but none of them is necessary, since even if the one was not present, the other factor would ensure the occurrence of the effect. For instance, both Billy and Suzy throw two rocks simultaneously at a bottle and they shatter it. They both caused the shattering, but the effect is not counterfactually dependent on either of them, since if Billy's rock had missed the bottle, Suzy's would have still shattered it. So there is causation without the cause being counterfactually dependent on the effect. (Clearly, the converse is also true: there are cases of counterfactual dependence that are not cases of causal dependence, as, for instance, when I say that 'if I were to write down 'Psillos', I would have to write down two 'l's'.) Even harder difficulties stem from cases of late pre-emption, as, for instance, when both Billy and Suzy throw rocks at a bottle, Suzy's rock shatters the bottle, but Billy's gets where the bottle used to be *after* it was smashed. Cases such as this show that a) causation does not involve counterfactual dependence and b) an appeal to causal chains cannot restore the link between causation and counterfactual dependence.

Lewis and his followers have made ingenious and lengthy efforts to render the counterfactual theory of causation foolproof. But all this is at the price of making the original, intuitively very plausible, account of causation very complicated and counter-intuitive. To stress a point made by Armstrong (1999, 181), the counterfactual theory might be salvaged, but only at the price of becoming *ad hoc* and circular. For the claim that *c* causes *e* does not imply that if *c* hadn't occurred, then *e* wouldn't have occurred. If anything, it implies the following: if *c* hadn't occurred, then *e* wouldn't have occurred, *unless e* was *causally* overdetermined, or *e* came to exist *uncaused* etc. Then, it's clear that analysing causation in terms of counterfactuals would either have to eliminate the unless-clause in an ad hoc way, or (inclusively) to appeal to the causal notions involved in the unless-clause.

All this is not to imply that there is no connection between causation and counterfactuals. The difference platitude does get captured by some notion of counterfactual dependence.[4] But the above does imply that there is not a straightforward and foolproof analysis of causation in terms of counterfactuals. Many philosophers have tried to find the resources for an analysis of causation elsewhere, and have revived the view that causation is a productive-mechanistic condition.

PRODUCTION THEORIES

Transference Accounts

There is a sense in which the view that causation is a productive relation goes back to Aristotle, who took causation to be efficient, that is to produce things. As is well-known, Aristotle distinguished between four types of causes. The material cause is "the constituent from which something comes to be"; the formal cause is "the formula of its essence"' the efficient cause is "the source of the first principle of change or rest"; and the final cause is "that for the sake of which" something happens (194b23-195a3). For instance, the material cause of a statue is its material (e.g., bronze); its formal cause is its form or shape; its efficient cause is its maker; and its final cause is the purpose for which the statue was made. Aristotle thought that, *ceteris paribus*, a complete causal explanation has to cite all four causes: the efficient cause is the active agent that puts the form on matter for a purpose. The four causes do not explain the same *feature* of the object (e.g., the material cause of the statue – bronze – explains why it is solid, while its formal cause explains why it is only a bust), yet they all contribute to the explanation of the features of the very same object.

Most modern philosophers revolted against all but efficient causation. The latter was taken to be the *only* type of causation by (almost) all those who advocated, in one form or another, the mechanical philosophy. Where Aristotle saw goals and purposes in nature, mechanical philosophers either excised all purpose (and all final causation) from nature (Hobbes) or placed it firmly in the hands of God (Descartes). It was mainly Gottfried Leibniz who tried to reconcile efficient (mechanical) causation with final causation. In a sense, Hume was the first to remove the efficiency from efficient causation: causation just is regular succession – one thing following another.

Descartes (1984) articulated what can be called the *transference model* of causation: when x causes y, a property of x is communicated to y. He thought this view is an obvious consequence of the principle "Nothing comes from nothing". As he put it, "For if we admit that there is something in the effect that was not previously present in the cause, we shall also have to admit that this something was produced by nothing" (p. 97).

But Descartes failed to explain how this communication was possible. Indeed, by taking matter to be an inert extended substance, he had to retreat to some external cause of motion and change. Descartes

[4] Another way to capture the *difference* platitude has given rise to what has been called probabilistic causation which is based on the thought that causes raise the *chances* of their effects, viz., the probability that a certain event happens is higher if we take into account its cause than if we don't. There are various theories of probabilistic causation, but we shall not discuss them here. Suffice it to say that, in broad outline, c causes e iff (i) the probability of e given c is greater than the probability of e given not-c and (ii) there is no other factor c' such that the probability of e given *c and c'* is equal to the probability of e given *not-c and c'*.

treated forces with suspicion, since they did not quite fit within his tight scheme of the two distinct substances (mind and body) and their two essential attributes (thought and extension). So in his *Principles of Philosophy* (1644) he retreated to God, whom he took to be the efficient cause of everything. The transference model was severely criticised by Leibniz (he called it the "real influx" model) on the basis of the claim that no impetus or qualities are transferred from one body to another (or between matter and mind) but that instead, each body is moved by an innate force. For Leibniz causes are required to explain why objects exist and change and they do this by providing a reason for this existence and change. But the reason (and hence the cause) is to be found in the primitive active force of each body, viz., in its power of acting and be acted upon.

This endowment of objects with active (and passive) powers became quite the standard way to account for the productivity of causation – without implying that anything gets transferred from the cause to the effect; it's just that causation is productive because the cause has the power to produce the effect. But Hume criticised relentlessly the view that causation can be understood in terms of the powers of causes to bring about their effects, claiming that this casts no light on the supposed productive relation between cause and effect – it merely redefines causation.

More recently, transference models have become quite popular again by being tied to physical properties such as energy-momentum. They have given rise to mechanistic theories of causation according to which there is a mechanism that connects cause and effect. According to Wesley Salmon's (1984) *mechanistic* approach, an event *c* causes an event *e* if and only if there is a causal process that connects *c* and *e*. A causal process is characterised by causal unity, e.g., the persistence of a quality or the possession of some characteristic. For Salmon, causal processes are the fundamental elements of the mechanistic approach to causation: they constitute the mechanisms that link cause and effect and transmit causal influence. Borrowing an idea of Hans Reichenbach's (1956), Salmon characterised 'causal' those processes that are capable of transmitting a mark, where a mark is a modification of the structure of a process by means of a single local interaction.

However, Salmon's view was shown not to be sufficiently distinct from the counterfactual approach to causation, since it had to rely on counterfactual conditionals to specify what processes are markable and hence causal. Later on, Salmon took causation to consist in the exchange or transfer of some *conserved quantity*, such as energy-momentum or charge. But according to Phil Dowe (2000) the notion of transference is not clear enough. On his conserved quantity theory of causation, a process is causal if and only if it *possesses* a conserved quantity. Similarly, an interaction between two processes is causal if and only if there is an exchange of a conserved quantity. Conserved quantities are those whose behaviour is governed by conservation laws. Examples of conserved quantities are mass-energy, linear momentum and charge. Advocates of this view argue that it is an empirical matter, open to scientific investigation, what conserved quantities there are in the world. They insist that this theory offers an account of singular causation, since though laws are required to identify the conserved quantities, the causal process need not instantiate a regularity. This theory of causation can be seen as a species of the mechanistic approach. The possession and exchange of conserved quantities is supposed to be the local tie that connects cause and effect. Though this theory seems plausible when it comes to physical causation, it is questionable whether it can be generalised to cover all cases of causation and especially the cases in the special sciences (economics, psychology etc.).

Power-based Accounts

Power-based accounts of causation of the sort Leibniz envisaged have become quite popular in the last decade or so. A lot of contemporary philosophers think that causation should be best understood on the basis of causal powers, that is powers, dispositions and capacities things have to cause other things to happen. This kind of move has been part of the resurgence of Aristotelianism in the philosophy of science. With it came the fact the essentialism became once more a respectable approach. Essentialism is the view that there is a sharp distinction between essential and accidental properties; an object is what it is in virtue of its essential properties, which it holds necessarily. An interesting twist that has been given to this view, defended recently by Brian Ellis (2001), is to take natural kinds (or natural properties) to have dispositional essences, that is causal powers that they possess *essentially* and in virtue of which they are disposed to behave in certain ways. So, for instance, water has essentially the causal power to dissolve salt and it is in virtue of *this* power that it does dissolve salt and that it is a *necessary truth* that water dissolves salt.

The key idea, viz., that properties are powers was re-introduced by Rom Harré and Edward H. Madden (1975) and has been strengthened by Sidney Shoemaker (1984). Harré and Madden drew a distinction between Aristotelian individuals and Parmenidean ones. Aristotelian individuals have variable powers (that is, powers that can change, fade away, die out etc.). This variability was said to be grounded in the natures of these individuals: their nature can remain intact and yet their powers may change. This nature was understood as the atomic structure of the Aristotelian individual. Parmenidean individuals, on the other hand, have constant powers and this constancy is constitutive of their nature. A Parmenidean individual cannot change its powers and remain the same individual: the powers and the nature of Parmenidean individuals are the same. Harré and Madden's example of Parmenidean individuals were the elementary particles, such as the electron and its constant power of negative charge. Shoemaker went further by suggesting that all properties are best understood as powers since the only way to identify them is via their causal role (the things they can do). Accordingly, two seemingly distinct properties that have exactly the same powers are, in fact, one and the same property. And similarly, one cannot ascribe different powers to a property without changing this property. It's a short step from these thoughts that properties are not freely recombinable: there cannot be worlds in which two properties are combined by a different law than the one that unites them in the actual world. Actually, on the view currently presented, it does not even make sense to say that properties are united by laws. Rather, properties – qua powers – *ground* the laws.

Many philosophers, however, find these views unappealing, not least because they fail to explain the fundamental notion of causal power. It has been argued that the positing of power is, in Mackie's (1977) memorable phrase, "the product of metaphysical double-vision" (p. 366). Far from explaining the causal character of certain processes (e.g., the dissolution of a sugar-cube in water), "they just *are* the causal processes which they are supposed to explain seen over again as somehow latent in the things that enter into these processes" (ibid.).

CAUSATION AND MANIPULATION

The idea that causes are "recipes" goes back to Douglas Gasking (1955). He took it that "a statement about the cause of something is very closely connected with a recipe for producing it or for preventing it"

(p. 483) and claimed this idea, though not exhaustive of causation, captures causation in most ordinary cases and explains it by reference to "producing-by-means-of" relation: performing the manipulation of an object we produce in it a change which results in a(nother) change in it or in a different object. For instance, we make a piece of iron glow (that is, we produce the glowing) by heating it. Now, that's clearly a *loose* sense of production since all it involves is a rather stable dependence of the effect on the cause. But it does bring out something important, viz., that part of the usefulness of causation – and part of how we come to have the concept of causation – as to do with the fact that we can do things with or to causes as means to bring about effects.

Von Wright (1973) elaborated this idea by tying causation to the concept of human agency since manipulation *is* a distinctively human action. This way of putting things might sound too anthropomorphic. Don't we think there would be causal relations, even if there wouldn't be any humans around capable to manipulate the causes? Von Wright was careful to note that the dependence of causation on human action is "epistemological rather than ontological" (p. 17). It concerns how causal claims are established and not what causation is. He was quite clear that "causation [...] operates throughout nature independently of agency, also in regions of the world inaccessible to human interference" (ibid.). Yet, he thought, the connection between causation and human action is also "logical", in the sense that the concept of causation "is connected with features which are peculiar to the concept of action" (ibid.). So, von Wright's main point was that the very concept of causation is modelled on the concept of human action, where someone acts freely to bring about something.

This might well be the case. We might come to possess the concept of causation because we can make things happen – though, I take it that how exactly we might come to possess the concept of causation is an empirical issue. But what can we learn about causation as it is in the world from this? Not much. It is one thing to say that we can know what causes what (at least in favourable circumstances) by manipulating the putative cause and bringing about the effect; it is quite another thing to say that what confers on a sequence of events the character of causal connection is the possibility of subjecting causes to experimental test by interfering with the 'natural' course of events. To say the least, this would not explain how we can meaningfully talk of causation when there is no possibility of experimental manipulation (as in the case of an earthquake, or as in the case of things that happen in remote regions of space-time).

Recently, there have been important attempts to give a more objective gloss to the idea of manipulation. Following the pioneering work of the computer scientist Judea Pearl, James Woodward (2003) has advanced a notion of intervention that is not restricted to human action. He has argued that a relationship among some magnitudes X and Y is causal if, were one to intervene to change the value of X appropriately, the relationship between X and Y would remain invariant *but* the value of Y would change, as a result of the intervention on X. According to Woodward, causation is based on counterfactual manipulation. His theory is *counterfactual* in the following sense: what matters is what *would* happen to a relationship, *were* interventions to be carried out. Interventions need not be actual. They can be hypothetical or counterfactual. And invariance is not understood in terms of stability under actual interventions. The causal relationship (generalisation) should be invariant under hypothetical or counterfactual interventions. To use a stock example, the force exerted on a spring *causes* a change of its length, because if an intervention changed the force exerted on the spring, the length of the spring would change too (but the relationship between the two magnitudes – expressed by Hooke's law – would remain invariant, within a certain range of interventions).

There is no space here to look into the details of the theory.[5] What should be noted is that though Woodward's theory rightly has been seen as a species of the counterfactual approach discussed above, the interventionist theory is not concerned with any kind of counterfactual conditionals whatever but rather with what Woodward has called 'active' or 'experimental' counterfactuals' viz., those that are related to *interventions* which can make the antecedent of the counterfactual conditional true. This is certainly a step in the right direction since Woodward's theory does not rely on an abstract metaphysical theory for the truth of causal claims but it is mostly interested in the epistemological and methodological circumstances under which causal facts can be ascertained.

The drawback is that there seems to be more to causation than invariance-under-actual-and-counterfactual-experimental-intereventions. Consider the true causal claim – discussed by Woodward (2003, 127-33): changes in the position of the moon with respect to the earth and corresponding changes in the gravitational attraction exerted by the moon on the earth's surface cause changes in the motion of the tides. As Woodward adamantly admits, this claim cannot be said to be true on the basis of interventionist (experimental) counterfactuals, simply because realising the antecedent of the relevant counterfactual is physically impossible. His response to this is an alternative way for assessing counterfactuals. This is that counterfactuals can be meaningful if there is some "basis for assessing the truth of counterfactual claims concerning what would happen if various interventions were to occur." Then, he adds, "it doesn't matter that it may not be physically possible for those interventions to occur" (p. 130). But then, it's no longer clear that we have been given a theory of causation with a distinctive reference to experimental and active counterfactuals. The latter, though epistemically useful, cannot exhaust the scope of causation.

In any case, the theory we have ended up with is not sufficiently different from the standard counterfactual approach. More generally, it seems that features such as manipulability and invariance under interventions are *symptoms* of a causal relation, but not constitutive of it.

CONCLUSION

What then is the answer to the title question? We do not quite know. There is no single account of causation – no theory of what causation *is* – that is free of counterexamples. Nor is there any theory of causation that tallies best with all our intuitions about what causes what. The persistent failure to find a fully adequate philosophical theory of causation may well make us sceptical about the prospects of such a theory. Perhaps, we are looking for unity where there is plurality – for an analysis of a single concept, where there are many.

In the meantime, we can carry on doing business as usual. We can find out a lot about what causes what without having a philosophical theory of causation; or we can look for common factors (or elements) in the various theories which can help us ground causal claims (and causal knowledge) as whatever is licensed by all of the various philosophical theories. We seem to be lucky enough to live in a world in which there are many ways to identify the presence of a causal connection – that is to ascertain that *c* causes *e*. In most ordinary cases, e.g., when we say that the ball broke the window or that aspirin causes headache relief or that smoking causes lung cancer, but in more scientific cases as well, e.g., when we say that tides are caused by the moon's attraction or that increases of unemployment rates cause a rise of the crime rates, it seems that whenever *c* causes *e*, it is the case that: there is a regularity (deterministic or

[5] For a fuller discussion see my (2004).

statistical) under which *c* and *e* fall *and* there is a relation of counterfactual dependence between *e* and *c* *and* there is some causal process (mechanism) that connects *c* and *e* and there is something (like energy) that gets transferred from *c* to *e*. In fact, it is because there are many ways to identify a causal fact that there is, typically, agreement about what causes what, even if there might be (philosophical) disagreement about what causation consists in.

The platitudes of causation suggest that there is something – what we call causation – that can do a lot of serious work for us (make a difference, be used as a recipe, explain and predict), even though we do not quite know what exactly this something is; nor whether it is a single something.

Appendix

In his "The Notion of Cause" (1912) Bertrand Russell argued that the concept of causation was incoherent. But this was just as well for him, since, as he claimed, physics has stopped looking for causes: for "there are no such things". Here is his famous dictum:

> *The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm.*

Russell's argument is simple but powerful. Take two events *c* and *e* such that *c* is said to cause *e*. Either they are *temporally contiguous* (*A*) or there is a *temporal gap* between them (*B*).

Take first option (*A*), and suppose that the two events *c* and *e* have a finite duration. So, *c* lasts for some time and then it ceases to exist, being followed immediately by the coming into existence of *e*. Then, there are two sub-options available.

> (*A1*): the cause is a dynamic process: the earlier parts of the cause will stand in causal relations to its later parts. But then, Russell notes, the earlier parts of the cause will be irrelevant to the effect since they are not temporally contiguous to it. Then, no part of the cause, no matter how temporally close to the effect might be, can be *the* cause of its effect. For no matter how much we diminish the duration of the cause, there always will be an earlier part that is not temporally contiguous with the effect. So, given (*A1*), it does not make sense to talk of *the* cause *c* of the effect *e*.
>
> (*A2*): the cause *c* of *e* is purely static: it involves no change within it. Against this sub-option, Russell argues that a) no such causes can be found in nature; and b) even if there were such causes, in order to give rise to their effect, they would then have to suddenly explode into the effect. Given the implausibility of sub-option (*A2*), Russell dismisses it altogether.

What about option (*B*) then? If there is a temporal gap between the cause *c* and the effect *e*, then, Russell argues, the cause is no longer sufficient for the effect. For, no matter how small the gap (temporal interval) might be, it's always possible that something else – call it *c*' – happens within this time-interval which *prevents e* from occurring. Given the failure of both options (*A*) and (*B*), Russell concludes that it is not really intelligible to talk of causes.

Russell's argument can be contested. For instance, one might notice that his point against (*A1*) relies on viewing time as infinitely divisible, that is, as being a real-line continuum. So, if one denies this assumption, then the argument against *A1* fails. Similarly, one may argue against Russell's critique of option (*B*) that to say that event-type *C* causes event-type *E* is not to say that there are no circumstances under which it doesn't. Be that as it may, Russell's suspicion of the concept of causation was inherited by the movement of Logical Positivism, which set the agenda for most of the philosophy of science in the twentieth century.

References

Armstrong, D. (1999). The open door: Counterfactual versus singularist theories of causation. In Howard Sankey (Ed.), *Causation and laws of nature* (175-85). Dordrecht: Kluwer.

Descartes, R. (1984). *The philosophical writings of Descartes*: *Volume II*. (J. Cottingham, R. Stootfoff & D. Murdoch, Trans.). Cambridge: Cambridge University Press.

Dowe, P. (2000). *Physical causation*. Cambridge: Cambridge University Press.

Ducasse, C. J. (1969). *Causation and types of necessity*, New York: Dover.

Ellis, B. D. (2001). *Scientific essentialism*. Cambridge: Cambridge University Press.

Gasking, D. (1955). Causation and recipes. *Mind, 64*, 479-87.

Hall, N. (2004). Two concepts of causation. In J. Collins, L. Paul and N. Hall (Eds.), *Counterfactuals and causation* (225-276). Cambridge: MIT Press.

Harré, R. & Madden, E. H. (1975). *Causal powers: A theory of natural necessity*. Oxford: Blackwell.

Hume, D. (1974). An enquiry concerning human understanding. In L. A. Selby-Bigge & P. H. Nidditch (Eds.), *Enquiries concerning human understanding and concerning the principles of morals.* Oxford: Clarendon Press.

Hume, D. (1978). *A treatise of human nature*. (L. A. Selby-Bigge & P. H. Nidditch, Eds.). Oxford: Clarendon Press. (Original work published 1739)

Lewis, D. (1986). Causation. In his *Philosophical Papers, Vol. II* (159-213). Oxford: Oxford University Press.

Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford: Clarendon Press.

Mackie, J. L (1977). Dispositions, grounds and causes. *Synthese, 34*, 361-70.

Psillos, S. (2002). *Causation and explanation*. Chesham, UK: Acumen.

Psillos, S. (2004). A glimpse of the secret connexion: Harmonising mechanisms with counterfactuals. *Perspectives on Science*, *12*, 288-319.

Reichenbach, H. (1956). *The direction of time*. Berkeley, CA: University of California Press.

Russell, B. (1918). On the notion of cause. In *Mysticism and Logic*. London: George Allen & Unwin.

Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.

Shoemaker, S. (1984). *Identity, cause, and mind*. Cambridge: Cambridge University Press.

von Wright, G. H. (1993). On the logic of the causal relations. In E. Sosa & M. Tooley (Eds.), *Causation*. Oxford: Oxford University Press.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.

❑ ❑ ❑

DISCUSSION

Chairperson: K.Subramaniam, Homi Bhabha Centre for Science Education

General comment by the speaker:

Before the discussion starts in earnest, I wanted, if I may, to make a very general comment. Since most of the participants in this conference are science educators, you might wonder whether the kind of enterprise I am engaged in, viz., philosophy of science, matters at all to science education. The short answer is: No. But this answer is too short. A longer answer would be much more interesting – and it would be, by and large, positive. If you think of it, philosophy of science matters to science education because it matters to science (not necessarily in a practical sense). It can

i. elucidate why we should value science while at the same time take a critical view of it;
ii. cast light on the nature of science – what, if anything, makes science distinctive as a cognitive enterprise?
iii. ground and explain the fertility and complexity of scientific method;
iv. explain the successes of science and its convergence to a stable scientific image of the world;
v. embed scientific theorising in a historical context, thereby showing continuities and breaks with the traditions;
vi. analyse the complex relations (or sometimes the absence of relations) between theory and experiment;
vii. reveal the idealised and abstract element in scientific theorising and explain how theories represent the world.

There is understanding in science, an area in which science education is focused on. But there is also understanding science (that is, understanding what it is we are doing and why we are doing it when we do science, at least from the cognitive point of view). This is a characteristically philosophical enterprise – even though it is not exclusively the domain of professional philosophers of science. More specifically understanding causation matters to explanation and understanding – and science is all about explanation and understanding. Understanding causation – better: coming to terms with some basic facts about the plurality of the ways in which causation reveals itself and can be known – can become a tool in explaining how we can come to know the causal structure of the world and how this causal structure can be modelled in different ways and can reveal itself in many and variegated ways.

Comment 1: Just a comment and perhaps the point where you reached last, is where I am taking off from. I think it's very important when we deal with children to realize that all of us are forced into a philosopher's role when a child asks, "What causes rain?" As an adult, a parent, a teacher or an educator we have to find an answer to what causes rain. To me it seems like, "Clouds cause rain" is not a good enough answer. And while it may meet most of the classical test of causation – clouds followed by rain, there's no rain without clouds, and so on, one would say that it's not an adequate explanation for rain. The point where you reached last, is really the point where most of us find ourselves engaged really.

Response: To reinforce what you said, that's a good way to highlight the fact that in most cases – even apparently simple and straightforward ones – the actual causal story is much more complicated than it seems. It's important to actually tell the people (the students) what the causal structure is and that it is much more complex than it appears. I think it's important to make people aware that what causes what is not always a straightforward matter.

Q. 1. When it comes to higher mathematics, the various platitudes of causation, how to work with the paradoxical hypothesis etc; the manipulative structures of causation bring about a different scientific angle in mathematics education.

Ans. I won't say anything about the well-known paradoxes in the foundations of mathematics. When it comes to mathematics itself, there are two issues to worry about. One is related to teaching mathematics and in particular, to how to better teach mathematics. Here there is a lot of causal modeling and experimental work aiming to improve the students' ability to understand mathematics. But this is a general point – it concerns teaching any subject whatsoever. The other issue concerns mathematics itself. There is no causation in mathematics – there are no causal relations among mathematical entities. Mathematical entities like numbers or triangles are abstract entities outside space and time; they are causally inert; they do not cause anything. At least that's what is the case according to a very influential and long-standing philosophical tradition – that goes back to the Pythagoreans and Plato. The upside of this view is that it provided a clear semantics for mathematical language and especially a uniform semantics for mixed sentences that involve mathematical terms and physical terms. The downside is that it is then unclear how we get to know mathematical entities and mathematical truths – if, that is, we do not causally interact with them, as we causally interact with tables and protons. Some think that there is a special faculty – intuition – that engenders mathematical knowledge. But this is quite mysterious. In any case, if causation is irrelevant to the mathematical realm, it is not irrelevant to how mathematics is taught and understood.

Q. 2. Is it the difficulty of arriving at a consensus about causation theory that pushes some cause-effect relationships into the realm of superstition?

Ans. I think that's a fair point. There has been a lot of superstition concerning causation. And we should not forget that causation has been, at least partly, an anthropocentric (shall we say anthropomorphic?) concept – it seems to originate in the thought that the human will has the power to produce something. The regularity view of causation gained momentum in the seventeenth century (in the aftermath of the scientific revolution) because it promised to demystify causation – by removing the alleged efficiency of the cause and the alleged necessity by which it brings about the effect. It was based on a revolt against hidden powers. On the regularity view, there is no mystery in causation – causation, as it is in the world, is regular succession: one thing follows another in a regular fashion. If the regularity theory worked when it comes to *everything* we want to say about causation, we would go home and have a drink and we wouldn't have to think hard about causation anymore. But the problem is precisely that regularity theory doesn't seem to capture everything we think causation is and does. Hence, old ideas about powers and efficiency and necessity crop up again and again, even giving rise to claims that there is a bond, a tie, between cause and effect.

Q. 3. Is silence a cause or an effect?

Ans. I will pass on this. Actually I should have stayed silent! I don't know how to answer this.

Q. 4. What are the philosophical speculations about the beginning and the end of causation?

Ans. I am not sure how exactly to understand this question. I presume you are worried about there being an ultimate cause of everything. This, then, is a tough question. There are well-known arguments for the claim that causal sequences must stop somewhere, on pain of infinite regress. But then there are well-known models of infinite sequences, which have no beginning and no end. So it at least makes sense to say that a causal chain might not have a beginning. My own view is that explanations (but not necessarily causation) must stop somewhere and hence that at each stage of inquiry we have to admit some unexplained explainers – they explain everything else without being themselves explained.

Q. 5. I meant to say we don't attempt to deliberate or speculate on unknown things philosophically. Do we not attempt as philosophers – regarding the topic, whatever the question is, at an ontological level?

Ans. Kant thought that we go beyond the bounds of reason (and the limits of possible experience) when we try to answer limiting questions like the divisibility of space or the origin of the universe. I am not so sure. I think that if there are no limits to science, there are no limits to philosophy.

Q. 6. The theory of causation is limited to the regularities which are observable, e.g. Laws of nature. But Indian epistemology like in Buddhism talks about theory of nothingness to explain about creation of universal system where there is cause. Do you have any comments?

Ans. Causation is not just about observable regularities Clearly, there are non – observable regularities – say regularities in microphysics.

Comment 2: I know this is a very daunting task to put all the theories of causation together and I think you've done it very well, but particularly for students who are interested perhaps from here and elsewhere – I want to draw your attention to the fact that there are very mature theories of causation from other traditions which the philosophy of science community in the world seems to not take account of. Whether it's African philosophy or particularly Indian philosophy – since there was a question earlier –where the ideas of inherent cause are very central to Sankhya theories as well as other schools. Nyaya and Buddhist traditions had complex philosophical analysis of causation, a lot of which is relevant to contemporary debates in understanding causality. I am just adding to the richness of the accounts you have given, which might be of interest particularly to students who are trying to engage with this problem and especially in the context of science education because it adds a very different way of approaching the question of causation.

Response: I take the point. I presented the standard thinking about causation in the tradition I have been brought up in, but you are right that there are many different things to be said in different traditions and many different approaches and they are all respectable and should be studied carefully.

Q. 7. I want to get back to the question about causation in mathematics because you said very quickly there is none and I am not quite ready to swallow that. I was thinking for instance about implication in mathematics, how that might or might not be related to causation. I was thinking about children working on algebra problems and saying *this* is equal to *this* is equal to *this* is equal to 5. It certainly feels like there is some sort of a causal chain. In doing higher level mathematics when you are doing proofs you have a certain sort of logical structure and you come up with some sort of a sense that something is true and it certainly feels like the logical structure gives rise to that conviction. So I'm not quite willing to believe there is no causation in mathematics.

Ans. I'll tell you what I meant, and I think what I meant was right – at least *partly* right! Take Physics. There is causation in the physical world itself and there is, of course, causation in how educators teach people how to understand the physical world. Take mathematics. There is causation in how educators teach people how to understand mathematics, in how to reason about mathematics etc. But is there causation in the mathematical world itself? That's the question I wanted to highlight. To be precise, the answer one gives to it depends on one's philosophical theory of mathematics. There are quite a few philosophical theories of mathematics – that is, of what mathematical entities are. A Platonist tradition takes mathematical entities to be abstract entities outside space and time; as such, they are causally impotent; there is no causation in the platonic universe which numbers and sets inhabit. But there are also somewhat extreme naturalistic views of mathematics which declare that mathematical objects are *in* the natural world. On views such as this, when one sees two chairs, one literally sees the *set* of two chairs (that is a mathematical object) – and not just the two chairs (that is, two physical objects). In response to your question then, I am inclined to say that if you take a naturalistic view of mathematics, of what kinds of things mathematical objects are, then you can tell a causal story of the mathematical world, which is then just a part of the physical world.

Comment 3: With regard to relevance to science education, there were four phrases that you mentioned, which are quite relevant: Making a Difference, Recipe, Explain, and Understanding. That's the sequence that should be followed in science education.

Q. 8. There is an assumption that understanding the cause is a kind of science; in a sense that there is nothing more in natural science to look at. Is it a sufficient condition or is that the reason why you are so obsessed with the notion of causation and how much of it is the substance of science?

Ans. It seems clear to me that we do understand things around us by finding their causes. Of course, there can be non-causal explanation. For instance, mathematical explanation is non-causal explanation; and explanation of laws by subsuming them under more fundamental laws is non-causal explanation. So: not all explanations are causal. It is also clear to me that causal explanation should stop at some point – we should take some things as unexplained explainers. If I were a strict Aristotelian, my reply would be fully positive: there is understanding *only* when there is causal explanation, and causal explanation is demonstration from first principles that state the fundamental causal connections in the world. But I am not an Aristotelian; so I am not saying this. Still, citing causes is sufficient for explanation – but we should be careful to bear in mind that in most typical cases we just pick out a portion of the causal net that has led to an effect and which portion we pick out and highlight might depend on our interests.

Comment 4: I think that I can explain why people are puzzled about causation and mathematics because what they are trying to say is to explain human reasoning about mathematics. And so I could say that 2 plus 3 is 5 because... and this is the way I reason. I reason because I have two fingers and three fingers and I put together and I get five. So I think that causality in mathematical thinking is about human reasoning and not about mathematics per se.

Response: Thank you for this point. I think it really reinforces what I said above. When we talk about reasoning, we can think of it as a psychological process with a form, and that's clearly a causal description of it. But we can also think of it by reference to the abstract logical form (or pattern) of a mode of inference. This is not a causal description. For instance, that say *p* implies *p* or *q* is a valid inference irrespective of whether or not, in an actual case, thinking of p causes me to think (that is to move to) *p* or *q*. Now, mathematics is not about fingers, sticks and pencils – though the toughest question in the philosophy of mathematics is how and in virtue of what mathematics (which is not about the physical world) is marvelously applicable to the physical world. As Frege famously said, if everything was permanently pinned down, and nothing could be moved about, wouldn't we have arithmetic? In such a world, we couldn't naturally move the sticks next to each other and say 2 & 3 made 5 (and 3 & 2 makes 5). But arithmetic would be possible even in this strange world of immovable objects.