Causation

Max Kistler

Max Kistler

Causation

in Anouk Barberousse, Denis Bonnay et Mikaël Cozic (eds.), *Philosophy of Science. A Companion*, New York: Oxford University Press, 2018, p. 95-141.

In 1912, Bertrand Russell recommended that philosophers eliminate causation from their stock of concepts. His argument relied on the premise that advanced sciences do not contain any concept corresponding to the intuitive notion of causation. However, Russell also argues that the notion of causation cannot possibly be reduced in purely scientific terms either. Now, if there is a conflict between an intuition of common sense and science, the naturalist attitude consists in resolving the conflict by following science instead of intuition. Thus, concludes Russell, philosophers should stop speaking of "causes". The debate launched by Russell's article continues to this day. On the one hand, many philosophers argue along lines similar to Russell's that the notion of causation has no equivalent in fundamental physics. One way to understand to explain why causation plays such an important role in common sense without having any equivalent in physics is to interpret is as belonging to "folk science" (Norton 2003). However, the debate concerning the presence of causation in fundamental physics continues[1]. It is for example argued that the distinction between timelike and spacelike distances in special relativity expresses a causal distinction: a distinction between distances that can be bridged by signals, which can be interpreted as causal processes, and distances that cannot be so bridged. On the other hand, there is now much less confidence that it is possible to generalize from physics to all other sciences. To the extent that nothing guarantees the effective reduction of all sciences to fundamental physics, causation might well be and remain a legitimate and even indispensable concept in other sciences even if it is not in physics.

The plan of this chapter is as follows. In the first section we will analyze Russell's reasons for holding that there can be no analysis of the concept of causation that is compatible with 20th century physics. We will see that the debate between "eliminativists" following Russell and philosophers holding that the concept of causation is as central to science as it is to common sense is structured by two distinctions: between microscopic and macroscopic entities and between concrete events and their measurable properties. It turns out that the debate on the legitimacy of the concept of causation is linked to the debate on the existence of laws of nature outside fundamental physics, laws that allow for exceptions, often called *ceteris paribus* laws. We will see that, even if it were correct that causation plays no role in the theoretical content of fundamental physics, it may be argued that the concept of causation is nevertheless legitimate and useful in many contexts. It does seem to be central not only for common sense, e.g. in the context of our planning actions in light of their consequences, but also for all other sciences outside fundamental physics, such as biology and neuroscience, as well as for many projects involving the analysis of philosophical concepts in naturalistic terms. Thus, causation plays a central role in philosophical theories of intentionality, perception, knowledge and action.

After having thus justified the project of a philosophical analysis of the concept of causation, we shall examine the most important approaches that have been put forward and developed: in terms of counterfactual conditionals, in terms of probability raising, in terms of manipulability, and in terms of processes.

1 The central idea of the counterfactual analysis of causation is that, for any two events *c*

[1] See, e.g. the debate between Frisch (2009a; 2009b) and Norton (2009).

and *e* that have actually occurred, *c* causes *e* if and only if it is true that: if *c* had not occurred, *e* would not have occurred[2].

2 The central idea of the probabilistic analysis is that factor C exercises a causal influence on factor E if and only if an event of type C raises the probability of an event of type E.

3 The central idea of the manipulability analysis is that there is a causal relation between two variables C and E if and only if interventions modifying the value of C modify the value of E.

4 Finally, the central idea of the process analysis is that an event *c* causes another event *e* if and only if there is a physical process of transmission between *c* and *e*, e.g. of a quantity of energy.

One difficulty one faces in comparing these approaches stems from the fact that they conceive of the terms of the causal relation in different ways: for some analyses, causes and effects are singular events, whereas for others it is rather properties of events or "factors", which can be instantiated by numerous events.

Moreover, to understand the complex debate between the advocates of these theories, it is important to be conscious of the aims they pursue and of the criteria they use to judge their success. One can conceive of the task of a philosophical analysis of the concept of causation in at least two ways. Its aim can be taken to be 1) pure a priori analysis of the concept of causation as it is used by subjects, independently of the features of the actual world, as it is described by contemporary science, or 2) a partly empirical and partly conceptual enquiry on the "real essence" of causation, as it is in the actual world. According to this second interpretation of what it means to understand causation, causation is a natural kind of relation analogous to natural kinds of substances, such as water, gold or, for common sense, tigers. Common sense presupposes that such kinds of substances or animals possess a real essence that can be discovered by empirical science. In an analogous way, causation might have a real essence specific to our actual world. However, rather than beginning with these methodological reflections, we will take them up after having presented the debate on the counterfactual analysis: it is easier to think about the "metaphilosophical" question of the aim, method and criteria of adequacy of an analysis after having studied a sample of the debate.

1. Russell and the elimination of the concept of causation

Russell's arguments are mainly directed against what is now often called "generic causation"[3]. Singular causal judgments, such as "the fact that I have rubbed this match (i.e. the match that I see before my eyes) is the cause of the fact that it has lit", differ from generic causal judgments, such as: "in general, rubbing matches causes them to light". In Hume's conception[4], the truth of a singular causal proposition depends on the truth of a generic causal proposition. The truth of the proposition that the singular event *c* causes the singular event *e* presupposes the truth of the generic proposition that events of the same type as *c* are followed by events of the same type as *e*. In other words, there can be no causation between *singular* events without an appropriate regularity at the level of *types* of events. We will see later that this thesis has been challenged, so as to dissociate singular causation from generic causation. If the existence of singular causal relations does not presuppose the existence of generic

2 Lower-case variables represent concrete particular events; upper-case variables represent properties of objects or events.

3 For a recent reevaluation of Russell's arguments against the possibility of constructing a concept of causality compatible with contemporary science, see Price and Corry (2007); Spurrett and Ross (2007).

4 Many contemporary approaches to causation are deeply influenced by David Hume's (1739-40; 1777) conception of causation.

causal relations they instantiate, singular causation is no target for Russell's arguments. However, only a minority of contemporary analyses take singular causation to be independent of nomological relations at the level of types of events, factors or properties. To the extent that philosophical analyses of causation aim at explaining and justifying the use of causal concepts in science, the generic concept remains the most relevant: it is generally taken for granted that the fact that this match has lit at time t can only be explained in terms of general propositions that apply to rubbings of matches at any place and at any time. Such an explanation might mention the general proposition that the energy produced in the form of heat by sufficiently strong rubbing triggers the chemical reaction of exothermic oxidation of any sample of phosphorus sesquisulfide (P_4S_3), which happens to be the substance that covers the head of ordinary matches.

1.1. The principle of causality and the repetition of events

Russell tries to establish the vacuity of the traditional "principle of causality" according to which "the same causes always have the same effects", or more precisely: "Given any event e_1, there is an event e_2 and a time-interval τ such that, whenever e_1 occurs, e_2 follows after an interval τ." (Russell 1912/1992, p. 195). This is a "meta-law", stating that there are laws of succession involving types of events. Russell argues against the existence of such laws of succession – and thus against the principle of causality – in advanced sciences, by noting first that there can be recurring types of events only if these types are conceived 1) vaguely and 2) narrowly; and secondly that vaguely conceived events cannot be the target of scientific explanations whereas generalizations bearing on narrowly conceived events are not strictly true.

1) Events that recur are conceived *vaguely*: to use Russell's own example, throwings of bricks – followed by breakings of windows – recur only if they are conceived in a way that abstracts away from microscopic details. There are no two throwings that resemble each other exactly in all microscopic details. The problem is that scientific explanation in its mature form requires one to be able to *deduce* the *explanandum* from the description of the situation playing the role of the *explanans*, together with statements of the laws of nature (see chapter ??? of this volume). Now, such a deduction is possible only if first the *explanans* contains a quantitative description of the cause, i.e. is conceived "precisely" (*ibid.*, p. 200), second the laws of nature are also quantitatively precise, and third the *explanandum* is a quantitatively precise description of the effect. However, to the extent that events are conceived in this quantitatively precise manner – which is what makes their scientific explanation possible – they do not recur. To the extent that the antecedent of a universal conditional applies only to one event, the truth of the conditional is almost trivial: it is true if and only if its consequent is true in the unique situation in which its antecedent is true. Such a statement cannot be used to explain other events, which is a major function of laws. There cannot be strict laws containing quantitatively precise predicates that can be used for the explanation and prediction of new situations; there is room for strict regularities only in common sense and "in the infancy of a science" (*ibid*, p. 201).

2) Events that recur are *narrowly* conceived. There can only be recurring events if they are conceived locally, i.e. as the content of a well-delimited region of space-time. There are many rubbings of matches of the same type only to the extent that the circumstances are not included in the rubbing events. However, to the extent that one abstracts away from the person who rubs, the weather and other contextual factors, the regular lighting of matches when rubbed has exceptions: there may be factors present in the surroundings of the first event (the rubbing) that prevent the second event (the lighting) from occurring; in other words, the regularity exists only insofar as "all other things are equal", or *ceteris paribus*. The

dialectic is similar to the case of vagueness: it is possible that a narrowly conceived event recurs, but insofar as the circumstances of the events are not taken into account, the regularity with which event *c* is followed by event *e* is not exceptionless because factors in the circumstances may interfere and prevent *e* from occurring even though *c* has occurred.

Generalizations bearing on narrowly conceived events cannot be used in scientific explanations because that requires strictly true universal propositions. "The sequence […] is no more than probable, whereas the relation of cause and effect was supposed to be necessary." (*ibid*., p. 201)[5] On the other hand, to the extent that the possibility of interference by factors present in the spatio-temporal vicinity of the antecedent event is diminished by including the surroundings of the events, the probability of their recurrence diminishes. "As soon as we include the environment, the probability of repetition is diminished, until at last, when the whole environment is included, the probability of repetition becomes almost *nil*." (*ibid*., p. 197)

Note that the first argument against the principle of causation questions only the existence of successions of *macroscopic* events conceived with *common sense* concepts: microscopic events, such as the interaction of an electron and a photon or the radioactive decomposition of an uranium-238 nucleus, recur even if they are precisely conceived. However, the second argument questions the strict recurrence of both microscopic and macroscopic events: if one considers a set of localized cause-events that are of strictly the same type but does not take their surroundings into consideration, such cause-events are not necessarily followed be the same effect-events, because these effects can be influenced by events occurring in the neighborhood of the cause-events.

Thus, Russell's conclusion also covers microscopic events: "As soon as the antecedents have been given sufficiently fully to enable the consequent to be calculated with some exactitude, the antecedents have become so complicated that it is very unlikely they will ever recur." (*ibid*., p. 198). In sum, there are no macroscopic events that are both precisely conceived and recur; microscopic events may recur even when they are precisely conceived; however, the succession of microscopic events only recurs to the extent that the events are conceived locally, without taking their surroundings into consideration. Thus, the principle of causality "same cause, same effect" is according to Russell, "utterly otiose" (*ibid.*, p. 198), to the extent that what would allow for repetition ("same cause"), i.e. conceiving of macroscopic events vaguely, or including spatio-temporal surroundings for microscopic events, either makes them inappropriate for being used in the exact sciences (for the former) or prevents them from recurring (for the latter).

1.2. The functional laws of mature science

Russell's second argument against the possibility of finding scientific legitimacy for the notion of cause consists in showing that the laws that are used in the explanations of mature sciences cannot be interpreted as causal laws. The laws used in mathematical physics, e.g. in "gravitational astronomy" (*ibid*, p. 193), have the form of functions: in a system of masses subject only to the force of gravitational attraction, it is possible to represent the configuration[6] of the system at a given moment as a function of that moment and of the configuration and speeds at some other moment (or as a function of the configurations at two other moments). Although it is true that such a function "determines" the configuration of the system, this does not justify the idea that this determination is *causal*. Russell has two reasons for holding that "in the motions of mutually gravitating bodies, there is nothing that can be

5 Russell does not consider probabilistic causation because he takes necessitation to be a defining condition of causality.

6 The configuration of a system is the set of the positions and speeds of each of its components.

called a cause, and nothing that can be called an effect" (*ibid*., p. 202).

The first is that this determination is purely logical and indifferent to the direction of time: Newton's laws, together with the law of gravitational attraction, make it possible to calculate the configuration of a system of masses at some time in the past as a function of its configuration at some future time, in exactly the same way in which they make it possible to calculate the characteristics of the system at some future time on the basis of its characteristics at some moment in the past. Given that the traditional concept of causation requires that the cause precedes the effect[7], this functional determination cannot be interpreted as being causal.

The second reason concerns the terms of the relations: causality relates particular, or concrete events, whereas functional equations relate values of measurable quantities. In other words functional equations relate *properties* of concrete events rather than events themselves. The equation expressing the law of gravitation – or law of universal attraction – indicates the value of the force of gravitational attraction between two massive bodies as a function of their masses and distance. The equation expressing Newton's first law says that the numerical value of the product of the acceleration of a massive object and its mass equals the numerical value of the total force acting on the object. These laws hold for all massive objects, however diverse they may be in other respects. Although the problem of induction is one obstacle to the knowledge of a law, there is another problem concerning our knowledge of functional laws such as the two just mentioned: It is practically impossible to test a hypothesis bearing on a law expressing a constant proportion of the values of certain magnitudes because these magnitudes are not instantiated in isolation, but by concrete events which themselves also depend on other properties.

There are two reasons why a law such as the law of gravitation cannot be tested directly. 1) The first is that there is no system of two masses that is not also subject to the attraction of other masses, in general, at a greater distance. 2) The second is that massive objects also have other properties that can give rise to other forces. Russell concludes that the quantitatively exact laws of mature sciences are not causal because the referents of their terms are not – as causes and effects would have to be – directly accessible to experience. "In all science we have to distinguish two sorts of laws: first, those that are empirically verifiable but probably only approximate; secondly, those that are not verifiable, but may be exact". (*ibid*., p. 203) The first type of laws corresponds to the "causal laws" of common sense and of sciences at the beginning of their development, whereas the laws of mature sciences belong to the second type: they cannot be interpreted as causal since their terms do not refer to concrete events.

1.3. *Ceteris paribus* laws

The problem raised by Russell has been the object of a rich literature on so-called *ceteris paribus* laws[8]. It has been noted that the interpretation of many quantitative laws presents us with a dilemma.

Either

1) One supposes that laws bear on concrete objects or events that are directly accessible to experience. If so then it turns out that these laws have exceptions or, in other words, hold only *ceteris paribus*;

Or

2) One supposes that laws bear neither on particular objects nor on particular events.

---

7 This traditional assumption has been challenged by the elaboration of the concept of *backward causation*, which is intended to apply in particular to certain processes in particle physics. Cf. Dowe (1996). Simultaneous causation raises its own problems.

8 See, e.g. the special issue of *Erkenntnis* 57(3) (2002).

Then it becomes hard to understand how it is nevertheless possible that such laws are being used to produce scientific explanations and predictions.

Hempel gives the following example. For every bar magnet *b*, "if *b* is broken into two shorter bars and these are suspended, by long thin threads, close to each other at the same distance from the ground, they will orient themselves so as to fall into a straight line" (Hempel 1988, p. 20). This generalization is not true without exception of the movement of concrete bar magnets: in certain circumstances, like when a strong air current blows in the direction perpendicular to the orientation of the magnet or when there is a strong external magnetic field, the two halves of the magnet do not align. Similarly, if one takes the law of gravitational attraction to bear on concrete massive objects, so that it determines the net force acting on them (which in turn determines their acceleration) as a function of their masses and their distances, the law has numerous exceptions:[9] an object with mass m_1 that is at distance d of a second object with mass m_2, is in general not subject to a net force $\boxed{}$ in direction of this second object (nor accelerated with $G\frac{m_2}{d^2}$ in its direction).

However, it is not necessary to conclude from this, with Cartwright (1983), that the laws "lie"[10]. Several strategies are available for reinterpreting functional equations and other nomological statements in such a way that they turn out true, despite the fact that the evolution of concrete objects and events often does not (strictly speaking) match with these equations and statements. One strategy consists in taking laws to bear only on systems that are in *ideal* situations, which means in particular that they are isolated[11]. For certain laws, such as the law of gravitational attraction, this has the consequence that the laws bear on no real system (because no real system is ideal in the sense of being isolated from external gravitational influences). Moreover, even if there were isolated systems this strategy faces the difficulty of explaining how a law that is true only of idealized situations can nevertheless be used for the prediction and explanation of facts concerning real systems.

Another strategy consists in taking laws to bear on abstract models rather than on real systems. Smith (2002) proposes to solve the problem of interpreting *ceteris paribus* laws by distinguishing between fundamental laws and equations of movements. Fundamental laws do not directly apply to real concrete systems. The law of universal gravitation determines the force with which two masses attract each other. However, this law cannot be used to *directly* calculate the movement of real objects, to the extent that no real object is exclusively subject to the gravitational attraction due to its interaction with a single other object. Every real object is attracted by many other massive objects, over and above being in general subject to other forces. Smith presents the law of universal gravitation as featuring in an algorithm or "recipe" for constructing a model. The last step of the algorithm leads to an equation of movement that is specific for a concrete system. In this sense, it does not have, according to Smith, the generality required for a law. Smith's fundamental laws correspond to the laws of which Russell says that they are not verifiable but can be exact. Among these fundamental laws, there are in particular the laws determining the different forces that are exerted on an object as a function of its properties and the other objects represented in a model A that contains a partial specification of the properties of a concrete system C under consideration. If C does not evolve as predicted by model A, this indicates simply that A represents C only incompletely. In this case, it may be necessary to improve A by including in it additional

9 Cartwright (1983), p. 57/8; Hempel (1988), p. 23; Pietroski et Rey (1995) p. 86; Smith (2002).

10 The title of Cartwright's book says, ambiguously, "How the Laws of Physics Lie", which could also mean "How the laws of physics stand". However, in her introduction, Cartwright explains that this is not the intended interpretation: "laws in physics […] must be judged false" (Cartwright 1983, p. 12).

11 Silverberg (1996); Hüttemann (1998).

objects, properties and interactions. The equations of movement that are calculated (on the basis of model A) in order to represent the evolution of sets of concrete systems C correspond to the laws, of which Russell says that they are "empirically verifiable but probably only approximate (*ibid*., p. 203), because nothing prevents a certain concrete system C to be subject to the influence of factors not represented in A.

In a similar spirit, Cummins (2000) has suggested distinguishing between "general laws of nature", whose domain of application is unlimited, and "*in situ* laws", which apply only to systems of a particular type, such as planetary systems or living beings, by virtue of the constitution and organization of these systems. If such a system, which Cartwright (1999) calls a "nomological machine", evolves according to a (system) law, its evolution can be seen as a causal process. In contrast with general laws of nature, system laws are not strict. Exceptions result from influences that perturb the evolution of the system from outside.[12] These perturbations can be the objects of causal judgments. According to Menzies (2004), every causal statement presupposes a model (constituted by a natural kind and a law applying to that kind). A factor is judged to be a cause if it makes a difference to the evolution of the system, relative to the background of the normal evolution of the model[13]. In one of Menzies' examples, a person who has been smoking for years develops cancer. Intuitively, the fact that the person is born and the fact that she has lungs are not causes of her cancer although both are necessary conditions. Menzies explains this intuition by suggesting that the identification of a cause normally constitutes the response to a "contrastive why-question" (Menzies, 2004, p. 148), of the form: "why did the man get lung cancer rather than not?" (Menzies, 2004, p. 149). The real history is compared with a fictive (or "counterfactual") history, in which the person does not develop any cancer. The facts of being born and of having lungs are not causes because they also feature in the fictive history.

Russell's analysis shows that laws having the form of quantitatively precise functional dependencies as they are used in mathematical physics cannot be interpreted as directly expressing regularities among observable events; more particularly, they cannot be interpreted as generalizations expressing the succession of causes and effects. This raises the general problem of understanding the relation between laws or models as they are used in the advanced sciences and their use for the prediction and explanation of real concrete systems. As the contemporary debate on *ceteris paribus* laws shows, this difficulty is not specific to the scientific justification of *causal* judgments. The same difficulty arises e.g. in the context of the determination of the spatial conformation of a macromolecule, on the basis of its components and the laws governing their interactions by virtue of their properties. Here, the notion of causality does not come into play because the dependence at issue of the macroproperty on the microproperties is simultaneous dependence between different properties of the same object. While the difficulty of understanding the application of models to real systems raises an important challenge to philosophy of science, it is not specific to the justification of causal judgments. The same can be said of the problem of induction: As Russell notes, it poses a principled obstacle to the knowledge of causal generalizations. However, the problem of induction is a general problem that arises just as well in the context of the knowledge of non-causal generalizations.

2. The reduction of causation to deductive-nomological explanation

12 Cf. Kistler (2006).

13 Menzies' idea that a cause is a factor that "makes a difference" relatively to a background makes use of Mill's (1843) analysis of the distinction between causes and conditions, and of Mackie's (1974) conception of the background as the "causal field". Similar ideas can be found in Lewis' (2000) analysis of causation in terms of influence, and in Hitchcock's (1996) and Woodward's (2003; 2004) work.

The most specific challenge raised by Russell's arguments is the justification of the characteristic features of causality, first and foremost its asymmetry, i.e. an event *c* cannot be both the cause of a second event *e* and its effect. Russell argues that no asymmetry of this sort exists at the level of the functional laws of physics. However, this does not show that there are no asymmetric relations in reality; it only shows that the scientific explanation of the source of this asymmetry must be found somewhere other than these functional laws.

The fact that the notion of cause does not appear in fundamental physics does not make the project of a philosophical analysis of this notion illegitimate. The laws of fundamental physics and causal judgments do not apply to the same objects: the values of the variables that figure in the former are determinate quantities that characterize certain *properties* of substances or events, whereas the terms of causal relations are concrete events. Given that causal judgments regularly occur not only in the judgments of common sense but also in many philosophical projects and in judgments bearing on the experimental testing of scientific theories[14], the project of a naturalistic analysis of causation has been very actively pursued during the 20th century, beginning with Russell himself[15].

The so-called deductive-nomological (DN) analysis of causation has been dominant during the first half of the 20th century. It can be seen as a contemporary version of the traditional reduction of causality to regularities and laws of nature. However, this reductive analysis of causation in the tradition of 20th century logical empiricism takes a form that distinguishes it from its philosophical predecessors. Instead of beginning, like Hume, with the analysis of the *idea* of causality that arises from the *experience* of the regular repetition of certain successions of events, and instead of suggesting, like Galileo, Newton, and many others, to substitute the notion of law for the notion of cause, the DN analysis aims at analyzing first of all causal *explanation*, as it is accomplished in the sciences (see chapter ??? of this volume). According to this analysis, it is equivalent to say that C causes E and to say that C figures as a premise in a DN explanation of E: the effect E is the *explanandum* – what is to be explained – and occupies the role of the conclusion of the argument, and the cause is the content of one of the premises that together constitute the *explanans* – that which explains. Here is how Carnap justifies his analysis of causation in terms of DN explanation: "What is meant when it is said that event B is caused by event A? It is that there are certain laws in nature from which event B can be logically deduced when they are combined with the full description of event A." (Carnap 1966/1995, p. 194)[16] It is essential for a scientific explanation that the link between the premise designating the cause and the conclusion designating the effect be provided by one or several laws of nature. If E were a *logical* consequence *of C alone*, their link would be logical or conceptual, which would be incompatible with the generally accepted Humean thesis that causation is a contingent relation. In retrospect, the attempt to reduce causation to deducibility with the help of laws appears as an attempt to *eliminate* causality and to replace it by mere laws. Such an analysis may well keep the word "causality" but the DN analysis deprives the word of its content: to say that C figures in a *causal* explanation of E means nothing more that to say that C figures

14 Cf. Putnam (1984).

15 In 1914, Russell explains that "there is, however, a somewhat rough and loose use of the word 'cause' which may be preserved. The approximate uniformities which lead to its pre-scientific employment may turn out to be true in all but very rare and exceptional circumstances, perhaps in all circumstances that actually occur. In such cases, it is convenient to be able to speak of the antecedent event as the 'cause' and the subsequent event as the 'effect'" (Russell 1914/1993, p. 223). Russell (1948/1992, p. 471ff.) presents a more elaborate theory of causation.

16 Popper also identifies causal explanation with scientific explanation, in the framework of the DN model: "To give a *causal explanation* of an event means to deduce a statement which describes it, using as premises of the deduction one or more *universal laws*, together with certain singular statements, the *initial conditions*." (Popper 1935/2002, p. 38; italics Popper's).

in a *scientific* explanation of E. If all scientific explanations are causal, the concept of causation loses its discriminative content.

The main reason why the DN analysis has widely been abandoned[17] is that it has become clear that some scientific explanations are *not* causal: there is a specific difference between non-causal and causal explanations that the DN analysis denies. Many physical explanations using functional dependences do not intuitively correspond to causal relations: when the thermal conductivity of a copper wire is deduced from its electric conductivity or vice versa (according to the Wiedemann-Franz law, which says that the values of these two properties of metals are proportional), none of them appears to be the cause of the other. In the same way, when the temperature of a sample of gas that can be considered to be "ideal" (in the sense of falling in the domain of validity of the ideal gas law according to which the product of pressure P and volume V of a sample of ideal gas equals the product of the volume V it occupies, the number n of moles contained in the sample and the universal gas constant R: pV=nRT) is deduced from its pressure, given the volume it occupies, it seems intuitively clear that the pressure of the gas is not the cause of its temperature. Pressure and volume characterize the same individual sample at the same time; their correlation can be explained by processes at the level of the molecules composing the gas. The ideal gas law being symmetrical, DN explanations that can be constructed on its basis cannot be causal without contradicting the asymmetry of causation. If the fact that P(x,t) (the pressure of sample x of gas at time t) is proportional to T(x,t) sufficed to establish that P(x,t) causes T(x,t), T(x,t) would cause P(x,t) for the same reason.

3. The analysis in terms of counterfactual conditionals

Given the number and the diversity of the counterexamples that have been found against the analysis of causation in terms of DN explanation, many philosophers have found it judicious to abandon that analysis. In a passage that marks a turning point in philosophical thinking on causality, David Lewis writes in 1973: "I have no proof that regularity analyses are beyond repair, nor any space to review the repairs that have been tried. Suffice it to say that the prospects look dark. I think it is time to give up and try something else. A promising alternative is not far to seek." (Lewis 1973/1980, p. 160). The basic alternative idea Lewis has in mind can be found in Hume's *Enquiries Concerning Human Understanding*. After his famous definition of causation in terms of succession, Hume offers a second definition: a cause is "an object, followed by another, […] where, if the first object had not been, the second never had existed." (Hume 1777, p. 76)[18] This second definition contains the leading idea of what is now known as the counterfactual analysis of causation: the proposition "*c* causes *e*" means that "if *c* had not occurred, *e* would not have occurred either". The latter proposition is often represented by the expression "C □→ E"[19]. This analysis is intended to be *a priori*, in the sense that its aim is not to discover the physical nature of real causal processes, but rather something that is implicitly known by every competent speaker of English (or any other language containing a synonym of the word "cause"), namely the meaning of the concept expressed by the predicate "causes". In the tradition of logical empiricism, the use of counterfactuals was considered methodologically suspect. Indeed, determining the truth value

17 I cannot develop here the reasons that have led to abandoning the classical conception of logical empiricism, i.e. the assimilation of causation to scientific explanation in the form of a deductive-nomological argument. See chapter 4 of Barberousse, Kistler, Ludwig (2000).

18 Hume does not develop this new idea, nor does he comment on the fact that it is not equivalent to the analysis of causation in terms of regularity.

19 In Lewis' terminology, upper case C represents the proposition that the event named by the corresponding lower case letter *c* has occurred. Except when quoting Lewis, I stick to the usual convention of using lower case letters like *c* and *e* for events and upper-case letters for predicates and propositions.

of a counterfactual proposition requires evaluating possibilities, which are not observable[20]. However, the elaboration of a formalism in which modal and counterfactual propositions can be interpreted in terms of possible worlds has given new life to the project of an analysis of causation in counterfactual terms. The strength of the counterfactual approach rests on the initial plausibility of the idea that a cause "makes a difference", an idea that can be expressed in a quite straightforward way by a counterfactual conditional[21].

David Lewis' contribution to the counterfactual analysis of causality has determined the orientation of all subsequent research in this framework. Lewis proposes to conceive of the semantic evaluation of counterfactuals in terms of the similarity of possible worlds. The terms of causal relations and of counterfactuals are events, where "event" is understood "in the everyday sense" (1986b, p. 161) of a particular happening at a determinate place and time.

The strategy adopted by Lewis for determining the truth conditions of counterfactuals consists in comparing different possible worlds with respect to their global similarity with respect to the actual world, where "actual" is understood in the modal sense. It starts with the thesis according to which the counterfactual proposition expressed by "if C were the case, E would be the case" is true in a world *w* if and only if 1) C is not true in any possible world or 2) if some world in which both C and E are true is closer to *w* than all possible worlds in which C is true but E false. When one asks whether *c* causes *e*, one presupposes that *c* has occurred, and that C is therefore true in the world *w*. On the basis of this presupposition, the second clause determines the truth value of the counterfactual.

Lewis' analysis of the causal relation in counterfactual terms is indirect; it uses causal dependence as an intermediate concept. If *c* and *e* are two distinct actual events[22], *e* depends causally on *c* if and only if it is true that "if *c* had not occurred, *e* would not have occurred". Causation is then defined by the existence of a set of intermediate events constituting a chain reaching from the cause *c* to the effect *e*: *c* is a cause of *e* if and only if there is a finite chain of intermediate events $e_1, e_2, \ldots e_k$, between *c* and *e*, such that the second link of the chain depends causally on the first, and in general if, for every n, the n^{th} link depends causally on the preceding $(n-1)^{th}$ link. The events *c* and *e* must be distinct in the sense that the space-time region in which *c* occurs must not overlap the region in which *e* occurs. With this restriction, the analysis avoids the problem of wrongly classifying non-causal dependence relations as causal: it is clear that the truth of the counterfactual "if John had not said 'hello', he would not have said 'hello' loudly" does not reveal the existence of any causal relation[23].

The counterfactual analysis can account for both deterministic and indeterministic causality. In a world in which there are indeterministic laws, *e* depends causally on *c* (where *c* and *e* are distinct events occurring in the actual world) if and only if, if *c* had not occurred, the probability of the occurrence of *e* had been much less than it actually was (Lewis 1986c, p. 176).

Several objections have been raised against Lewis' analysis of causation. Two sorts of counter-examples have been found: "false positives" seem to show that counterfactual dependence is not sufficient for the existence of a causal relation, whereas "false negatives" seem to show that it is not necessary either. We will look at some of these counterexamples and the lessons to be drawn from them. However, rather than taking these criticisms as

---

20 J. St. Mill (1843) analyzes the counterfactual "if A occurred, then B would have occurred" in terms of the possibility to deduce B from A together with a set of auxiliary propositions S, which must necessarily contain laws of nature. Thus understood, the counterfactual analysis is equivalent to the DN analysis.

21 Mackie (1974, chap. 2) has enriched the counterfactual analysis by the distinction between the background "causal field" and the salient factor that appears intuitively to be the cause insofar as it "makes a difference" with respect to the background.

22 In the general case where *c* and *e* are possible events, it must be true both that "if *c* had not occurred, *e* would not have occurred" and "if *c* had occurred, *e* would have occurred".

23 Cf. Kim (1973); Lewis (1986a).

refutations, advocates of the counterfactual analysis regard these problems as indications of a need for improvement.

A first difficulty for the counterfactual analysis stems from the existence of so-called *backtracking* counterfactuals, according to which a past event depends counterfactually on a present or future event. Take a wave on the ocean. It seems correct to say: "if a given wave summit had not been at x at time t, it would not have been at x-dx at time t-dt", where "x-dx" represents the location of the wave summit at a moment t-dt preceding t. Such backtracking counterfactuals seem to be true in conditions in which some event *c* is a sufficient condition for some later event *e*, in the sense that, once *c* had happened, nothing could have intervened to prevent *e* from happening. In such a situation, it seems true that, if *e* had not occurred, *c* would not have occurred either. Take a situation in which a bomb explodes at instant t after having been triggering by a detonator, and suppose that the triggering is sufficient for the explosion, in the sense that the explosion could not have been prevented once the triggering had occurred. It seems correct to say: if the bomb had not exploded, its detonator would not have been triggered. Now, if there are true backtracking conditionals, counterfactual dependence is not sufficient for (nor, a fortiori, equivalent to) causal dependence, because the future event cannot be the cause of the past event[24], although the past event depends counterfactually on the future event. The wave summit at (x, t) does not cause the wave summit at (x-dx, t-dt), although the wave summit at (x-dx, t-dt) seems to depend counterfactually on the wave summit at (x, t); similarly, the triggering of the detonator depends counterfactually on the explosion of the bomb but the explosion of the bomb does not cause the triggering of the detonator. In other words, the counterfactual analysis seems to predict wrongly that effects sometimes cause their own causes.

Lewis solves this problem by arguing that the use of backward counterfactuals does not correspond to our "standard" (Lewis 1979/1986, p. 35)[25] strategy of judging the similarity among possible worlds. The justification of this thesis depends on a contingent but real asymmetry of our actual world. According to Lewis (1979/1986, p. 49), a set of conditions is a "determinant" of a given event if these conditions, together with the laws of nature, are sufficient for the occurrence of the event. Among the determinants of an event, there are its causes as well as the traces it leaves behind. The asymmetry of the actual world is grounded on the fact that events have in general few determinants preceding it (its causes) but a large number of determinants following it (its traces). Lewis calls this fact the "asymmetry of overdetermination" (Lewis 1979/1986, p. 49): ordinary events have in general only one cause. It is a contingent fact characteristic of the actual world that events are only exceptionally overdetermined by many causes. If one considers the waves that propagate from a perturbation localized at a point on the surface of a lake, there is only one common cause of numerous perturbations on the surface of the water, whereas the event at the origin of the wave has numerous traces: the origin of the wave is overdetermined by the traces in its future, whereas these traces are not overdetermined by the point-like cause in the past.

Here is how Lewis justifies his thesis that backward counterfactuals are not relevant for the analysis of the meaning and truth value of causal statements. To judge whether *e* depends counterfactually on *c*, it is necessary, according to the counterfactual analysis, to evaluate the counterfactual "if *c* had not occurred, *e* would not have occurred". This requires considering

[24] I put the possibility of backward causation to one side here. It remains controversial whether and how backward causation might be conceived and whether such a concept can be applied to certain physical processes. Cf. Faye (2010).

25 Given that counterfactuals are in general vague and given that that their evaluation depends on the context, Lewis (1979/1980, p. 32-5) acknowledges that there are particular contexts, in which we take backward counterfactuals to be true. However, he argues that these particular contexts should be excluded from the evaluation of those counterfactuals that must be used for the analysis of causal dependence.

possible worlds in which *c* does not occur. Such worlds differ from the actual world, for in the actual world, both *c* and *e* occur. Among those possible worlds in which *c* does not occur, those that determine the truth value of the counterfactual by determining the truth value of the consequent *e*, are the worlds that are closest to the actual world. Lewis gives several weighted criteria for determining whether a world is "closer" to the actual world. The first two criteria in order of decreasing importance are

1) avoiding "big, widespread, diverse violations" (Lewis 1979/1986, p. 47) of the laws of the actual world;

2) maximizing the spatiotemporal region in which there is perfect match with respect to particular facts[26] of the actual world[27].

Recall that the relevant possible worlds all differ from the actual world by the fact that *c* does not occur in them. In the framework of events that are determined according to deterministic laws, this divergence is accompanied either by a vast divergence of states of affairs with respect to the causal histories leading respectively to *c* (in the actual world) and to non-*c* (in the possible worlds under consideration), or by a violation of the laws, i.e. by the fact that the possible worlds under consideration do not perfectly obey the laws of the actual world. Lewis argues that the analysis of our practice of making and evaluating counterfactuals shows that we consider to be closest to the actual world those worlds that resemble the actual world perfectly for their entire history up to the time of *c*, and differ from the actual world by virtue of a localized violation of the laws of nature at a moment just before the time of *c*. We judge such worlds to be closer to the actual world than worlds that do not contain any such "miracles", but differ from the actual world by a great number of facts concerning a large part of their history.

At this point, the "asymmetry of overdetermination" comes into play to guarantee that counterfactuals are evaluated according to the "standard" interpretation, i.e. in such a way that the future depends counterfactually on the past but not vice versa. Given the asymmetry of overdetermination, the worlds in which the miracle takes place in the *past* of *c*[28] are closer to the actual world than worlds in which the miracle takes place in the *future* of *c*. A miracle that would be sufficient to make a non-*c* world "reconverge" towards the actual world so as to resemble the actual world perfectly for the *future* of *c*, would have to be much more extended than the miracle required to prevent *c* in a world that resembles perfectly the actual world with respect to the *past* of c. From this reasoning, Lewis concludes that the relevant possible worlds always contain a miracle occurring at a moment immediately preceding the antecedent of the counterfactual. This "standard" choice of the relative importance of the criteria of similarity between possible worlds, taken to be implicit in our practice of evaluating counterfactual propositions, together with the contingent asymmetry of the actual world, guarantees that all backtracking counterfactuals are false. Consider a "backward" counterfactual of the form "if *e* had not occurred, *c* would not have occurred" where *c* and *e*

---

26 The technical sense of the expressions "fact" and "state of affairs" as they are used in contemporary philosophy has its origin in Wittgenstein's *Tractatus* (1921). According to an important interpretation, a fact ("Tatsache" in German) is what makes true a descriptive statement: the satisfaction of a predicate by an object. The concept of a "state of affairs" ("Sachverhalt" in German) is more general in the sense that it also applies to what is possible, what could be the case. If it is possible that object *a* satisfies predicate P, then "*a* is P" expresses a "state of affairs". If *a* is actually P, "*a* is P" also expresses a fact.

27 Lewis mentions avoiding small divergence with respect to laws or facts as separate criteria: "(3) It is of the third importance to avoid even small, localized, simple violations of law. (4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly" (Lewis 1979/1986, p. 48).

28 The past with respect to the moment at which *c* occurs in the actual world. An event *e* in world w_1 appears as a miracle with respect to world w_2 if the circumstances in which *e* occurs (in w_1) are not in conformity with the laws of w_2. Then *e* is a miracle in w_1 relative to w_2.

are events that occur in the actual world and where *e* occurs *later* than *c*. The possible worlds that are relevant for its evaluation are those in which the antecedent non-*e* is true by virtue of a "tiny miracle" that occurs *immediately before* the occurrence of *e* in the actual world. Thus, the miracle occurs *after* the occurrence of *c*; therefore, *c* occurs in the closest possible world in which the antecedent of the counterfactual is true; therefore, the consequent of the backtracking counterfactual is false, and thus the counterfactual itself is false as well.

The argument that establishes that backward counterfactuals are systematically false also provides a solution to what Lewis (1986b, p. 170) calls "the problem of epiphenomena": consider an event *c* that causes two effects, *e* and *f*, but where *e* does not cause *f* nor does *f* cause *e*. Lewis' analysis seems to predict wrongly that *e* causes *f* because there seems to be a chain of counterfactual dependences between *e* and *f*: if *c* is necessary in the circumstances for *f* then *f* depends counterfactually on *c*; and if *c* is sufficient for *e* then *c* seems to depend counterfactually on *e*: if *e* had not occurred, *c* would not have occurred. Now if Lewis' argument is correct to the effect that our criteria for evaluating counterfactuals guarantee, in the context of the asymmetry of overdetermination, that backward counterfactuals are always false, then the latter counterfactual is false, and there is not after all any chain of counterfactual dependence between the epiphenomena *e* and *f*.

Several objections have been raised against this reasoning. Horwich (1987, p. 10) notes that the asymmetry of overdetermination is known only by science and *a posteriori*. Insofar as it is not an aspect of reality that is known *a priori* by all competent speakers, a conceptual analysis of the concept of causation cannot make use of it[29]. Several authors have questioned the scientific correction of Lewis' (and Popper's 1956) thesis according to which events have typically few determinants preceding them but many determinants following them, or, in other words, few causes and many traces. Concerning the deterministic and symmetric laws of classical mechanics, this difference is in fact illusory. Elga (2000) has shown that, even for counterfactuals whose antecedent expresses an irreversible event in the thermodynamic sense (of an increase in entropy), Lewis is wrong to say that worlds in which the antecedent is true by virtue of a miracle that occurs immediately *before* the antecedent are closer to the actual worlds than worlds in which the miracle occurs *after* the antecedent. Elga illustrates this point with a situation in which Gretta smashes, in the actual world w_1, an egg in her pan at 8 o'clock. Consider the closest worlds in which Gretta does not smash any egg at 8 o'clock. According to Lewis, it needs only a tiny miracle, e.g. in a process taking place in Gretta's brain just before 8 o'clock, say at 7:59, that guarantees that she does not smash any egg. Such a world w_2 containing a miracle at 7:59 resembles the actual world perfectly with respect to all facts in the whole of history right up to 7:59, and diverges from it only after the time of the miracle. However, Elga shows that there is a world w_3 that shares, contrary to the actual world, the whole set of facts pertaining to the future beginning from a moment just after 8, say from 8:05, so that there is in w_3, after 8:05, a smashed egg just as in the actual world w_1. These are worlds in which Gretta smashes no egg but in which the miracle that guarantees the convergence with respect to the actual world is not larger than the miracle that occurs in world w_2. Elga has us consider a process that corresponds to the process taking place in the actual world from 8 to 8:05 but which evolves in the opposite direction, like when one watches a film in the wrong direction. The egg that has been smashed in the pan "*uncooks*" beginning at 8:05 and returns in the eggshell. This process is in agreement with the laws of physics although it is very improbable because it depends in an extremely sensitive manner on its initial conditions: if one produces an tiny change in the positions and speeds of the molecules at 8:05, a more banal process will take place, in which the egg remains in the pan and starts cooling down. Thus it suffices have a tiny miracle at 8:05, to guarantee that the

29 Lewis answers this objection in (1979/1986, p. 66).

entire past changes, including Gretta's act of smashing an egg at 8 o'clock. With such a small miracle at 8h05, the whole past in w_3 is different from what it is in the actual world, and does in particular not contain Gretta's smashing any egg at 8 o'clock. However, worlds w_1 and w_3 resemble each other perfectly for all times after 8:05. Thus, w_3, in which the miracle that ensures that the smashing does not occur happens *after* 8 o'clock (the time of the smashing in w_1), does not differ more from w_1 than world w_2, in which the miracle occurs *before* 8 o'clock.

We have seen that Lewis defines causation indirectly, using the notion of causal dependence as an intermediary between counterfactual dependence and causation: *c* is a cause of *e* if and only if there is a finite chain of intermediate events $e_1, e_2, \ldots e_k$, between *c* and *e*, such that the second link of the chain depends causally on the first, and in general if, for every n, the n^{th} link depends causally on the preceding $(n-1)^{th}$ link. Causal dependence is then, as we have seen, reduced to counterfactual dependence.

This analysis solves two difficulties: first it guarantees the transitivity of the causal relation, and second it allows justifying the intuition that a "pre-empted" cause is only a potential rather than an actual cause.

1) Counterfactual dependence is in general not transitive: it is easy to find examples where it is true that A □→ B and that B □→ C, but false that A □→ C. The reason is that the evaluation of a counterfactual depends on the background circumstances of the antecedent, and that the antecedents in a series of counterfactuals do not in general share their backgrounds. When the causal relation is reduced to a chain of events in which each link depends counterfactually on the preceding link (instead of reducing it directly to causal dependence), the first and the last link of a causal chain are guaranteed to be linked as cause and effect, whereas the last link does in general not counterfactually depend on the first. However, this aspect of Lewis' analysis has also given rise to an objection. Several authors claim that there are counter-examples to the transitivity of causation. In particular, such counter-examples concern judgments in which an absence, or a particular aspect of an event, play the role of cause or effect, or judgments in which the causal link is grounded on a double prevention[30]. In an example offered by Ehring (1987), someone puts potassium salts in the fireplace, which brings about a change of the color of the flame from orange to purple. Later, the flame lights a piece of wood next to the fireplace. There is a causal chain between the act of putting potassium salt in the fireplace and the lighting of the piece of wood. However, it seems false to say that the first event causes the last[31]. The transitivity of causation can be defended against certain counter-examples by showing that the appearance of the existence of a causal chain is due to too coarse a conception of the terms of the relevant causal relations. If the terms of the causal relations are not concrete events, but *facts* bearing on these events, there does not appear to be any chain linking the act of throwing salt in the fire to the lighting of the piece of wood: the salt is causally responsible for the fact that the flame changes color; however, the cause of the lighting is not the fact that the flame changes its color but rather the fact that its gives off heat[32]. It can also be defended by denying that there are causal relations with "negative" terms, such as absences or omissions: such relations correspond often to non-causal explanations, which can give an illusory impression of causality. Such statements describe a situation lacking any causal process, which is implicitly contrasted with a background situation in which there is such a causal process[33]. If this is correct, explanatory chains containing double prevention do not in general indicate the existence of a causal chain.

30 See Bennett (1987); Hall (2004a).

31 Other examples of this kind can be found in McDermott (1995), Hall (2000/2004b), Paul (2004).

32 See Kistler (2001). Paul (2004) offers a similar analysis, in which she argues that causation relates aspects of events, rather than events themselves.

33 Cf. Kistler (1999/2006); Hall (2004a); Kistler (2006).

To use an example that Hitchcock (2001) attributes to Ned Hall[34], a hiker sees a rock falling, which causes him to duck so as to avoid being hit by the rock. The fact that he has not been touched might seem to be a cause of the pursuit of the trek. This is a case of *double prevention*, in the sense that the ducking prevents the rock from preventing the pursuit of the hiker's trek. It seems wrong to say that the falling of the rock caused the pursuit of the trek, although there seems to be a causal chain from the first event to the last. However, it can be denied that it is a causal chain, thereby defending the transitivity of causation, by denying that the negative fact of *not* being touched by the rock can be either an effect or a cause.
2) The second problem that the introduction of a chain of intermediate events solves arises in the context of situations of "preemption" and cases involving "redundant causation". Such situations are frequent, for instance in biology. For example, sometimes it is said that evolution brings about both a main mechanism important for an organism's survival, and a backup mechanism, something that takes over in case of failure of the main mechanism[35]. Other examples involve human actions. One of the paradigm cases of preemption in the literature involves two snipers, S_1 and S_2, who aim at the same victim at the same time. S_1 decides to fire (event *a*); this decision causes her shot, which causes the death of the victim (event *c*). S_2 who sees S_1 shoot does not shoot and thus does not cause *c*; S_2's determination to fire (event *b*) is not followed by S_2's firing: the process is interrupted by S_2's seeing S_1 fire. This situation shows that counterfactual dependence is not necessary for causation: *a* causes *c* although *c*, the victim's death, does not counterfactually depend on *a*. If *a* had not happened, S_2 would have fired. The event *b*, corresponding to S_2's determination to fire, would have caused S_2's firing, which would have caused *c*; in short, *c* would have happened even without *a*.

The requirement of the existence of a chain of intermediate events solves this difficulty: for event *a*, the positions of the bullet on its trajectory from *a* to *c* constitute such a chain. By contrast, given that S_2 does not fire, there are, for all times following S_2's noticing that S_1 has fired, no intermediate events between *b* and *c* on which the death of the victim depends counterfactually and which depend on *b*. Lewis' analysis yields the intuitively correct result that *b* is no cause of the death of the victim. This type of situation is called "early preemption", because, insofar as the potential causal chain between *b* and *c* is interrupted early, i.e. a sufficiently long time before *c*, there exists a chain of events between *a* and *c* to which no parallel chain between *b* and *c* corresponds.

However, this solution is ineffective in cases of what has been called "late preemption", in which there is a continuous chain of events between events *b* and *c*, but where *b* still does not cause *c*. Hall (2004a, p. 235) for instance considers the situation in which two children (Suzy and Billy) throw rocks at a bottle. Suzy throws her rock a little earlier than Billy, so that her rock smashes the bottle (event *c*). However, Billy's rock follows closely behind Suzy's rock, so that there is not only a chain of events between Suzy's throw and *c*, but also between Billy's throw and *c*. Nevertheless, to the extent that Suzy's rock reaches the bottle a moment before Billy's, Suzy's but not Billy's rock is the cause of *c*.

In "Postscripts to 'Causation'", Lewis (1986c) introduces the concept of "quasi-dependence" to solve the problem of late preemption. In cases of late preemption, in spite of the presence of the preempted event *b*, and in spite of the fact that there is an entire parallel chain from *b* to *c*, the "preempting" event *a* causes *c*. The reason why the presence of the

34 According to Hitchcock (2001, p. 276), this example features in an unpublished version of Hall (2004a).

35 The main mechanism for the orientation of honey bee workers relies on the perception of the location of the sun, but backup mechanisms are available for situations where the sun is not directly visible: one relies on the perception of patterns of (the ultraviolet component) of polarized light, another on the perception of landmarks (Winston 1991, p. 163/4).

redundant cause *b* does not deprive *a* of being efficacious in causing *c* is the fact that causality is an *intrinsic* quality of the process localized between *a* and *c*. According to Lewis, each event in the chain between *a* and *c* is *quasi-dependent* on its predecessor because the process intrinsically resembles – i.e. if only the events localized on the chain linking *a* to *c* are taken into account – processes whose elements are fully counterfactually (and therefore causally) dependent on their predecessors. Event *a* (Suzy's throw) is the cause of *c* because *a* intrinsically resembles possible throws that Suzy executes in the absence of any of Billy's throws. Event *c* is quasi-dependent on Suzy's throw because *c*'s counterpart in such possible situations (where Suzy throws but Billy doesn't) is counterfactually dependent on the counterpart of Suzy's throw.

However, there are even more problematic cases of preemption that involve a chain of intermediate events that makes the effect *c* "quasi-dependent" on an earlier preempted event *b* (which is *not* a cause of *c*). Schaffer (2000) calls this sort of situation "trumping preemption": a major and a sergeant shout orders at a corporal. Both shout "Charge!" at the same time, and the corporal decides to charge. Given that a soldier obeys the orders of the higher-ranking soldier, the cause of the corporal's decision is the major's order, not the sergeant's. However, the corporal's decision is quasi-dependent both on the sergeant's and on the major's order. The chain reaching from one of the orders to the corporal's decision is intrinsically similar to a chain that, in the absence of the second order, guarantees counterfactual dependence along the links of the chain and therefore the existence of a causal relation. Quasi-dependence is therefore not, after all, sufficient for causation.

This difficulty has led Lewis (2000) to devise a new version of his counterfactual account, in terms of "influence". Lewis suggests that *the fact* that the occurrence of *e* is counterfactually dependent on the occurrence of *c* is not by itself sufficient for *c* being a cause of *e*; there is the further requirement that *the way* in which *e* occurs and *the moment* at which *e* occurs also depend counterfactually on the manner and the moment in which *c* occurs. Lewis' new analysis employs the notion of the alteration of an event. An alteration of an actual event *e* is a possible event that differs slightly from *e*, either by its properties or by the moment at which it occurs. If an event *c* influences another event *e*, there is "a pattern of counterfactual dependence of whether, when and how on whether, when and how" (Lewis 2000/2004, p. 91). More precisely: "Where C and E are distinct actual events, let us say that C *influences* E iff there is a substantial range C_1, C_2,… of different not-too-distant alterations of C (including the actual alteration of C) and there is a range E_1, E_2,… of alterations of E, at least some of which differ, such that if C_1 had occurred, E_1 would have occurred, and if C_2 had occurred, E_2 would have occurred, and so on" (Lewis 2000/2004, p. 91; emphasis Lewis')[36]. Just as in his original analysis, the fact that *c* causes *e* is reduced to the existence of a chain of intermediate events in which each link influences the following link.

Another objection against the counterfactual analysis concerns the fact that it does not respect the common sense distinction between causes and background conditions. Now one might consider rejecting this distinction (as did Mill) since the distinction only reflects the interests of human observers; but "philosophically speaking", background conditions are causes in the same sense as salient factors that common sense recognizes as causes. However, to the extent that the aim of the counterfactual analysis is not the nature of causation as it is in reality, but the structure of our naïve concept of causation, it seems essential that the analysis respects this distinction. To accomplish this, one can hypothesize that ordinary causal statements like "c causes e" in fact contain implicit comparisons to a "normal" background situation. This can be made explicit in a paraphrase of a form such as "*c*

[36] I have kept Lewis' notation, where the upper case letter "C" represents "the proposition that *c* exists (or occurs)" (1986b, p. 159), where lower case "*c*" represents a particular event.

rather than *c** has caused *e* rather than *e**". The correct counterfactual analysis would then be: "if *c** had occurred rather than *c*, *e** would have occurred rather than *e*"[37]. This idea is closely related to the intuition that a cause makes a difference with respect to its effects: one compares, though often implicitly, the situation as it is when the cause is present to the situation, as it would have been if the cause had been absent. If the effect is present in a situation in which the cause is present but absent where the cause is absent, one has good reason to think that the cause is responsible of this difference. To use Achinstein's (1975) example, the cause of Socrates' death is his drinking hemlock, because this is the factor that makes the crucial difference with respect to his death. Many other characteristics of the situation, such as the fact that Socrates drank hemlock occurred *at dusk*, are not causes of his death. The time at which the drinking occurred made no difference to the hemlock's fatal effect.

4. Methodology

The successive modifications of the counterfactual analysis are motivated by the attempt to avoid two sorts of counter-examples. "False positives" for a proposed analysis are situations featuring two events that the analysis presents as being related as cause and effect, where intuitively they are not so related. "False negatives" are on the contrary situations in which an event *c* is intuitively the cause of another event *e*, but where the analysis yields the result that it is not. These are the two possible forms of mismatch between a given analysis and intuition. The research on improving the counterfactual analysis is driven by the presupposition that the main criterion of adequacy of a philosophical analysis of the concept of causation is agreement with common sense intuitions. However, this choice of the criterion of adequacy is controversial. The diversity of extant analyses of the concept of causation can be explained at least in part by the existence of different ways of conceiving the aim and method of such an analysis. A major disagreement opposes a priori and a posteriori analyses.

1 Advocates of the counterfactual analysis want to provide a "conceptual analysis" of a concept mastered by everyone (at least everyone within the language community of speakers of some natural language containing causal vocabulary). Just like other common sense concepts, people use causal concepts to reason about possible or counterfactual situations in addition to reasoning about actual situations. For example, causal concepts are also used to reason about the consequences of science fiction novels, where facts and even laws of nature may differ widely from the actual world. If the aim of the philosophical analysis of causation is an analysis of this common sense concept, the analysis must be such that it applies to all possible worlds to which the concept of causation applies. Moreover, insofar as the common sense concept of causation is not informed by scientific knowledge about the physical nature of the causal processes of the actual world, scientific knowledge appears irrelevant to the philosophical analysis of the concept. Therefore a conceptual analysis can be conducted in a purely a priori manner. The adequate method consists in carefully spelling out "from the armchair" one's spontaneous intuitions on a certain number of fictitious situations. And although these situations can reflect real world scenarios, such as children throwing rocks at bottles or soldiers shouting orders, the a priori analysis of our naïve concept of causation can just as well make use of intuitions concerning unreal or even physically impossible situations, such as situations in which magicians cast spells. In a situation conceived by Schaffer (2004, p. 59), Merlin casts a spell that transforms a prince into a frog. Magical causal interactions of this sort are not constrained by

37 Cf. Hitchcock (1996a); (1996b); Maslen (2004); Schaffer (2005).

physical laws and can act at spatial and temporal distance without any causal intermediaries.

2 A theory can start with the analysis of the common sense concept, but then make corrections in order to obtain better coherence and systematicity without thereby abandoning the framework of a priori constraints. It is, e.g. intuitively correct to judge both that an ice cube (more precisely the melting of the ice cube) in a glass of water causes the water to cool down, and that the cooling of the water (more precisely the fact that the water gives off heat) causes the melting of the ice cube. Taken together, the set of these two judgments violates the asymmetry of causation, which is, as we have seen, a central component of the concept of causation. It can be concluded that at least part of the naïve intuitions on this situation must be incorrect. However, there does not seem to be any reason to take one to be incorrect rather than the other.

3 There is an alternative way of conceiving of the aim of the philosophical analysis of causation. Causation can be taken to be a concept of a "natural kind" of relation whose real essence must be discovered *a posteriori*. This is the way in which process theories of causation conceive of their task. From such a perspective, the causal relation whose "real essence" one tries to discover does not exist in all possible worlds. In this framework, one may look for a scientific reason for following one intuitive judgment rather than the other in the case of the two judgments that together violate the asymmetry of causation. The judgment that the cooling of the water causes the melting of the ice cube corresponds to the physical transference of heat, whereas there is no physical process corresponding to the other judgment[38].

From the point of view of the project of conceptual analysis, an approach that takes into account physical constraints on possible causal interactions seems to "suffer from a lack of ambition" (Collins et al. 2004, p. 14). For a priori approaches, the analysis of the concept of causation must apply in all possible worlds to which the concept of causation applies, and in particular in "worlds with laws very different from our own" (Collins at al. 2004, p. 14). Limiting one's reflection to those causal processes that are possible in the actual world given its laws, appears as "not merely unfortunate but deeply misguided" (Collins et al. 2004, p. 14) from the point of view of advocates of conceptual analysis who aim at finding an "account that has a hope of proving to be not merely true, but necessarily so." (Collins et al. 2004, p. 14).

Defenders of the idea that the causal relation is a natural kind of relation whose nature needs to be discovered on the basis of both conceptual and empirical constraints, can reply that we have here two different though related projects. The difference between the research on the naïve concept of causation and the research on what the essence of causation is in the actual world is analogous to the difference between the psychological research on "naïve physics", or "folk physics" and research in physics, or between psychological research on "folk biology" and biological research. Naïve physical concepts and naïve convictions on the properties and the evolution of physical objects determine only very partially the concepts and theories of scientific physics. In an analogous way, our a priori convictions on the nature of causation might only partially constrain the theory of causation as a natural relation existing in the actual world. The nature of such a natural relation must at least in part be discovered by empirical research.

One may try to reconcile the project of a priori conceptual analysis with the project of discovering the nature of causation as a natural kind of process (as it is in the actual world) in

38 One may of course describe the process of diffusion of heat in a negative way. Instead of saying that the water transfers heat onto the ice cubes, on can say that the presence of a colder object diminishes the heat contained in the water.

the framework of what has been called the "Canberra plan"[39]. It proceeds in two steps, the first of which belongs to conceptual analysis: one discovers the constraints that a real relation must satisfy so as to be a candidate for being the causal relation. Transitivity and asymmetry are among these conceptual constraints. In a second step, which is empirical, one discovers which actual relations or processes satisfy the constraints discovered in the first step. The idea is to apply to the concept of causation a general strategy for reducing common sense concepts to scientific concepts, which is known as functional reduction (Jackson 1998, Kim 1998). In the first conceptual step, one shows, e.g., that the concept of water is a functional concept that applies to a substance insofar as it satisfies a certain number of functional conditions: it is liquid at temperatures between 10°C and 30°C, it is transparent but refracts light with a characteristic refraction index, it freezes at 0°C and boils at 100°C under atmospheric air pressure at sea level etc. In the second step, it is empirically discovered that substances that satisfy these conditions in the actual world are mostly composed of H_2O molecules.

5. Causation as a process

As we have seen, an important motivation of the counterfactual analysis has been the discovery of various sorts of "false positives" for the deductive-nomological analysis. Some facts can, on the background of laws of nature, play the role of premises and conclusions of deductive arguments, without being linked as causes and effects. However, certain situations that refute the deductive-nomological analysis are also false positives that refute the counterfactual analysis. In certain background conditions, given two effects e_1 and e_2 of a common cause c, e_1 can serve as a premise in an argument whose conclusion describes e_2, and vice versa. Now, in appropriate circumstances, e_1 and e_2 can also be counterfactually dependent on each other. This parallel is certainly no coincidence: nomological dependence (which is according to the DN analysis a crucial part of what makes causal propositions true) creates counterfactual dependence. This is the case both when the nomological dependence goes together with causation and when it does not. For this reason, counterfactual dependence seems to be too weak to guarantee causation. We have already considered the debate about Lewis' suggestion that the counterfactual dependence between e_1 and e_2 is not sufficient for causation because it is grounded on causal dependences between e_1 and the common cause c and between c and e_2, and because the second counterfactual dependence is backward. This solution does not apply to cases of counterfactual dependence between aspects of an event or situation: given a sample g of gas (which approximately satisfies the conditions for being an "ideal" gas) and the ideal gas law pV=nRT (where p represents pressure, V Volume, T temperature, n the number of moles of gas, and R the universal gas constant), if g had not been at temperature T (supposing its volume to be held fixed), it would not have had pressure p. If the kinetic energy of the molecules contained in g had not been E, the temperature of g would not have been $T=2E/3k_B$ (where k_B represents Boltzmann's constant). It is one of the central conceptual constraints on the causation relation that its terms must occupy distinct spatio-temporal regions. "C and E must be distinct events – and distinct not only in the sense of nonidentity but also in the sense of nonoverlap and nonimplication" (Lewis 2000, p. 78). Pressure and temperature of the same sample of gas at the same moment cannot be linked as cause and effect because there is no spatio-temporal distance between these instances of properties. The same is true of the relation between the temperature of the sample of gas and the mean kinetic energy of its molecules. These examples of dependence between different

39 This expression has been introduced by O'Leary-Hawthorne and Price (1996) by reference to the Australian National University at Canberra, in the context of the analysis of the concepts of truth, reference and belief. Lewis (2000/2004, p. 76) applies it to the analysis of the concept of causation.

properties of a given system at a time show that for such properties, counterfactual dependence is not sufficient for causation.

This problem (as well as the problem that counterfactual dependence is not necessary for causation either, as preemption scenarios seem to show) can be avoided by analyzing causation in terms of a local process that stretches between two events that are localized in space and time. There are several versions of such process accounts of causation. One of its historical sources is Russell's (1948) analysis of causation in terms of "causal lines", which is inspired by the physical notion of a world line. The concept of a world line can be obtained from the spatio-temporal trajectory of an object. In a three-dimensional representation of the position of the Earth in space, its trajectory around the Sun appears as an ellipse. In a four-dimensional representation, in which the temporal dimension is represented as a fourth dimension alongside the three spatial dimensions – following at this point the unification of the spatial and temporal dimensions required in physics by the theory of relativity – the Earth's trajectory appears as its world line, which is an open curve in 4-dimensional space-time.

A causal line is a world line that satisfies an additional condition: along the line there are qualities or structures that are either constant or change in a continuous and smooth manner: "Throughout a given causal line, there may be constancy of quality, constancy of structure, or gradual change in either, but not sudden change of any considerable magnitude." (Russell 1948, p. 477) This condition is supposed to guarantee that causation grounds our acquisition of knowledge. For Russell, as for Hume, the only way in which we can justify beliefs whose subject matter goes beyond what is immediately given to our senses consists in relying on causation. The perception of a table provides knowledge of the table, and not only of the sensory impressions from the table. This is so because these sense impressions are linked by a causal chain to the table, or more precisely to events of interaction between light and the surface of the table. Russell defines the notion of a causal line with respect to the possibility of justifying our inferences to what happens at some distance from ourselves: "A 'causal line', as I wish to define the term, is a temporal series of events so related that, given some of them, something can be inferred about the others whatever may be happening elsewhere." (Russell 1948, p. 477) Any inference of this sort is inductive, and therefore fallible. In this context, Russell notes that an inference to an effect from a given cause is more reliable than a "backward" inference from an effect to a cause. The reason is that events of the same type can have different causes. Now, the inferences that provide us with knowledge of the world external to our sense organs belong to this second and more fragile sort of inferences.

Russell defines causal lines as world lines whose qualitative continuity can serve as inductive justification to enhance our knowledge beyond our perceptions. The fact that causal lines are defined by an epistemic requirement makes them inadequate as a basis for a metaphysical account of causation because they would make the existence of causal processes and relations dependent on human inferences. The fallibility of inferences grounded on the continuity of causal lines shows that such a causal line can only be a fallible indicator of the existence of a real causal process; however, being a causal line is neither necessary nor sufficient for being a real causal process. It is not sufficient because the continuity of structure or quality can also characterize "pseudo-processes" (Salmon 1984). Pseudo-processes are world lines that give human observers the illusory impression of a causal process. Their qualitative continuity qualifies them as Russellian causal lines, even though they are not real causal processes. Take Salmon's (1984, p. 141/2) spot of light cast on the inner wall of a hollow cylinder by a projector rotating at its centre. The world line characterized by the series of places on the wall at the times at which the light spot appears on them is a causal line without being a causal process. The trajectory of the spot of light along the inner wall of the

cylinder can exhibit perfect qualitative continuity. However, it is no causal process because spots of light at successive moments do not exercise any causal influence on one another: the light spot that appears at *x* at *t* does not cause the spot that appears at the immediately following place and time; rather, each spot is the end point of a causal process originating in the projector. Being a causal line is not necessary for being a causal process either because continuity of structure is not necessary: some causal processes are characterized by large and fast qualitative changes, e.g. when several particles of different types follow each other in a "cascade" of radioactive decomposition.

Taking his inspiration from Russell's causal lines and Reichenbach's (1956) concept of a mark, which is defined as a local modification of structure, Salmon (1984) has suggested analyzing the concept of causal process as a process that 1) has structure or qualities that are either permanent or only changing continuously and 2) is capable of transmitting a mark. The light spot gliding along the wall of the cylinder is not a causal process because, if one modifies its color by inserting a red filter between the projector and the wall at one point, this modification will not propagate to the subsequent evolution of the spot.

This analysis in terms of continuity of structure and mark transmission raises several difficulties[40]: causal processes that are characterized by large and fast qualitative changes are counterexamples to the requirement of continuity of structure. Insofar as a world line is subject to changes that are fast relative to the scale of human observation, so that its observation does not give to an ordinary human observer the impression of qualitative constancy or of continuous change, it is neither a Russellian causal line nor a causal process as defined by Salmon. Salmon begins with the Russellian concept of a causal line, which requires the existence of a structure that is preserved along the line, and adds the additional requirement of mark transmission. "A given process, whether it be causal or pseudo, has a certain degree of uniformity – we may say, somewhat loosely, that it exhibits a certain structure. The difference between a causal process and a pseudo-process, I am suggesting, is that the causal process transmits its own structure, whereas the pseudo-process does not." (Salmon 1984, p. 144) A world line that is subject to fast and important qualitative changes, relative to the scale of what it observable by an ordinary human, does not even satisfy the conditions that Salmon imposes on processes: "processes can be identified as space-time paths that exhibit continuity and some degree of constancy of character" (Salmon 1994, p. 298; repr. Salmon 1998, p. 249). A fortiori, it cannot be a causal process. On the other hand, there seem to be pseudo-processes capable of transmitting marks. Kitcher (1989, p. 463) mentions derivative marks: when a passenger in a car holds a flag out of the window, the shadow cast by the car as it passes along a wall bears the mark of the flag. Moreover, the analysis of the notions of mark and of causal interaction seems to be circular: A mark is a modification of structure introduced into a process by a causal interaction, but an interaction is causal if it leads to the introduction of a mark.

A tradition going back to the 19th century[41] identifies causal processes with processes of transmission of energy, momentum (Aronson 1971, Fair 1979), or more generally, of a quantity of a conserved quantity (Salmon 1994; Kistler 1998; 1999/2006). This approach is motivated by a "mechanist" intuition, according to which causal influence propagates only by contact and with finite speed. This intuition manifests itself when one considers certain situations that are problematic for theories analyzing causation in terms of nomological regularity or counterfactual dependence. Thunderstorms follow regularly upon sudden falls of barometer readings. They also depend counterfactually on them: if the barometer had not fallen, there would not have been a thunderstorm. However, the reason for which the barometer reading is nevertheless not a cause of the thunderstorm is that the barometer does

40 These difficulties have led Salmon (1994) to abandon it.

41 See Krajewski (1982).

not take part in the mechanism of the genesis of the thunderstorm. Some authors deny the possibility that a quantity of energy can be transferred in the strict sense: the reason is that particular quantities of energy lack the individuality required to give sense to the idea that it remains the same quantity across time (Dieks 1986). For this reason, the most elaborate version of the process theory in terms of conserved quantities (Dowe 1992; 2000) does not make use of the concept of transmission, but uses instead Russell's concept of the "continuous manifestation" of a conserved quantity. By the continuous manifestation of a property by a world line, Dowe means that this property characterizes all points on the line, which does not require any form of transmission. This makes his account vulnerable to the objection that certain pseudo-processes manifest conserved quantities, without thereby being causal[42]. We have already considered the light spot gliding over the internal wall of a hollow cylinder. The trajectory of this spot constitutes a perfectly homogeneous world line: in the conditions stipulated by this thought experiment, the light spot contains, or manifests, exactly the same energy at each instant; each instant is qualitatively perfectly similar to each other. Nevertheless, the world line constituted by the trajectory of the light spot is not a causal process. The causal process responsible for the light spot is the process of propagation of light from the projector to the wall.

Theories that analyze causation in terms of transmission or continuous manifestation of conserved quantities avoid the problems mentioned above, of the relation between two effects of a common cause and of redundant or preempted processes. The fact that two events are effects of a common cause does not entail that there is a causal relation between those events, since no process of transference may relate them. Moreover, the fact that a process P_1 is accompanied by a second redundant (preempted) process P_2 does not prevent P_1 from transmitting conserved quantities. Consider again two snipers shooting at the same victim from which they are separated by the same distance. Imagine that sniper S_1 shoots a tiny moment earlier than sniper S_2, so that the bullet shot by S_1 kills the victim. In this case, S_2's shot (event *b*) does not cause the victim's death (event *c*). Neither the probabilistic[43] nor the counterfactual analysis can account for the intuition that what the makes S_1's shot (event *a*) the cause of the victim's death must be some feature that is localized at the process linking *a* to *c*. Both the probabilistic and the counterfactual analysis make the existence of a causal relation between *a* and *c* depend on factors that are *not* localized between *a* and *c*. If sniper S_1's shot takes place in a situation in which sniper S_2 also shoots, there is no counterfactual dependence between *a* and *c*: given S_2's shot, it is not true that, had S_1 not shot, the victim would not have died. One of our intuitions seems to indicate that the existence of a causal relation between *a* and *c* can only depend on processes situated between *a* and *c*, and that it cannot depend on events and processes that do not interfere with the processes between *a* and *c*.[44] On the other hand, the analysis according to which causation is grounded on a process of transmission takes into account this intuition of locality, according to which the existence of a causal relation between *a* and *c* only depends on processes between *a* and *c*. If *a* transmits something, say an amount of energy, to *c*, *a* is a cause of *c*, whether or not other events such as *b*, also have a causal impact on *c*.

However, transference theory encounters several important problems.

1 We have already mentioned the objection that the transmission analysis suffers from a

42 See Salmon (1994, p. 308); Kistler (1998); (1999/2006).

[43] The probabilistic analysis will be presented in the next section.

[44] Lewis' (1986c) notion of quasi-dependence makes whether *c* causes *e* depend on possible worlds in which there is a process between *c** and *e** that is intrinsically similar to the process between *c* and *e* and where *e** depends indirectly (through a chain of dependence) counterfactually on *c**. However, whether *c** causes *e** in those possible worlds is not only a matter of the intrinsic characteristics of the local process between *c** and *e**.

lack of ambition, because its target is causation as it is in the actual world, rather than the general concept that applies to all possible worlds. However, this is only an objection to the extent that one shares the presupposition that conceptual analysis is the only legitimate or at least the only sufficiently ambitious aim of philosophical theories of causation.

2 Transference analyses can also be suspected of a lack of ambition of another sort: they seem to apply only to physical causal processes. Therefore the transference analysis seems inadequate for ordinary causal judgments involving non-physical properties, arguably for example psychological properties. To illustrate: the fact that the doorbell rings wakes Peter up. The noise of the doorbell seems to be the cause of his waking up, but it does not seem to be relevant to consider the underlying causal process from the point of view of energy transmission[45]. Indeed the application of the analysis to causal judgments of common sense presupposes that all causes and effects are physical. In reply, there are several ways of articulating the content of ordinary causal judgments with transference theory. The causal judgment that the doorbell wakes Peter up does not directly make reference to energy transmission. The dependence of his awakening on the propagation of sound waves, their transduction in nerve signals and the transmission of the latter to Peter's auditory cortex is the object of several "special" sciences, such as acoustics, psychophysics, physiology and neurophysiology. In a physicalist framework, it is supposed that all these facts supervene[46] on the set of physical facts. If this is correct, the process of the doorbell waking Peter up may supervene on a physical process of transmission. The relevant properties of which the causal judgment states the causal dependence may even be specific forms of conserved quantities. The picture that emerges from this possibility has two parts: two conditions together make true the judgment that the fact that *c* (the activation of the doorbell at time t) is F (makes a specific sound) is causally responsible for the fact that *e* (Peter at the moment immediately following t) is G (wakes up). It is made true by 1) a process of transmission from cause *c* to effect *e* and 2) a law of nature expressing the dependence of G on F (Kistler 1999/2006). To judge that the doorbell wakes Peter up there must be an "in situ" law according to which, in ordinary, non exceptional circumstances, doorbells wake sleeping people up, or at least raise the probability of their waking up. A different approach consists in articulating the condition of transmission with a counterfactual condition: according to Menzies (2004), the two facts that 1) the cause "makes a difference" to the effect and that 2) there is a process from cause to effect are both necessary and together sufficient for the existence of a causal relation. Transmission guarantees the existence of a process between *c* and *e* (Menzies' condition 2). The fact that *c* is F makes a difference with respect to the fact that *e* is G, to the extent that, if *c* had not been F (if the doorbell had made no sound), *e* would not have been G (Peter would not have waken up) (Menzies' condition 1).

3 The ordinary concept of transmission being causal, the transference approach seems condemned to circularity. However, circularity can be avoided by redefining the concept of transmission. Given two distinct spatio-temporal regions x and y, a quantity

45 See Collins et al. (2004), p. 14.

46 Roughly, a first set of properties (or predicates) M is said to "supervene" on a second set P if and only if it is impossible that two objects differ with respect to a property of set M, without differing with respect to any property of set P. Physicalism is the doctrine according to which the set of mental properties supervenes on the set of physical properties. The truth of physicalism implies that a person cannot change mentally without changing physically and that there cannot exist a copy (or "clone") of a person p that differs from p mentally without differing from p physically. Several concepts of supervenience have been elaborated. One important difference between them concerns the interpretation of the concept of necessity (or impossibility) that is used in their definition. Cf. Kim (1990) and the introduction to Savellos and Yalcin (1995).

A is transmitted between x and y if and only if A is present both at x and at y.

4 If transmission is construed in this way, causality is not asymmetric. However, it can be argued that the asymmetry of causation is a physical characteristic of causality as it is in the actual world, rather than flowing from a conceptual constraint. Our region of the universe contains a plethora of irreversible processes that are all oriented in the same direction, as is guaranteed by the second law of thermodynamics. Such a physical ground of the asymmetry of causation can also ground the direction of time (Reichenbach 1956; Lewis 1979/1986; Hausman 1998; Savitt 2006).

5 Transmission processes are everywhere. Events that are spatio-temporally sufficiently close to each other are e.g. often linked by transmissions of photons. Therefore, transmission theory seems condemned to lead to an inflation of true causal judgments. A first reply to this objection is that those plethoric causal judgments are true but lack communicational relevance. A second reply is that the relevant causal processes can be chosen on perfectly objective grounds, on the basis of the properties of the effect that is indicated in the *explanandum* of the causal explanation one is looking for. If one asks for the cause of Peter's waking up, the relevant causal process is at the physiological and psychological level and leads to the instantiation of the physiological and psychological properties constitutive of waking up.

6 It has been argued (Curiel 2000, Lam 2005) that the theory of general relativity does not guarantee global energy conservation, so that energy cannot be transmitted. In reply, it may be said that local conservation of energy is sufficient to guarantee the existence of local transmission and local causation, even if it turns out that the applicability of the concept of causation to large scale cosmological events and processes is more restricted than common sense would have expected.

7 Transmission theory seems to be refuted by a much less technical problem: there are many true causal propositions both in common sense and in science where negative facts play the role of causes or effects. Important types of propositions of this sort involve omission or prevention. If I kill a plant by *omitting* to water it, it seems that I have caused its death without having transmitted anything to it[47]. If on the contrary I *prevent* the plant's death by watering it, the event of the plant's death does not take place and cannot therefore be the object of any transmission. Schaffer (2000a) argues that there are many common sense causal propositions bearing on situations in which no transmission seems to be involved. Striking cases are propositions expressing double prevention, in which something or someone prevents the prevention of an event. Schaffer (2006) offers the example of the terrorist who prevents the sentinel in the control tower of the airport from preventing a collision of two airplanes.

Causal propositions in which the cause and/or the effect is/are a negative fact(s) are incompatible with three intuitive properties of causation noted by Hall (2000): a causal process is local (in the sense that the cause is linked to the effect by an intermediate series of events), intrinsic (it does not depend on what happens or is the case elsewhere), and transitive. If *a* can cause *b* by omission, prevention, or double prevention, then certain causal relations obey neither to locality nor to intrinsicality nor to transitivity. Three (incompatible) consequences can be drawn from this.

1 Omissions are not instances of causality although they appear to us as such, e.g. because we tend to conflate causal and non-causal explanation or because we conflate moral responsibility with causality (Dowe 1999; 2000, Armstrong 2004; Beebee 2004; Kistler 2006).

47 The example is Beebee's (2004). More precisely, I do not transmit anything relevant to the plant, although there are no doubt innumerable irrelevant processes linking me to it, such as transmission of photons.

2 Propositions involving omission and prevention can be truly causal, which means that locality, intrinsicality and transitivity are not after all necessary conditions for causation (Schaffer 2000; 2004).

3 There are two concepts of causation or two aspects of the concept of causality: One corresponds to counterfactual dependence (or to probability raising or to nomological dependence), the other corresponds to the existence of a transmission process. According to Hall (2000), these two concepts of causality are even independent of each other.

6. The probabilistic analysis

There are two strategies for discovering laws in general and causal laws in particular on the basis of data bearing on complex situations. The first uses statistical correlations expressed in conditional probabilities that can be found in the data; the second uses controlled experiments. Each of these methods can be used to construct an analysis of causation: the former has inspired the probabilistic analysis of causation that will be discussed presently; in the next section, we will examine the analysis of causation in terms of intervention or manipulation.

In the complex situations explored by such sciences as economics, sociology, epidemiology or meteorology, laws and causal relations do not manifest themselves as exceptionless regularities: not all smokers get lung cancer. In macroeconomics, the so-called Phillips curve represents the dependence between the rate of inflation and the unemployment rate; it implies that the higher the unemployment rate is, the slower is the raise of salaries, and that if on the contrary unemployment is decreasing, salaries and indirectly inflation tend to rise; however, it turns out that that a high unemployment rate can coexist, for quite long periods, with strong inflation.

In the perspective of improving the analysis of causation in terms of regularity, the probabilistic analysis is built on the idea of associating causation with the influence of one factor on a second factor, where this influence need not be universal but must only be statistically significant. The fundamental hypothesis is that factor A has a causal influence on factor B if and only if the probability of B given A is greater than the probability of B given the absence of A.

(PR Probability raising) A is a cause of B if and only if $P(B|A)>P(B|\text{non-}A)$

There are two sorts of motivations for switching from an analysis of causation in terms of universal regularities to an analysis in terms of probability raising. The first reason is that lawful and causal influences are, in complex situations, often masked by other influences and therefore do not manifest themselves in the pure form of a universal regularity, as it happens in the examples just mentioned. The second reason is the hypothesis that there are intrinsically statistical laws, in the sense that, even in a situation in which nothing interferes, some causes only raise the probability of their effects without necessitating them. It is controversial whether there are any laws of this kind outside of quantum physics, but the capacity of the probabilistic analysis to take laws of this kind into account gives it an advantage over analyses of causation in terms of universal regularities.

Two remarks before we consider the development of the fundamental hypothesis (PR). The first is that the probabilistic analysis assimilates ontology to epistemology: the causal relation is identified with what allows us to discover causal influences in complex situations, i.e. the inequality of conditional probabilities. The second is that the probabilistic analysis does not apply – at least not directly – to causal relations and processes between particular events, but only to relations of causal influence between “factors”, properties or types of events. The formalism that is a central part of this approach presupposes that the terms of the

causal relation can be subjected to the operations of propositional logic, such as negation and conjunction. This requires construing the terms of the causal relation as facts (Vendler 1967a, 1967b, Bennett 1988, Mellor 1995) or types of facts rather than as particular events (Davidson 1967).

Condition (PR) is faced with two difficulties that it shares with the DN and the counterfactual account.

1) Probability raising is symmetrical: if A and B are statistically positively correlated, so that P(A|B)>P(A|non-B), it is also true that P(B|A)>P(B|non-A).

2) The effects of common causes are generally statistically correlated although one effect is no cause of the other. If smoking (F) raises both the probability of lung cancer (C) and the probability of heart attack (I), C and I are *ceteris paribus* also positively correlated with each other. One of the reasons of the success of the probabilistic analysis is that this second problem can quite straightforwardly be solved with the condition of the absence of a "screening factor"[48]. If A and B are statistically positively correlated, a third factor C is called a "screening factor" with respect to A and B if the positive correlation between A and B disappears if the probabilities are calculated conditionally on the presence or absence of C. Formally, in such a situation we have P(B|A)>P(B|non-A), but P(B|A & C)=P(B|non-A & C) and P(B|A & non-C)=P(B|non-A & non-C).

The concept of a screening factor can then be used to complete the probabilistic analysis. Factor A, instantiated at instant t, is cause of factor B, instantiated at the same time or later, if and only if two conditions are satisfied:

1 P(B|A) > P(B|non-A)
2 There is no factor C, instantiated at t or earlier, which screens off the correlation between A and B.

This condition solves the problem that positive statistical correlation is in general not *sufficient* for causation, as shown by the correlation between effects of common causes. However, there are also situations in which such a positive correlation is not *necessary* for causation. There are situations in which the presence of factor A, which is a cause of factor B, nevertheless *diminishes* the probability of B. If smokers (M) practice more sport (S) than non-smokers, making M positively correlated with S, it is possible that the beneficial effect of S, which diminishes the risk of cardio-vascular illness (CV), overcompensate for the negative effect of M, which enhances the risk of CV. In such situations a factor M may diminish the probability of its effect CV:

P(CV|M)< P(CV|non-M).

There is a solution to this problem, different versions of which have been proposed by Cartwright (1979, p. 423) and Skyrms (1980). In Cartwright's version, A causes B if and only if the probability of B is higher in the presence of A than in its absence, in all sets that are homogeneous with respect to all causes of B that are not effects of A.

A causes B if and only if P(B|A & Ci) > P(B|¬A & Ci) for all Ci, where Ci are causes of B that are not caused by A.

A "test situation" is characterized by holding fixed the set of factors that cause B but are not caused by A. Insofar as a test situation excludes all indirect causal influence from A on B, it provides a means for evaluating by purely statistical means whether A causes B. This strategy may e.g. justify the intuitive judgment that M causes CV: in a test situation, the conditional probability of CV given M is evaluated within a set of persons who all have the same level of sports practice (S). In such a situation, the probability of CV given M is greater than given not-M.

However, the proposal to analyze the causal influence from A on B in terms of the

48 This concept has been introduced by Reichenbach (1956).

raising of probability in test situations changes the nature of the project of probabilistic analysis. First, in the form proposed by Cartwright and Skyrms, the analysis cannot any more serve as a basis for the reduction of the concept of causality: indeed, the *analysans* essentially contains the concept of cause. In order to determine whether A causes B, it is already required to know all other causes of B, or more precisely all factors that cause B independently of A.

Second, the requirement of measuring conditional probabilities in sets that are homogeneous with respect to all factors that can influence the probability of B but are not correlated with A is incompatible with one of the major motivations of the probabilistic approach: its aim was to provide a method for detecting causal influences in situations where correlation is imperfect, because the presence of interfering factors prevents the universal correlation of cause and effect. However, insofar as intrinsically indeterministic laws are not taken into account, in a situation in which all causes of B that are independent of A are held fixed, if A causes B, P(B|A)=1. Indeed, probabilities lower than 1 measure the net effect of unknown factors that are independent of A and influence B negatively or positively.

We have already mentioned another important problem for the probabilistic analysis: statistical correlation is symmetrical, so that if the probability of B is larger in the presence of A than in its absence, the probability of A is also larger in the presence of B than in its absence. There are several proposals for what should be required in addition to probability raising, in order to distinguish cause and effect. One possibility is to simply stipulate that the factor that is instantiated earlier in time is the cause, and the factor instantiated later, the effect. However, this idea does not fit well with a theory first of all devised for causal relations between general factors, rather than between particular instances of these factors. Moreover, such a stipulation precludes the possibility of so-called backward causation, i.e. causal processes evolving in the direction opposite to the direction of time. Finally, it makes it impossible to reduce the direction of time itself to the direction of causation. A traditional approach to explaining the origin of the asymmetry of time consists in making the hypothesis that it derives from the asymmetry of causation: the fact that instant t_2 is later than instant t_1 is grounded on the fact that an event occurring at t_1 may cause an event occurring at t_2, but that the opposite is not possible[49]. However, the probabilistic analysis can be defended against this objection if the direction of time can be grounded on something other than the direction of causation. According to one hypothesis, the asymmetries of causation and time both derive from the asymmetry of some fundamental physical processes. These are often taken to be thermodynamically irreversible processes, characterizing the evolution of systems whose entropy rises. Other processes that have been suggested as possibly grounding the asymmetry of causation are intrinsically asymmetric microphysical processes, such as the disintegration of K-mesons, or "kaons"[50].

It has also been suggested that the difference between cause and effect might be an effect of the perspective of an observer or human agent, in the sense that, independently of the perspective of the agent, at the level of the objective dependence among factors in the world, causation is symmetric[51].

The most influential proposal to account for the asymmetry of causation in terms of probabilistic conditions is due to Reichenbach (1956) who has suggested using common causes in order to determine the direction of causation (and time). If A and B are positively correlated and if C is a screening factor, such that the correlation between A and B disappears both in the presence and in the absence of C, and such that the presence of C raises both the probability of A and of B, the triplet ACB is called a "conjunctive fork". If the factor C is

[49] This would require some refinement to take account of special relativity.

50 These decomposition processes "violate" the symmetry with respect to temporal inversion ("T"). Cf. Dowe (1992a, p. 189).

51 Fair (1979), Price (1992), Menzies and Price (1993), Price (2007).

instantiated in the past of A and B, and if there is no factor D satisfying the same conditions as C but instantiated in the future of A and B, ACB constitute an open fork in the direction of future (and C is a common cause of the two effects A and B); if the only factor D that satisfies these conditions is instantiated in the future with respect to A and B, ADB constitute an open fork directed towards the past; if finally there is both a factor C in the past and a factor D in the future that satisfy the indicated conditions, ACBD constitute a closed fork. Reichenbach's hypothesis is that the direction from cause to effect (which is also the direction of time) is the direction in which open forks dominate.

Finally, there are numerous attempts to improve the analysis of the notion of causation by a synthesis of conceptual elements of different approaches. One such analysis does so in terms of probabilistic counterfactuals. This theory, suggested by D. Lewis (1986c) and elaborated by Noordhof (1999; 2004), analyzes the causal relation between particular events in the following way: "For any actual distinct events, e_1 and e_2, e_1 causes e_2 iff there are events $x_1,\ldots, x_n$ such that x_1 probabilistically depends upon e_1, …, e_2 probabilistically depends upon x_n." (Noordhof 1999, p. 97) Probabilistic dependence is then analyzed in terms of a counterfactual condition on the chances[52] of the corresponding types of events: "e_2 *probabilistically-depends* on a distinct event e_1 iff it is true that: if e_1 were to occur, the chance of e_2's occurring would be at least x, and if e_1 were not to occur, the chance of e_2's occurring would be at most y, where x is much greater than y." (Noordhof 1999, p. 97)

7. Manipulability and structural equations

One of the most fruitful recent developments in this field is the philosophical analysis of models that have been elaborated in artificial intelligence. The relevant models represent research strategies for analyzing causal structures that are employed in sciences like economics that study causal influences in complex systems. This approach makes use of statistical analysis of conditional probabilities, and in some versions at least (Pearl 2000) analyzes causation in terms of counterfactuals involving experimental interventions or manipulations[53]. As with the probabilistic approach, the analysis of causation in terms of interventions or manipulations is grounded on an analysis of the logic implicit in scientific research on causes. In the social sciences like sociology, economics, and also psychology, the analysis of conditional probabilities is used to extract information on causal influences among different factors. However, in experimental sciences, interventions are a crucial additional method for discovering causal influences. The experimenter manipulates a given "cause" variable under conditions in which other variables are under control, to observe subsequent variation in "effect" variables, which indicates causal influence. Causal graphs and structural equations are formal tools that have been developed to build models of causal structures on the basis of information obtained in this way. The philosophical analysis of such models of the logical form of the scientific research for causes has led to a complete renewal of older philosophical theories of causation in terms of "manipulation" or "intervention".

According to one traditional analysis of causation not yet mentioned so far, a cause C of an effect E is an action that would give a human agent a means to obtain E if she decided to make C happen[54]. However, in this form, such an account suffers from two major defects, circularity and anthropocentrism. The latter is implicit in the thesis that an event can be a

[52] Chances are single-case probabilities, "as opposed to finite or limiting frequencies" (Lewis 1986c, 177/8).

53 Another version has been worked out by Spirtes, Glymour and Scheines (2000). Woodward (2003) has elaborated a philosophical analysis on causation on the basis of the works of Spirtes, Glymour and Scheines (2000) and Pearl (2000). Keil (2000, 2005) has offered an original analysis of causation in terms of manipulation that makes no use of the technical apparatus of structural equations and directed graphs.

54 Cf. Gasking (1955), Menzies and Price (1993).

cause only if its occurrence can be the result of the decision of a human agent. Von Wright (1971) has argued that although the fact that the human capacity to intervene in events in the experimental sciences is indispensable for the analysis of our *knowledge* of causal relations, we should not conclude from this that human action is essential to the *ontology* of causation. It will be shown how recent manipulationist (or interventionist) accounts reply to the objection of anthropocentrism. As for circularity, it seems impossible to build a non-circular analysis of causation that is grounded on the notion of intervention, insofar as an intervention is a causal process. For this reason, recent manipulability theories of causation such as Woodward's (1993) do not aim at a reductionist analysis of the notion of causation, but only at analyzing the logic of causal reasoning in the context of experimental interventions.

Here are some key ideas that structure the approach to causation in terms of interventions, using the formal tools of structural equations and causal graphs. The causal structure of a complex system is represented by a model built from a set of variables V and a set of structural equations that express functional relations among these variables. Let us use Menzies' (2008) analysis of a toy situation often used in the philosophical literature: two kids throw rocks at a bottle to smash it. We have already encountered this situation as an example of preemption: Billy's throw does not smash the bottle although it would have had Sally not thrown her rock an instant earlier, so that it smashed the bottle before Billy's rock could. To represent the relevant actual and possible causal influences in this situation, the following variables can be used. In this case, all variables have only two values ("1" in case the event described by the variable occurs, "0" in case it doesn't), but the formalism can also be used with variables with more than two and also continuous values.

- *BT* = 1 if Billy throws a rock, otherwise BT = 0;
- *ST* = 1 if Sally throws a rock, otherwise ST = 0;
- *BH* = 1 if Billy's rock hits the bottle, otherwise BH = 0;
- *SH* = 1 if Sally's rock hits the bottle, otherwise SH = 0;
- *BS* = 1 if the bottle shatters, otherwise BS = 0.

Each variable is associated to a structural equation. A variable is called "*exogenous*" if its value is determined by factors external to the causal system whose model is being built. In the example, BT and ST are exogenous variables, insofar as their values are not determined by the values of other variables within the model. Therefore, the structural equations for this variables, BT=1 and ST=1, do not contain any other variables, but simply stipulate their values. By contrast, the value of an *endogenous* variable is a function of other variables within the system. The equation for the endogenous variable SH is SH=ST, which means that the value of SH is determined by the value of ST: if Sally throws a rock, the rock reaches the bottle (ST=1 and SH=1) and if she doesn't, the rock doesn't reach the bottle (ST=0 and SH=0). The preemption of the process beginning with Billy's throwing his rock is expressed by the equation for BH: BH=BT & non-SH. Billy's rock reaches the bottle only if 1) Billy throws the rock and if 2) the rock thrown by Sally does not reach it. The variable representing the smashing of the bottle is also endogenous: BS = SH or BH. The bottle gets smashed either if Sally's rock reaches it or if Billy's rock reaches it.

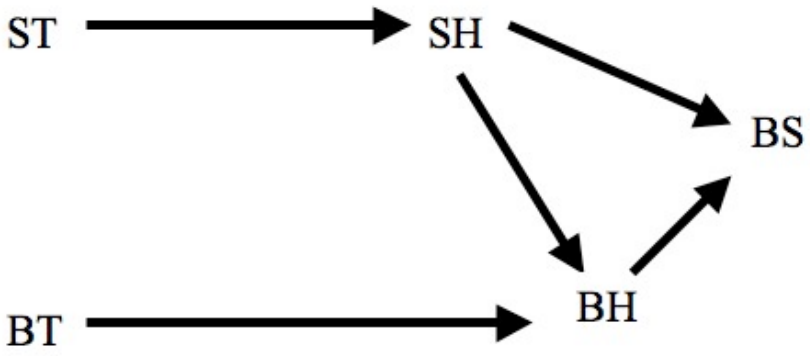

Figure 1 (from Menzies 2008)

The content of a set of structural equations can also be represented in a structural graph. Figure 1 shows a graph representing the structural equations defining our situation: each variable corresponds to a node in the graph. An arrow going from variable X to variable Y represents the fact that the value of Y depends on the value of X; in this case, X is called a "parent" of Y. A *directed path* from X to Y is a set of arrows leading from X to Y. Each arrow and each structural equation represents a set of counterfactual conditionals. Once a model is constructed, it can be used to determine the truth-value of new counterfactuals that do not simply correspond to one arrow. Say we want to know what would have happened if the rock thrown by Sally had not reached the bottle. To find this out, one sets the variable corresponding to the antecedent of the counterfactual to the value it would have if the antecedent were true. In this case, one sets SH=0. This represents an "atomic intervention" (Pearl 2000, p. 70). It is equivalent to what Lewis calls a "miracle": One does not take into consideration the past that might have led to the truth of the antecedent. Rather, the value of the antecedent (here, SH) is set while the values of all variables corresponding to the past of the antecedent keep the values they have in actuality. In the graphical representation, this means that all arrows leading to the variable SH are erased, which is equivalent to transforming SH into an exogenous variable. In the manipulationist interpretation of this formalism, this corresponds to a localized experimental intervention on variable SH, which comes from outside the system and is direct in the sense that it is not obtained indirectly by intervening on factors that influence SH within the system. As with Lewis' concept of a miracle, this guarantees that no "backtracking" counterfactual can be true. When the value of variable X is modified, the variables situated in the past of X are left untouched. In the standard representation, these are the variables figuring at the left of X. The values that the variables to the right of X take in a situation in which X takes the stipulated value can then be determined on the basis of the equations corresponding to the arrows starting at X.

Pearl (2000, p. 70) defines the causal effect of X on Y, written "P(y/do(x))", as the probability distribution of the different values y of Y, given that an intervention ("do") has fixed x as the value of variable X. This has the consequence that all factors different from X that also influence Y are included in X's impact on Y. To avoid this result, Woodward (2003) imposes additional constraints on interventions I appropriate to determining whether X causes Y. 1) I must be the only cause of X, in the sense that all other influences on X must be cut. 2) I must not cause Y through any paths that do not go through X. The administration of a placebo pill in the following situation does not fulfill this condition. I is the ingestion of the pill; X is the action of the pill on the body after its ingestion; Y is recovery. By the definition of a placebo, if I is efficacious in changing the value of Y, its influence does not flow through X, i.e. changes in the body brought about by the absorption of the pill. In such a situation, the fact that I influences Y does not mean that X causes Y. 3) I must not be the effect of a cause

that influences Y through any path that does not go through X. If, in order to find out whether the indication X of a barometer causes the thunderstorm Y, my interventions I on X depend on (my knowledge of) air pressure, then Y may vary as a function of the values that I imposes on X, whereas X does of course not cause Y. 4) The values of all possible causes of Y that are not situated on a path from I through X to Y must be held fixed.

In this context, Woodward (2008) defines the causal effect of X on Y by the difference of the values of Y that corresponds to the difference between two values x and x* of the variable X, on which one intervenes via I.

(CE) ("causal effect") $Y_{do(x),\ Bi} - Y_{do(x^*),\ Bi}$

where "$Y_{do(x),\ Bi}$" represents the value of the variable Y given that an intervention has set variable X to value x, in circumstances B_i.

If the relation between X and Y is deterministic, X is a cause of Y if and only if there are pairs of values x and x* ($x^* \neq x$), such that (CE) differs from zero; if the relation is indeterministic, X is a cause of Y if and only if there are pairs of values x and x*, such that there are values of Y whose probability is different for the two values of X.

The structural equations model shares with the counterfactual analysis the idea that causation must be defined in models in which the past corresponds to actuality but the putative cause has a counterfactual value; however, it avoids at least some of the counterexamples to the counterfactual analysis. Thus, it yields the intuitively correct result in the preemption case considered above.

In figure 1, BH is the only intermediate variable between ST and BS that is not on the path ST – SH – BS. Thus, in order to judge whether ST causes BS, BH must be held fixed at its actual value BH=0. If one considers the counterfactual situation in which the value of the putative cause ST is changed so as to become ST=0, the value of BS determined by the equations also differs from its actual value, to become BS=0. This means that ST causes BS.

Lewis' analysis fails to yield the correct result in this case because BS does not counterfactually depend on its cause ST, because BS=1 even if ST=0. The interventionist analysis avoids this difficulty by "freezing" on their actual values all variables that are not on the path connecting the putative cause to its putative effect.

Here is an interpretation of this formal difference. In the structural equations model, the antecedent can be taken to represent, not a fact in a different possible world, but a situation resulting from an experimental intervention. In Lewis' analysis, the evaluation of a counterfactual requires holding fixed all events in the past of the putative cause (described by the antecedent of the counterfactual), whereas the structural equations model requires holding fixed the values of all variables that are not situated on the path between the putative cause and effect. This difference has formal consequences: Lewis' analysis makes causation transitive, whereas it isn't necessarily transitive in the structural equations model[55].

The structural equations model provides the means of distinguishing different causal notions that can all be expressed by the common sense word "cause". The fact that it allows defining different causal notions shows the fecundity of this approach, although it cannot provide a non-circular analysis of causation. A variable X can influence another variable Y in two independent ways in such a way that these influences cancel each other out. Starting the engine of a car X raises the temperature of the engine Y[56], but X also causes the onset of the ventilation system Z, which lowers the temperature of the engine. It is possible that the positive direct influence from X on Y is exactly compensated by the negative influence of X on Y via Z, such that X has zero net influence on Y. In such a case, it seems both intuitively correct to say that starting the engine raises the temperature of the engine and that it does not.

---

55 Cf. Hitchcock (2001).

[56] Hesslow (1976) gives a structurally similar example.

However this involves no paradox insofar as the two judgments contain different notions of causation[57] that are expressed by the same common sense term. The former is correct if "raises" is taken to express the concept of being a *contributing* cause, the latter is correct if "raises" is taken to express the concept of being a *total* cause.

In the situation sketched, X is not a "total cause" of Y, as defined by condition (CE) above. However, X is a "contributing cause":

(CC) X is a contributing cause of Y if and only if the value of Y changes as a consequence of a change of the value of X, where the values of all variables different from X and Y are held fixed, and in particular those that lie on paths between X and Y.

Indeed, if we hold Z in our example fixed, we find that an intervention on X modifies the value of Y, so that starting the engine is a contributing cause of the rise of temperature of the engine, although applying condition (CE) shows that it is not a total cause of the rise of temperature: if Z is not held fixed, starting the engine does not make the temperature rise.

Older versions of the manipulability theory make the judgment "X causes Y" depend on the possibility of acting on X. This seems to make it impossible to apply the concept of causation to events that are in principle outside the sphere of influence of human interventions. However, eruptions of volcanoes and explosions of supernovae seem to be causes although no possible human action could ever bring them about or modify them. This problem is solved in recent versions of the interventionist analysis, in which the notion of intervention is defined without any reference to human action. Analyses of causation in terms of structural equations and directed graphs avoid anthropocentrism because the intervention that sets the value of the putative cause is no longer required to be the result of a human action. Natural events entirely independent of all intentional actions can satisfy the formal conditions on an intervention modifying the value of the putative cause. Such a "natural experiment" provides just as good a basis for judging causal influence as intentional interventions by human experimenters. Neuropsychology is one important field of research where hypotheses on the causal influence of the activation of specific brain regions are evaluated by such "natural experiments": the hypothesis that the activation of brain region X causally influences the activation of brain region Y is confirmed by the observation that a modification of X due to accident or illness is systematically followed by a modification of Y.

However, there seem to be causal relations on which even interventions as defined by these new theories seem to be impossible. To judge whether the gravitational attraction of the moon causes the tides, one must examine the consequences of an intervention on the position or the mass of the moon. It can be doubted whether "interventions" on the moon are physically possible: such an intervention would require modifying the position or the mass of the moon by some means that does not also *directly* influence the tides.

8. Conclusion

Philosophical research on causation has developed into a rich and complex field. Since the once dominant deductive-nomological analysis has been abandoned, several alternative approaches based on very different premises have been developed. Each can claim a certain extent of success insofar as it can account for intuitions or alleged facts about causation that provide counterexamples against rival accounts. But each also has its own counter-examples. The confusion that threatens can be greatly diminished by realizing that different approaches do often not share their goals. Most traditional philosophical analyses pursue the aim of a priori conceptual analysis, whereas others, such as analyses in terms of manipulability or process theories take as their criterion of success fit with the logic of scientific research about

57 Cf. Woodward (2003, p. 50sq.).

causal relations or with the structure of reality as described by present-day physics. Although these aims may seem incompatible, there are also efforts to construct a synthetic theory that can preserve what is correct from several seemingly incompatible theories. According to one hypothesis of this sort, different theories are applicable to different domains of phenomena and of scientific research. The probabilistic analysis may provide an adequate analysis of causal judgments in economy and other social sciences, whereas theories in terms of transmission processes and conserved quantities may be adequate for physical causation. The counterfactual conception may seem most adequate to account for common sense causal judgments. Such a "regionalist" conception is not the only form of pluralism or relativism, according to which there is more than one concept of causation[58]. Something may be a cause in the sense of one of these concepts, without being a cause in the sense of others. In the counterfactual sense, the rock thrown by Sally is not the cause of the bottle's breaking because the breaking does not depend on Sally's action. It would have broken anyway in the context of the backup cause constituted by Billy's well-aimed throw. However, in the sense of causation as a physical process, Sally's throwing her rock does cause the bottle's breaking. More ambitious syntheses aim at constructing a unified theory that can account for all situations, making use of conceptual ingredients taken from different (and incompatible) theories. Examples are the probabilistic counterfactual analysis (Noordhof 1999), and the theory according to which causation can be analyzed in terms of the raising of the probability of a process (Schaffer 2001). The conception of functional reduction provides another framework for a synthetic account. According to this approach, causation is a concept whose conditions of application are in part *a priori* and in part *a posteriori*. It applies to causation a two-stage model of reduction devised by Armstrong (1968) and Lewis (1972) to solve the mind-body problem. The first step of pure *a priori* conceptual analysis aims at making explicit the "functional profile" of a given concept: these are the constraints that determine the set of objects to which the concept applies. To use one of the paradigmatic examples from the mind-body problem, pain is a state of a subject A that is caused by damage to the body of A and causes characteristic mental states and behavior, such as the desire that the pain ceases and actions aiming at interrupting or diminishing what causes the damage. This first conceptual step of the analysis is independent of empirical research and aims at the *a priori* conditions of application of the concept. The second step aims at discovering those natural objects, states, or processes that possess, in the actual world, the functional profile found in the first step. For cognitive concepts such as pain, it is conceivable that one finds different natural states or processes occupying a given functional role in different cognitive systems, e.g. animals of different species. If this is correct, there would be a general concept of pain although the concept applies to different types of states or processes in animals of different kinds.

Applying this strategy to the analysis of causation, it is conceivable that different sorts of natural relations or processes play the role of causation in different fields. In this way one is led to a pluralist conception, in which it would be coherent to judge e.g. that probability raising occupies the conceptual role of causation in epidemiology and economy, that counterfactual dependence occupies it in the explanation of human actions, that the existence of a mechanism plays the role of causation in biology, whereas the existence of a transmission process plays the role in physics. There would be both a general concept of causation corresponding to a priori conceptual constraints, such as spatio-temporal distinctness of cause and effect and asymmetry, and "regional" concepts of causation, specific to different domains of inquiry[59].

[58] Hitchcock (2007a) provides a classification of types of pluralism about causation.

[59] I thank Andrew McFarland for helpful comments and for improving the language of this chapter.

References

Achinstein, Peter (1975), Causation, Transparency, and Emphasis. *Canadian Journal of Philosophy* 5, p. 1-23.
Aronson J.J. (1971), The Legacy of Hume's Analysis of Causation, *Studies in the History and Philosophy of Science* 2, p. 135-165.
Armstrong, David M. (1968), *A Materialist Theory of Mind*, revised ed., London: Blackwell, 1993.
Armstrong, David M. (2004), *Truth and Truthmakers*, Cambridge: Cambridge University Press.
Bennett, Jonathan (1987), Event Causation: the Counterfactual Analysis, *Philosophical Perspectives* 1, p. 367-368, repr. in Sosa et Tooley (1993), p. 217-233.
Bennett, Jonathan (1988), *Events and Their Names*, Hackett, Indianapolis/Cambridge.
Carnap, Rudolf (1966/1995), *An Introduction to the Philosophy of Science*, New York, Dover Publications. Original edition *Philosophical Foundations of Physics: An Introduction to the Philosophy of Science*, New York, Basic Books, 1966.
Cartwright, Nancy (1979), Causal Laws and Effective Strategies, *Noûs* 13, p. 419-427; repr. In *How the Laws of Nature Lie*, Clarendon Press, Oxford, 1983.
Cartwright, Nancy (1999), *The Dappled World, A Study of the Boundaries of Science*, Cambridge, Cambridge University Press.
John Collins, Ned Hall, and L. A. Paul (éds.), *Causation and Counterfactuals*. Cambridge (Massachusetts), MIT Press, 2004.
Cummins, R. (2000), How Does it Work? Vs. What Are the Laws? Two Conceptions of Psychological Explanation, in F. Keil & R. Wilson (eds.), *Explanation and Cognition*, Cambridge, MA: MIT Press, p. 117-45.
Curiel, Erik (2000), The Constraints General Relativity Places on Physicalist Accounts of Causality, *Theoria* (San Sebastian) 15, p. 33-58.
Davidson, Donald (1967), Causal Relations, in Davidson D. (1980), *Essays on Actions and Events*, Oxford, Clarendon Press 1980.
Dieks D. (1986), Physics and the Direction of Causation, *Erkenntnis* 25, p. 85-110.
Dowe, Phil (1992), Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory, *Philosophy of Science* 59, p. 195-216.
Dowe, Phil (1992a), Process Causality and Asymmetry, *Erkenntnis* 37, p. 179-196.
Dowe, Phil (2000), *Physical Causation*, Cambridge, Cambridge University Press, 2000.
Elga, Adam (2000), Statistical Mechanics and the Asymmetry of Counterfactual Dependence, *Philosophy of Science*, Supp. Vol. 68, PSA 2000, p. 313-324.
Fair D. (1979), Causation and the Flow of Energy, *Erkenntnis* 14, p. 219-250.
Faye J. (2010), Backward Causation, *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/causation-backwards/
Frisch, Mathias (2009a), "The Most Sacred Tenet?" Causal Reasoning on Physics, *British Journal for the Philosophy of Science* 60, p. 459-74.
Frisch, Mathias (2009b), Causality and Dispersion: A Reply to John Norton, *British Journal for the Philosophy of Science* 60, p. 487-95.
Gasking D. (1955), Causation and Recipes, *Mind* 64, pp. 479-487.
Ned Hall (2004a), Two Concepts of Causation, in Collins et al. (2004), p. 225-276.
Hall, Ned (2000/2004b), Causation and the Price of Transitivity, in Collins et al. (2004), p. 181-204.
Hausman, Daniel (1998), *Causal Asymmetries*, Cambridge, Cambridge University Press.
Hesslow, Germund (1976), Two Notes on the Probabilistic Approach to Causality, *Philosophy of Science* 43, p. 290-2.

Hitchcock, Christopher (1996a), The Role of Contrast in Causal and Explanatory Claims, *Synthese* 107, p. 395-419.
Hitchcock, Christopher (1996b), Farewell to Binary Causation, *Canadian Journal of Philosophy* 26, p. 335-364.
Hitchcock, Christopher (2001), The Intransitivity of Causation Revealed in Equations and Graphs, *Journal of Philosophy* 98, p. 273-99.
Hitchcock, Christopher (2007), Prevention, Preemption, and the Principle of Sufficient Reason, *Philosophical Review* 116, p. 495-532.
Hitchcock, Christopher (2007a), How to Be a Causal Pluralist, in Peter Machamer and Gereon Wolters (eds.), *Thinking about Causes*, Pittsburgh, University of Pittsburgh Press, p. 200-221.
Hume, David (1739-40), *Treatise of Human Nature*, L.A. Selby-Bigge et P.H. Nidditch (eds.), Oxford, Clarendon Press, 1978.
Hume, David (1777), *Enquiries Concerning Human Understanding and Concerning the Principles of Morals,* L.A. Selby-Bigge et P.H. Nidditch (eds.), Oxford, Clarendon Press, 1975.
Keil, Geert (2000), *Handeln und Verursachen*, Frankfurt a.M., Vittorio Klostermann.
Keil, Geert (2005), How the *Ceteris Paribus* Laws of Physics Lie, in: Jan Faye et al. (eds.), *Nature's Principles*, Dordrecht, Kluwer.
Kim, Jaegwon (1973), Causes and Counterfactuals, *Journal of Philosophy* 70, p. 570-2.
Kim, Jaegwon (1990), Concepts of Supervenience, repr. In Kim, Jaegwon, *Supervenience and Mind*, Cambridge, Cambridge University Press, 1993.
Kistler, Max (1998), Reducing Causality to Transmission, *Erkenntnis* 48 (1998), p. 1-24.
Kistler, Max (1999/2006), *Causation and Laws of Nature*, Londres, Routledge, 2006; translation of *Causalité et lois de la nature*, Paris, Vrin, 1999.
Kistler, Max (2001), Causation as transference and responsibility", *in* Wolfgang Spohn, Marion Ledwig & Michael Esfeld (eds.), *Current Issues in Causation*, Paderborn, Mentis, p. 115-133.
Kistler, Max (2006), La causalité comme transfert et dépendance nomique, *Philosophie* 89, p. 53-77.
Krajewski, W. (1982), Four Conceptions of Causation, in: W. Krajewski ed., *Polish Essays in the Philosophy of the Natural Sciences*, Reidel, Dordrecht, 1982.
Lam, Vincent (2005), Causation and space-time, *History and philosophy of the life sciences* 27, p. 465-478.
Lewis, David (1972), Psychophysical and Theoretical Identifications, *Australasian Journal of Philosophy* 50, p. 249-258, repr. in : David Chalmers (éd.), *Philosophy of Mind: Classical and Contemporary Readings*, New York, Oxford University Press, 2002, p. 88-94.
Lewis, David (1979/1986), Counterfactual Dependence and Time's Arrow, with Postscripts, in *Philosophical Papers, vol. II,* New York, Oxford University Press, p. 32-66.
Lewis, David (1986a), Events, in *Philosophical Papers, vol. II,* New York, Oxford University Press, p. 241-269.
Lewis, David (1986b), Causation, in *Philosophical Papers, vol. II,* New York, Oxford University Press, p. 159-172.
Lewis, David (1986c), Postscripts to "Causation", in *Philosophical Papers, vol. II,* New York, Oxford University Press, p. 172-213.
Lewis, David (2000), Causation as Influence, in Collins, Hall and Paul eds. (2004), p. 75-106.
Mackie John L. (1974), *The Cement of the Universe*, Oxford, Clarendon Press.
Maslen, Cei (2004), Causes, Contrasts, and the Nontransitivity of Causation, in Collins et al. (2004), p. 341-357.
McDermott, Michael (1995), Redundant Causation, *British Journal for the Philosophy of Science* 46: 523-544.

Menzies, Peter (2004), Difference-making in Context, in Collins, Hall and Paul (eds) (2004), p. 139-180.
Menzies, Peter (2008), Counterfactual Theories of Causation, in *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/causation-counterfactual/.
Menzies, Peter and Price Huw (1993), Causation as a Secondary Quality, *British Journal for Philosophy of Science* 44, p. 187-203.
Noordhof, Paul (1999), Probabilistic Causation, Preemption and Counterfactuals, *Mind* 108, p. 95-125.
Noordhof, Paul (2004), Prospects for a counterfactual theory of causation, in Phil Dowe et Paul Noordhof (eds.), *Cause and Chance*, Routledge, 2004, p. 188-201.
Norton, John (2003), Causation as Folk Science, *Philosopher's Imprint* 3 (www.philosophersimprint.org/003004/), repr. in H. Price and R. Corry (eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, Oxford: Clarendon Press, 2007, p. 11-44.
Norton, John (2009), Is There an Independent Principle of Causality in Physics?, *British Journal for the Philosophy of Science* 60, p. 475-86.
O'Leary, John et Price, Huw (1996), How to Stand Up for Non-Cognitivists, *Australasian Journal of Philosophy* 74, p. 275-292.
Paul, L.A (2004), Aspect Causation, in Collins et al. (2004), p. 205-224.
H. Price et R. Corry (eds) (2007), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, Oxford, Clarendon Press.
Price, Huw (1992) Agency and Causal Asymmetry, *Mind* 101, pp. 501-520.
Price, Huw (2007), Perspectival Causation, in H. Price et R. Corry (eds) (2007), p. 250-292.
Pearl, Judea (2000), *Causality. Models, Reasoning, and Inference*, Cambridge, Cambridge University Press.
Popper, Karl R. (1935/2002), *The Logic of Scientific Discovery*, London and New York: Routledge. Original edition: *Logik der Forschung*, Vienna: Julius Springer, 1935.
Popper, Karl R. (1956), *The Arrow of Time*, Nature 1977, p. 538.
Putnam, Hilary (1984), Is the Causal Structure of the Physical Itself Something Physical ?, repr. in H. Putnam, *Realism with a Human Face*, ed. by J. Conant, Cambridge, MA, Harvard University Press, 1990.
Reichenbach, Hans (1956), *The Direction of Time*, Berkeley, Univ. of California Press, 1991.
Russell, Bertrand (1912), On the Notion of Cause, *Proceedings of the Aristotelian Society*, 13 (1912-13) et *Scientia (Bologna)*, 13 (1913); repr. in *Mysticism and Logic* (1917), repr. Londres, Routledge, 2004 and in *The Collected Papers of Bertrand Russell, vol. 6 : Logical and Philosophical Papers 1909-13*, John G. Slater (éd.), Londres et New York, Routledge, 1992, p. 193-210.
Russell, Bertrand (1914), *Our Knowledge of the External World*, Londres, Routledge, 1993.
Russell, Bertrand (1948), *Human Knowledge, Its Scopes and Limits*, Londres, Routledge, 1992.
Salmon, Wesley (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton University Press.
Salmon, Wesley (1994), Causality Without Counterfactuals, *Philosophy of Science* 61, p. 297-312.
Salmon, Wesley (1998), *Causality and Explanation*, New York, Oxford: Oxford University Press.
Savellos, Elias E. et Yalçin, Ümit D. (eds.), *Supervenience: New Essays*, Cambridge, Cambridge University Press, 1995.
Savitt, Steven (ed.) (2006), Studies in History and Philosophy of Modern Physics 37, no. 3 :

"The arrows of time", p. 393-576.
Schaffer, Jonathan (2000), Trumping Preemption, *Journal of Philosophy* 97, p. 165-181, repr. in Collins et al. (2004), p. 59-73.
Schaffer, Jonathan (2000a), Causation by Disconnection, *Philosophy of Science* 67, p. 285-300.
Schaffer, Jonathan (2001), Causes as Probability-Raisers of Processes, *Journal of Philosophy* 98, p. 75-92.
Schaffer, Jonathan (2005), Contrastive Causation, *Philosophical Review* 114, p. 297-328.
Schaffer, Jonathan (2006), Le trou noir de la causalité, *Philosophie*, no. 89 (2006).
Smith, Sheldon (2002), Violated Laws, *Ceteris Paribus* Clauses and Capacities, *Synthese* 130, p. 235-264.
Spirtes, Peter, Glymour, Clark and Scheines, Richard (2000), *Causation, Prediction and Search*, Second edition, Cambridge (Mass.), MIT Press.
D. Spurrett et D. Ross (2007), Notions of Cause : Russell's Thesis Revisited, *British Journal for the Philosophy of Science* 58, p. 45-76
Vendler Z. (1967a), Causal Relations, *Journal of Philosophy* 64, p. 704-713
Vendler Z. (1967b), Facts and Events, in: *Linguistics and Philosophy*, Ithaca, N.Y., Cornell University Press.
Winston, Mark, L. (1991), *The Biology of the Honey Bee*, Cambridge (MA), Harvard University Press.
Wittgenstein, Ludwig (1921), *Tractatus logico-philosophicus*, trad. C. K. Ogden and Bertrand Russell, London, Routledge, 2001.
von Wright G.H. (1971), *Explanation and Understanding*, Ithaca, N.Y., Cornell University Press.
Woodward, James (2003), *Making Things Happen: a Theory of Causal Explanation*, Oxford, Oxford University Press, 2003.
Woodward, James (2004), Counterfactuals and Causal Explanation, *International Studies in the Philosophy of Science* 18 (2004), p. 41-72.
Woodward, James (2008), Causation and Manipulability, *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/causation-mani/