Causation in contemporary analytical philosophy

Max Kistler

Max Kistler

Causation in contemporary analytical philosophy

At the beginning of the 20th century, Bertrand Russell, in his famous paper *On the Notion of Cause*, pleaded for the «complete extrusion [of the word 'cause'] from the philosophical vocabulary»[1]. He gave three arguments for the thesis that «the reason that physics has ceased to look for causes is that, in fact, there are no such things» and that the whole conceptual apparatus of causation is «a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm»[2].

Russell's first argument is that the idea, inherent in the concept of causation, of an intrinsic difference between an active cause and a passive effect, is anthropomorphic. This makes the concept inadequate for the analysis of natural causal relations, with the exception of human actions. «Nothing can operate except a volition. The belief that causes "operate" results from assimilating them, consciously or unconsciously, to volitions»[3].

His second argument will prove particularly important in what follows. The concept of causation is associated with the "principle of causality" according to which causality is the source of regularities in nature. (For Hume, it even consists fundamentally in regularities.) The principle of causality says that the same cause always has the same effect. Now, it is essential to the concept of causation to be applicable to macroscopic events such as car accidents, avalanches, or deaths. The formidable complexity of such events contrasts with the simplicity

---

[1] B. RUSSELL, *On the Notion of Cause* [1912], in B. RUSSELL, *Mysticism and Logic* [1917], Routledge, London 1986, 173.

[2] RUSSELL, *On the Notion…*, 173.

[3] RUSSELL, *On the Notion…*, 183. Russell continues to condemn this anthropomorphic aspect of the common notion of causation, when he undertakes, two years later, to rehabilitate the concepts of causation and of causal law. «Historically, the notion of cause has been bound up with that of human volition. [...] The cause is supposed to be "active", the effect "passive". [...] But all such ideas, as applied to physics, are mere anthropomorphic superstitions» (B. RUSSELL, *Our Knowledge of the External World* [1914], Routledge, London 1993, 227).

of the conditions that are the object of the laws of physics. According to Russell, the requirements of applicability to complex macroscopic events and of following the strict regularity of a law of nature, are incompatible, although both are central to causation. The more the description of a complex event involves sufficient detail and completeness to allow the application of physical laws, the less it is possible that the event will ever recur. However, without recurrence, the idea of regularity is unintelligible[4]. This blocks the possibility to justify the scientific legitimacy of the concept of causation by its link to lawful regularity via the principle of causality, which might have appeared promising, given that modern science has substituted the search for laws for the search for causes.

Russell's second argument establishes that there are no regularities at the level of concrete events. His third argument goes a step further. It says that in 20th century physics, there are not any laws of succession at all, which might have played the role of grounding causation. The progress of physics consists in replacing them by laws of coexistence or by equivalences or functional dependencies expressed in equations. This thesis about the form of laws in mature science explains according to Russell the failure of the project to make the concept of causation respectable by reducing it to nomological regularity. According to the nomological conception of causation[5], two events c and e are linked as cause and effect, if and only if c and e are instances of types of events, such that there is a law that all events of type c are followed by an event of type e. However, according to Russell, the functional laws of contemporary physics are incapable of playing this role. We can reconstruct Russell's argument as a reductio. The concept of causation has two essential aspects. First, the cause is temporally separate from the effect and precedes it. Second, there is an asymmetry between cause and effect in that the former determines or produces the latter, but not vice versa, and this is intrinsic, i.e. not relative to a description of these events. The reduction of causation to regular succession according to law of nature can only use functional laws of coexistence, for mature physics has no other laws. However, this eliminates the two characteristic traits of causation, for the terms of functional laws are not separate in time, and there is no intrinsic asymmetry between them. Hence the reduction of causation to nomological regularity fails. Given this conclusion, the role of Russell's second argument is to give an addi-

---

[4] See RUSSELL, *On the Notion...*, 181. Nancy Cartwright has more recently taken up this observation of Russell's, to draw the opposite conclusion from Russell's own. «I hold just the reverse of Russell's view. I am in favour of causes and opposed to laws». N. CARTWRIGHT, *How the Laws of Physics Lie*, Clarendon Press, Oxford, 1983, 74.

[5] Kant has offered a transcendental justification of the grounding of causation in lawful regularity. Recently, Davidson (D. DAVIDSON, *Laws and cause*, «Dialectica», 49 (1995), 263-279) has undertaken to reformulate a transcendental justification which is Kantian in spirit.

tional reason for why there can be no laws of succession between complex events. It is important for Russell's demonstration of the bankruptcy of the concept of causation, to show that the absence of such laws is no historically contingent fact, but has a conceptual ground, namely the incommensurability of the terms of quantitative laws (a restricted number of measurable quantities) and the terms of causal relations (concrete events with an infinite number of properties).

1. The absorption of causation in the model of deductive-nomological explanation

Russell's arguments have not succeeded in convincing the philosophical community of the need to eliminate the concept of causation from a scientifically informed world view. Many continue to think that the concept of causation is necessary to understand not only our interaction with the environments of everyday life, but also science.

Nancy Cartwright accepts Russell's thesis that «causal principles cannot be derived from the causally symmetric laws of association»[6], but she objects to his thesis that «laws of association are all the laws there are»[7]. Her reason is the following: We are capable of intervening in the world. This capacity presupposes that those interventions abide by nomological constraints, or laws of nature, that we are able to know, at least approximately. Moreover, science allows us to develop and refine our «effective strategies»[8] of intervention, and above all to explain why they work.

Hilary Putnam has also drawn attention to the indispensability of the concept of causation for the analysis of scientific experimental and explanatory practice, and not only in sciences like biology that have not completely accomplished the quantitative formulation of all their laws. We need it for physics too, when it comes to understanding the interaction of concrete and complex systems:

«When we come to sciences less fundamental than fundamental particle physics, say sociology or history, or even evolutionary biology, or even to chemistry or solid-state physics, then, of course, we find that causation-as-bringing-about is invoked constantly. [...] Even in fundamental physics such talk becomes indispensable when we *apply* the physics to actual systems that are (of course) *not* the whole universe-regarded-as-a-closed-system»[9].

[6] CARTWRIGHT, *How the Laws…*, 21.

[7] CARTWRIGHT, *How the Laws…*, 21.

[8] CARTWRIGHT, *How the Laws…*, 21.

[9] H. PUTNAM, *Is the Causal Structure of the Physical Itself Something Physical?*, in H. PUTNAM, *Real-*

Even Russell himself, only two years after his radical critique of the concept of cause, partly rehabilitated it[10], however, after cutting out the anthropomorphic aspects of the common sense notion of cause. Science provides no grounding for the existence in nature of a compulsion or necessity, in which the cause produces the effect. The idea of such a necessity has its only origin in the experience of our own voluntary acts or acts of volition[11]. We must also refrain from thinking of the cause as being "active" and of the effect as being "passive"; furthermore, we must eliminate from the concept of causation «everything that differentiates between past and future»[12]. However, says Russell, we need the notion of causation in order to understand the evolution of physical systems that are «less than the whole state of the universe»[13], which is of course all we ever try to do in practice. According to the position articulated by Russell in 1914, we can know «that certain constant relations are observed to hold [...] between events of specified kinds with given intervals of time between them»[14], and he now thinks that such constant relations between events can be considered as «causal laws»[15].

The reintegration of the notion of causation within the conceptual apparatus of the philosophy of science has found its canonical expression in the assimilation of causation to the relation of explanation, conceived according to the so-called 'deductive-nomological' model. This conception of causation belongs to the core of the doctrine of logical empiricism. As such, it was the standard view on the nature of causation from the 1930's to the 1970's when criticism and alternative proposals began to proliferate. I shall present this conception in some detail because all later alternative conceptions are motivated by the attempt to

ism with a Human Face, J. CONANT (ed. by), Harvard University Press, Cambridge, MA 1990, 86; italics Putnam's.

Since then, Putnam has changed his mind about the objective reality of causation. He now regards it as an effect of perspective that makes its appearance relatively to our interest of searching explanations, which is governed by a pragmatic logic. See H. PUTNAM, *Renewing Philosophy*, Harvard University Press, Cambridge, MA 1992, 47ff.

[10] Russell maintains his thesis that «in a sufficiently advanced science, the word 'cause' will not occur in any statement of invariable laws». (RUSSELL, *Our Knowledge...*, 223). However, he now acknowledges 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, *Our Knowledge...*, 223).

[11] Cfr. RUSSELL, *Our Knowledge...*, 229.

[12] RUSSELL, *Our Knowledge...*, 227.

[13] RUSSELL, *Our Knowledge...*, 230.

[14] RUSSELL, *Our Knowledge...*, 230.

[15] RUSSELL, *Our Knowledge...*, 230.

overcome the difficulties encountered by the logical empiricist's absorption of the notion of causation in the deductive-nomological model of explanation[16].

According to the model of scientific explanation put forward by Hempel[17] and developed by Hempel and Oppenheim[18], an explanation is a deductive argument. Its conclusion is called the "explanandum" – that which needs to be explained –, its premises constitute the "explanans" – that which explains. Part of the premises consists in «statements of antecedent conditions» (often called "initial conditions"), and part in "General Laws", whereas the conclusion contains a «description of the empirical phenomenon to be explained»[19]. An adequate explanation must satisfy the following requirements: the explanandum must be a logical consequence of the explanans, for only the logical necessitation of the conclusion of a valid argument by its premises can sufficiently "ground" (our belief in the truth of) the explanandum. Furthermore, the explanans must contain an essential reference to general laws. In other words, an explanation is not adequate if it is possible to deduce the explanandum from the particular premises alone. However, Hempel and Oppenheim explicitly acknowledge the legitimacy of purely general explanations, in which the explanans contains only general laws and which have a law as their explanandum. This is meant to allow integrating in the model the derivation of laws from other, more general laws, such as the derivation of the general form[20] of Kepler's laws from Newton's laws. The deductive-nomological (from now on "D-N") model is only intended to cover the explanation of empirical phenomena. Therefore, it requires the explanans to have empirical content, in other words it must be capable of being tested by observation or experimentation. Finally, an explanation can only be correct if all statements in the explanans are true. This implies that we have to abandon the aim to know with certainty whether an explanation is correct. The problem of induction[21], i.e. the fact that inductive arguments are never logically valid, condemns us to construct explanations with hypothetical laws whose truth will for logical reasons remain forever uncertain.

[16] On the D-N model of explanation and its relation to causation, see M. KISTLER, *Causalité et lois de la nature*, Vrin, Paris 1999, chap. 5, and A. BARBEROUSSE, M. KISTLER et P. LUDWIG, *La philosophie des sciences au XXe siècle*, Flammarion, Paris 2000, chap. 5.

[17] C.G. HEMPEL [1942], *The Function of General Laws in History*, repr. in C.G. HEMPEL, *Aspects of Scientific Explanation*, The Free Press, New York 1965, 231-243.

[18] C.G. HEMPEL and P. OPPENHEIM, *Studies in the Logic of Explanation*, repr. in HEMPEL, *Aspects…*, 245-290.

[19] HEMPEL and OPPENHEIM, *Studies…*, 249.

[20] Kepler's laws mention the sun, which is an individual object. Hence, these laws cannot be derived without particular premises, such as a statement of the sun's mass. However, the general laws describing the relative movement of two masses can be derived from general premises only.

[21] See BARBEROUSSE, KISTLER, LUDWIG, *La philosophie…*, chap. 2.

According to Hempel and Oppenheim, any explanation that satisfies the requirements of the D-N model is *ipso facto* a causal explanation. Hence, it is possible to use the model to analyse the causal relation: the conjunction of the initial conditions is taken to be a description of the cause, and the event described by the explanandum is taken to be its effect[22]. Insofar as the laws appearing in the explanans are explicitly considered to be empirical regularities, the assimilation of the cause-effect relation to the explanatory relation in the D-N model leads to a version of the regularity theory of causation. To that extent, the account still has to face Russell's second and third arguments sketched above. Interestingly, Hempel later acknowledged, to a certain extent, the inadequacy in principle of the D-N model to the cause-effect relation, and this for several reasons.

He acknowledges (in line with the conclusion of Russell's second argument) that, as long as the cause does not comprise the state of the entire universe at an instant, it is never strictly possible to deduce with logical necessity the description of the state of a limited system, from the description of a different state of the system and laws.

> «If the system is not isolated, i.e., if relevant outside influences act upon the system during the period of time from the initial state invoked to the state to be explained, then the particular circumstances that must be stated in the explanans include also those outside influences»[23].

Later, Hempel[24] even more explicitly acknowledged that the deduction of particular events (whose extent is narrower than the entire universe) is never strictly possible, for «if the lawlike premises of deductive-nomological explanations are not to be patently false, we must explicitly recognize the explanations' auxiliary reliance on qualifying clauses, commonly tacit, which govern such premises»[25], clauses that Hempel calls «*provisos*», and that others call «*ceteris paribus*» clauses[26].

---

[22] «The type of explanation which has been considered here so far is often referred to as causal explanation. If E describes a particular event, then the antecedent circumstances described in the sentences $C_1, C_2, ..., C_k$ may be said jointly to "cause" that event, in the sense that there are certain empirical regularities, expressed by the laws $L_1, L_2, ..., L_r$, which imply that whenever conditions of the kind indicated by $C_1, C_2, ... C_k$ occur, an event of the kind described in E will take place» (HEMPEL and OPPENHEIM, *Studies...*, 250). In short, «by virtue of thus presupposing general laws which connect "cause" and "effect", causal explanation conforms to the D-N model» (HEMPEL, *Aspects...*, 348).

[23] HEMPEL, *Aspects...*, 352.

[24] See C. G. HEMPEL, *Provisos: A Problem Concerning the Inferential Function of Scientific Theories*, in A. GRÜNBAUM and W. C. SALMON (ed. by), *The Limitations of Deductivism*, University of California Press, Berkeley 1988, 19-36.

[25] A. GRÜNBAUM, *Preface*, in GRÜNBAUM and SALMON (eds.), *The Limitations...*, x.

[26] On this problem, see KISTLER, *Causalité...*, chap. 3.

The second reason Hempel gives for the inadequacy of the D-N model as a means to analyze the cause-effect relation (and which is reminiscent of Russell's third argument) concerns «laws of coexistence»[27]. Hempel admits that the mere deducibility of the explanandum from initial conditions and general laws is no sufficient reason to consider those conditions as the cause of the event described by the explanandum. He gives the following counterexample to this thesis. Take the explanation of the period of an (ideal) pendulum by its length. This explanation uses the law relating the period T of an ideal pendulum to its length L: $T = 2\pi \sqrt{L/g}$ (where g is the acceleration of free fall at the surface of the Earth). It links two properties[28] the pendulum possesses at the same time; therefore the events or facts corresponding to the possession of these properties cannot stand in the relation of cause to effect, which requires temporal separation[29]. More generally, Hempel admits that there are a certain number of types of explanation that satisfy the criteria of the D-N model, but which cannot be interpreted as causal explanations: besides explanations using laws of coexistence, this is the case where the explanandum is not a particular event but a law. In the deduction of (the general form of) Kepler's laws from Newton's, no particular event is named that could be interpreted as cause or effect.

Finally, Hempel notes that there are explanations satisfying all the requirements of the D-N model, but where the "initial" conditions contain a reference to a state of the system that is posterior to the state described in the explanandum. One example is the explanation of the refraction of a light ray crossing the interface separating two media with different refractive index. The explanans mentions, besides Fermat's principle according to which the path of the ray is such as to minimize its overall duration, the points of departure and arrival of the ray, to deduce the point of its passage of the interface. Here, the arrival mentioned in the premises is posterior to the passage of the interface mentioned in the conclusion. Hempel notes that the «uneasiness»[30] produced by explanations of this kind stems from the apparent impossibility of interpreting them as causal explanations, for it seems a priori excluded that refractions and similar types of physical processes constitute cases of backwards causation where the effect pre-

---

[27] HEMPEL, *Aspects…*, 352.

[28] To be able to attribute its period to the pendulum at a given instant, one must consider it as «a dispositional characteristic» (HEMPEL, *Aspects…*, 352).

[29] R. CUMMINS, *The Nature of Psychological Explanation*, MIT Press, Cambridge, MA 1983, 3 uses this example to refute more generally the identification of scientific explanation with the discovery of a "covering law". Hempel (HEMPEL, *Aspects…*, 352 ff.) notes, concerning this example, that intuitively, the explanation of the pendulum's period by its length seems more acceptable than the explanation of its length by its period. However, none of the logical requirements of D-N explanation allows to justify this intuitive preference, the law linking these two variables being a symmetrical law of coexistence.

[30] HEMPEL, *Aspects…*, 353.

cedes the cause[31]. As a result, Hempel despairs of finding general conditions, which it would be sufficient to impose on D-N explanation, in order to yield a model of the causal relation.

Many expositions of the doctrine called the "deductive-nomological theory of causation", by the leading philosophers of the logical empiricist tradition take fewer precautions than Hempel, and identify the cause-effect relation directly with the relation between the initial conditions and the explanandum in a correct D-N explanation. According to Carnap,

> «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*»[32].

Popper, though a severe critic of logical empiricism in many regards, also identifies causal explanation with D-N explanation, without any further constraints:

> «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*»[33].

As we have seen, Hempel gives a number of reasons arguing against this D-N account of causation. The counterexamples can be divided in two categories[34]: the first contains D-N explanations in which the relation between explanans and explanandum cannot be taken to be causal because it is not a relation between

---

[31] It is controversial whether there is backwards causation in quantum mechanics. Cfr. H. REICHENBACH, *The Direction of Time* [1956], University of California Press, Berkeley 1991, and P. DOWE, *Backwards Causation and the Direction of Causal Processes*, «Mind», 105 (1996), 1-22.

[32] R. CARNAP, *Philosophical Foundations of Physics*, Basic Books, New York 1966.

[33] K.R. POPPER, *Logik der Forschung* [1934], J.C.B. Mohr, Tübingen 1994[10], 31; engl. transl. *The Logic of Scientific Discovery*, London, Routledge 1992[5], 59, italics Popper's. There are many variants of this doctrine. Stegmüller expresses the idea semantically, in terms of the truth conditions of causal statements. «Wo ein individueller Kausalzusammenhang festgestellt wird, da wird behauptet, dass eine spezielle Ereignisfolge unter ein (hypothetisch angenommenes) Naturgesetz subsumiert werden könne» (W. STEGMÜLLER, *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie*, Vol. I.: *Erklärung, Begründung, Kausalität* [1969], Springer, Berlin 1983, 512-513). Davidson defends the idea in terms of our justification for believing particular causal statements. «It does not follow that we must be able to dredge up a law if we know a singular causal statement to be true; all that follows is that we know there must be a covering law. And very often, I think, our justification for accepting a singular causal statement is that we have reason to believe an appropriate causal law exists, though we do not know what it is»: D. DAVIDSON, *Causal Relations* [1967], in D. DAVIDSON, *Essays on Actions and Events*, Clarendon Press, Oxford 1980, 160.

[34] Such counter-examples have been collected by P. W. HUMPHREYS, *The Causes, Some of the Causes, and Nothing But the Causes*, in *Minnesota Studies in the Philosophy of Science*, vol XII: *Scientific Explanation*, University of Minnesota Press, Minneapolis 1989, 283-306, 300-301, and W. SALMON, *Four Decades of Scientific Explanation*, University of Minnesota Press, Minneapolis 1990, 46-50.

spatio-temporally separated particulars. The second contains explanations concerning causal processes, but where neither can the initial conditions be identified with the cause nor the conclusion with the effect. The following belong to the first category.

a) The explanation of general laws, such as Kepler's laws from Newton's laws.

b) Logical and mathematical explanations, proceeding by deduction from axioms or theorems, which are not interpreted or which have abstract models as interpretations.

c) Explanations by laws of coexistence, such as the explanation of the period by the length of the pendulum.

d) Analytical or compositional[35] analysis, in particular functional analysis: an important type of scientific explanation does not focus on the occurrence of an event or on the acquisition of a property by an object, but rather on the nature of a property. For example, statistical physics explains what the property of an object O being at a certain temperature consists in, by indicating the underlying nature of that property: it identifies the mean kinetic energy of O's components as the microscopic reduction base of that property. There is no causal relation between the properties of the components and their interactions on the one hand, and the reduced property of the whole object on the other, if only because the instantiations of the reduced and reducing properties are not separated in space and time. An important type of such non-causal explanations concerns the nature of dispositional properties, as when the solubility of sugar in water is explained by properties of the sugar molecules and their interactions with water molecules.

e) Some explanations satisfy the purely formal requirements of the D-N model, but use a trivial generalization in the place of a genuine law of nature. Thus, one can "explain" the fact that John has not become pregnant, by deducing it from the fact that he has taken contraceptive pills for the whole year, together with the "law" that no man who regularly takes contraceptive pills will become pregnant. There is of course neither a link of relevance nor of causation between men's taking birth-control pills and their avoiding pregnancy.

The second category of counterexamples to the D-N theory of causation consists of cases where the explanation is about a causal process, but where nevertheless the initial conditions do not cause the explanandum.

a) Explanations using extremal principles, such as the explanation of the path of a refracted light ray mentioned above.

---

[35] Cummins distinguishes explanations of these types from causal explanation by subsumption under a causal law. Furthermore, Cummins distinguishes the compositional analysis of a *system* from the analysis of a *property*. He calls the latter "functional analysis" (CUMMINS, *The Nature of Psychological …*, 15) if the property is dispositional, otherwise "property analysis".

b) The D-N model allows the explanation of the cause by the effect. To take a famous example[36], a tower projects a shadow. The (non-causal) explanation of the height of the tower by the length of the shadow satisfies the requirements of the model just as well as the (causal) explanation of the length of the shadow by the height of the tower. However, only in the second case can the initial conditions be interpreted as a cause of the explanandum.

c) The D-N model allows the explanation of an event by another, where both are effects of a common cause, although neither is a cause of the other. The explanation of the storm by the fall of the barometer is legitimate insofar as the link between these events is nomological, rather than accidental. However, the initial conditions do not indicate a cause of the explanandum; rather, both are effects of a common cause, the fall of air pressure.

d) Henry Kyburg[37] has noted that the D-N model makes the following explanation legitimate: a pinch of salt dissolves in warm water. This can be explained by the fact that the salt has been hexed and that by necessity, all samples of hexed salt dissolve in warm water. In this case, there really is a causal relation, but the presence, among the initial conditions, of a piece of information without relevance for the effect, makes it impossible to identify the whole set of those initial conditions with the cause.

e) Michael Scriven[38] has discovered a counterexample similar to the preceding one in its structure, but which poses are more pressing problem. It raises the so-called 'preemption problem', which arises in cases of potential causal overdetermination. A bridge collapses as a result of being overloaded. A bomb that would have been sufficient to destroy the bridge explodes just after it has collapsed. The explanation of the collapse by the explosion meets the D-N model's requirements without indicating the cause of that event. This is expressed by saying that the efficacy of the bomb has been "preempted" by the overload. According to Hempel[39], it is possible to rescue the model from refutation by this case, by taking into account the *process* leading to the destruction of the bridge, as it lawfully evolves in time. On closer inspection, the law that allows deducing the destruction of the bridge from the explosion of the bomb governs the propagation of the shock wave, from the space-time location of the explosion and with finite speed. However, in the case described by Scriven, the D-N explanation

[36] Cfr. S. BROMBERGER, *Why-Questions* [1966], in S. BROMBERGER, *On What We Know We Don't Know*, University of Chicago Press, Chicago and CSLI, Stanford 1992, 75-100, 83.

[37] H.E. KYBURG JR., *Comment*, «Philosophy of Science», 32 (1965), 147-151.

[38] M. SCRIVEN, *Explanation, Prediction, and Laws*, in H. FEIGL and G. MAXWELL (ed. by), *Minnesota Studies in the Philosophy of Science*, III, University of Minnesota Press, Minneapolis 1962, 170-230.

[39] HEMPEL, *Aspects…*, 420.

will turn out not to be correct for at the moment of the shock wave's arrival at the location of the bridge, there already is no more bridge there to be destroyed.

These counterexamples do not only show that the deductive-nomological theory of causation, the contemporary version of Hume's regularity theory, is inadequate. They also set the agenda for the search of an alternative account[40]. The four major theories of causation, which are to be discussed here, will be evaluated first of all by their capacity to cope with those problem cases. We shall consider in turn the counterfactual, the interventionist, the probabilistic and the transference account.

2.   The counterfactual theory of causation

David Lewis has proposed returning to Hume's second definition of cause according to which «we may define a cause to be an object followed by another [...] where, if the first object had not been, the second had never existed»[41]. Lewis' proposal consists in reducing causation between two events to a finite chain of intermediate events, each link of which causally depends on the preceding[42]. The notion of *causal dependence* between events is then analysed with the help of the *counterfactual* dependence between the propositions saying that those events have taken place. According to the counterfactual theory of causation, event *e* depends causally on event *c* if the following two counterfactuals are true: if *c* had occurred, *e* would have occurred; if *c* had not occurred, *e* would not have occurred[43]. The distinction between causal dependence and the causal relation itself is introduced in order to guarantee the transitivity of the causal relation, for causal dependence is not transitive[44]. A situation is con-

---

[40] Lewis motivates his elaboration of a counterfactual theory of causation by the enumeration of problems encountered by the deductive-nomological account. «I have no proof that regularity theories 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». (D. LEWIS, *Philosophical Papers*, vol. II, Oxford University Press, New York 1986, 160).

[41] D. HUME, *An Enquiry Concerning Human Understanding*, in *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, L.A. Selby-Bigge and P.H. Nidditch (ed. by), Clarendon Press, Oxford 1975[3], 76.

[42] Precedence is here understood logically. It is determined by the order of the members of the chain, without prejudging the temporal precedence of the cause with respect to the effect. Indeed, Lewis gives an independent argument against the possibility of backwards causation, where the effect is earlier in time than the cause. I shall present and criticize that argument below.

[43] The first counterfactual carries the weight in cases where the events *c* and *e* do not occur, whereas the second carries the weight in cases in which they do occur. Cfr. LEWIS, *Philosophical Papers...*, 167.

[44] «Causation must always be transitive; causal dependence may not be so»: LEWIS, *Philosophical Papers...*, 167. Among the defenders of the counterfactual approach, some have criticized Lewis' move to

ceivable where the premises of the following argument are true but its conclusion false[45]:

If Otto had gone to the party, then Anna would have gone.

If Anna had gone, then Waldo would have gone.

Therefore, if Otto had gone, then Waldo would have gone.

The conclusion is false if Waldo is Otto's unlucky rival for Anna's affections, and does all in his power to avoid his rival. The non-transitivity of counterfactual dependence can be explained by the fact that their evaluation implicitly presupposes a concrete situation, and that the situation presupposed by the conclusion differs from that presupposed in the premises[46].

Lewis's proposal still faces some of the counterexamples to the deductive-nomological account.

a) There seem to be counterfactuals expressing dependence of the cause on the effect. Here is such an apparently true "backtracking" counterfactual, i.e. a counterfactual whose antecedent bears on an event that happens *later* than the event described by the consequent. If there had not been a solar eclipse on 11 August 1999, the sun, moon and earth would have had different positions or different speeds on 10 August 1999. Lewis calls this «the problem of effects»[47].

b) Similarly, two effects of a common cause seem to depend counterfactually on each other. Given that Smith and Jones watch all (and only) soccer matches broadcast by TV XY, it seems that if Smith had seen the match on day D, then Jones would have seen it (because XY would have shown it), and if Smith had not seen it, Jones would not have seen it either (because XY wouldn't have shown it). Lewis calls this «the problem of epiphenomena»[48].

c) Finally, the counterfactual analysis seems to provide an incorrect result in cases of preemption and causal overdetermination. If a prisoner is executed by two bullets *A* and *B* shot by two soldiers at the same time, each of which would alone have been sufficient for death, we have a case of causal overdetermina-

---

guarantee the transitivity of causation by stipulation. As Geert Keil points out (G. KEIL, *Handeln und Verursachen*, Vittorio Klostermann, Frankfurt am Main 2000, 297), within a theory aiming at reducing causation to counterfactual dependence, it seems plausible to accept the fact that counterfactual dependence is not transitive, and to conclude that causation is not transitive either. Ned Hall (N. HALL, *Causation and the Price of Transitivity*, «Journal of Philosophy», 97 (2000), 198-222) shows that if one accepts "negative events", such as absences and omissions among causes and effects, then the thesis that counterfactual dependence is sufficient for causation enters in conflict with the thesis of the transitivity of causation.

[45] The example is from David Lewis: D. LEWIS, *Counterfactuals*, Blackwell, Oxford 1973, 32.

[46] Cfr. KEIL, *Handeln…*, 295-296.

[47] LEWIS, *Philosophical Papers…*, 170.

[48] LEWIS, *Philosophical Papers…*, 170.

tion, or, in Lewis' words, of «redundant causation»[49]. Intuitively, each bullet seems to be a cause of death. However, it is not true that if, say, bullet *B* had not been shot, the prisoner would not have died. Or, to take the case of preemption mentioned earlier, if the bridge hadn't been overloaded, it would have collapsed nevertheless. Hence, the counterfactual analysis denies, wrongly, that the overload is, in the situation sketched above, the cause of the bridge's collapsing[50].

The counterfactual theory inherits, it seems to me, some of the difficulties of the nomological theory because it tacitly relies on the same foundations[51]. (This is a controversial thesis that I will try to justify below). Let us follow Lewis[52] (and Stalnaker[53]) in their conception of the truth-conditions of counterfactual conditionals, in terms of possible worlds. Let us assume that A is false in the actual world (or, equivalently, that event *a* does not occur in the actual world: A is the proposition that *a* occurs[54]). On which condition is the counterfactual "if it were that A, then it would be that C" (equivalent to: if *a* had occurred, then *c* would have occurred») true? According to Lewis, it is true in the actual world "if and

[49] LEWIS, *Philosophical Papers…*, 193.

[50] Lewis (LEWIS, *Philosophical Papers…*, 205-207) proposes to justify the intuition that the overload is the cause of the collapse although it does not counterfactually depend on it, by pointing out that the collapse is nevertheless *quasi-dependent* on the overload. The relation between the overload and the collapse is causal because it is intrinsically similar to other relations between overloads and collapses, in which no preempting cause is present. In those other situations, there is counterfactual dependence; hence (by definition of this concept) there is quasi-dependence in the situation in which the preempting cause is present. Causation being an intrinsic relation depending only on the process linking cause and effect, the relation is causal even in the situation in which the preempting cause is present. Hence, quasi-dependence is sufficient for causation, even in the absence of dependence. However, there are cases of preemption, in which the effect is also quasi-dependent on the preempted cause. Schaffer (J. SCHAFFER, *Trumping Preemption*, «Journal of Philosophy», 97 (2000), 165-181) shows that Lewis' theory of quasi-dependence systematically fails in cases of "trumping preemption", where the preempted process is never interrupted and goes to completion, just as the preempting process. In such cases, the theory of quasi-dependence yields the wrong result that the preempted cause is a cause because the effect is quasi-dependent on it. Schaffer (SCHAFFER, *Trumping Preemption*, 175) reports the following example, attributing it to Bas van Fraassen and Ned Hall: The major and the sergeant, both before the corporal, shout «charge!» at the same time and the corporal decides to charge. The higher officer's order "trumps" the lower officer's; hence the major's order, not the sergeant's, is the cause of the corporal's decision. However, it is at least conceivable that the competing process from the sergeant's order to the corporal's decision not be interrupted. In that case, the decision is quasi-dependent on the sergeant's order just as it is on the major's. Lewis (D. LEWIS, *Causation as Influence*, «Journal of Philosophy», 97 (2000), 182-197, 185) himself acknowledges that situations of this kind refute his theory of quasi-dependence.

[51] However, these problems do not arise in exactly the same way. In a case of preemption, e.g., the nomological theory yields the false result that the preempted event is a cause, whereas the counterfactual theory yields the false result that the real (preempting) cause is not a cause, because the effect would have occurred even if it had been absent. (See LEWIS, *Philosophical Papers…*, 200).

[52] D. LEWIS, *Counterfactuals*, Blackwell, Oxford 1973.

[53] R. STALNAKER, *A Theory of Conditionals*, in N. RESCHER (ed. by), *Studies in Logical Theory*, Blackwell, Oxford 1968.

[54] See LEWIS, *Philosophical Papers…*, 166.

only if some (accessible[55]) world where both A and C are true is more similar to our actual world, overall, than is any world where A is true but C is false"[56]. There is no universal algorithm allowing us to evaluate the comparative similarity of any pair of possible worlds. However, one can establish a hierarchy of criteria, in decreasing order of importance. What is, according to Lewis, responsible for the largest dissimilarity between a world w_0 and a world w_1, is that w_1 contains «big widespread, diverse violations of law»[57] with respect to the laws of w_0. In other words, in order to find the nearest world where the antecedent A of the counterfactual is true, one must eliminate all worlds containing major violations of the laws of w_0. According to Lewis' second most important criterion, the most similar worlds are those with a «perfect match» of all facts existing in the largest spatio-temporal region of w_0. The second least importance is given to the criterion according to which worlds are closer if they contain fewer «small localized simple violations of law»[58]. In other words, w_1 is closer to w_0 than w_2, if w_1 contains fewer events which appear as "miracles" from the point of view of the laws of w_0 than w_2. The least weight is given to similarity with respect to particular matters of fact.

Lewis tries to justify his thesis that backtracking counterfactuals are never true on the basis of the application of these criteria of similarity between possible worlds[59]. This thesis turns out to be crucial for the defence of the counterfactual theory of causation, because many (perhaps all) counterexamples it faces get their plausibility (which Lewis takes to be only apparent) from a (sometimes implicit) reliance on backtracking counterfactual dependence. The first counterexample that raises the "problem of effects" (*a* above) is explicitly based on

---

[55] The concept of accessibility of one possible world from another allows us to construct different concepts of possibility and necessity. In particular, physical necessity can be distinguished from logical necessity, on the basis of the distinction between possible worlds in which our actual laws of nature hold and worlds in which they do not. The former are accessible from the actual world with respect to physical possibility: a proposition is physically necessary if and only if it is true in all physically accessible possible worlds; a proposition is physically possible if and only if there exists a physically accessible world in which it is true. Cfr. LEWIS, *Counterfactuals*, 4-5.

[56] LEWIS, *Philosophical Papers…*, 41.

[57] LEWIS, *Philosophical Papers…*, 47.

[58] LEWIS, *Philosophical Papers…*, 48.

[59] The counterfactual theory might be tempted, like the nomological theory, to avoid the problem of having to judge falsely that effects are also causes of their own causes (the "problem of effects"), simply by stipulating that the cause always precedes its effects in time. It seems, however, preferable to avoid this easy solution, for three reasons: cfr. LEWIS, *Philosophical Papers…*, 170. 1) Such a stipulation solves neither the problem of epiphenomena nor the problem of preemption. 2) It prejudges a priori that backwards causation is impossible, whereas the controversial issue of the existence of backwards causation seems to be empirical and should therefore be decided on scientific grounds. 3) It makes a causal theory of time a priori inconceivable.

a backtracking counterfactual. The counterexample of the common cause (*b* above), the proposition that if Smith hadn't seen the match, then Jones hadn't seen it either, is grounded on the backtracking counterfactual that if Smith hadn't seen it, then TV station XY wouldn't have broadcast it[60].

Let us consider the following example[61]. In world w_0, during the cold war and Nixon's presidency of the USA, there is a red button that can trigger, with perfect reliability, a nuclear holocaust. We can imagine that w_0 is our actual world. In w_0, Nixon does not push the button. According to Lewis, the counterfactual "if Nixon had pushed the button, there would have been a nuclear holocaust" is true, whereas the backtracking counterfactual "if there had been a nuclear holocaust, then N. would have had pushed the button" is false. The reason why the former is true is that the closest world where N. pushes the button is, according to Lewis, a world that exactly resembles our own (w_0), until the moment immediately preceding N.'s fatal act, but in which «a tiny miracle takes place» at precisely that instant, such that «the deterministic laws of w_0 are violated at w_1 in some simple, localized, inconspicuous way»[62]. W_1 differs from w_0 only by «a few extra neurons [that] fire in some corner of Nixon's brain»[63]. Among the worlds meeting that requirement, the nearest to w_0 is that which contains no other miracles or any other violations of the laws of nature. The consequent of the "forward" counterfactual is therefore true in w_1, because our laws inevitably determine the occurrence of the nuclear holocaust, once the button is pushed.

What about the truth value of the backtracking counterfactual that "if there had been a nuclear holocaust, then N. would have had pushed the button"? According to Lewis, it is false, because the closest world to w_0 in which its antecedent is true, i.e. in which there is a nuclear holocaust, is a world in which that holocaust is caused by a tiny miracle immediately preceding it. In effect, Lewis shows that worlds in which the miracle is located *after* the holocaust, are necessarily more distant from w_0 than those where the miracle is located *before* it. A miracle in w_1 occurring immediately after the holocaust maximizes the similarity with w_0, as far as its final phase starting with the holocaust is concerned. However, such a miracle would not at all be «small, localized, inconspicuous»[64], because a very large miracle is needed to "wipe out" the large number of effects

---

[60] Lewis shows (Lewis, *Philosophical Papers…*, 200-201) that at least certain cases of preemption are only counterexamples to the counterfactual theory if there are true backtracking counterfactuals.

[61] Lewis, *Philosophical Papers…*, 43ff.

[62] Lewis, *Philosophical Papers…*, 44. On miracles, see also Lewis, *Counterfactuals*, 75.

[63] Lewis, *Philosophical Papers…*, 44.

[64] Lewis, *Philosophical Papers…*, 44.

spreading out from a single cause[65]. This asymmetry stems from the existence of irreversible processes, which is, according to a plausible hypothesis, also at the origin of the asymmetries of causation and time[66].

However, contrary to what Lewis asserts, the irreversibility of physical processes is no obstacle to the possibility of reasoning backwards in time. Lewis' observation that the miracles that must be postulated to allow a perfect match between w_0 and a time span of a possible world in which the antecedent of a counterfactual is true, are smaller if they are located earlier than the occurrence of the antecedent, rather than later, obliges us only to look for the nearest possible world, among the worlds where the miracle takes place *before* the holocaust described in the antecedent. However, this does not yet completely determine the temporal location of that miracle: Lewis thinks that it must be located *immediately before* the event described in the antecedent, i.e. the nuclear holocaust; this would indeed let the counterfactual come out false.

> «The proper solution to both problems [the problems of effects and of epiphenomena; M.K.], I think, is flatly to deny the counterfactuals that cause the trouble. If *e* had been absent, it is not that *c* would have been absent. [...] Rather, *c* would have occurred just as it did but would have failed to cause *e*»[67].

In our case, Lewis' verdict would be: if *e* (the holocaust) had been present, it is not that *c* (the event where N. pushed the button) would have been present; rather, *c* would still have been absent but *e* would have been produced by a miracle immediately before *e*.

Lewis' analysis of such cases presupposes that the counterfactual can be evaluated on the basis of a possible world in which a miracle occurs *within the stretch of time between the events described by the antecedent and the consequent* of the counterfactual. However, this destroys the capacity of that world to ground our counterfactual judgment. The possibility of reasoning in a counterfactual way presupposes that we consider that the laws of the actual world are valid without exception in the world on the basis of which we evaluate the counterfactual and in which its antecedent is true, *within the stretch of time between the events described by the antecedent and the consequent.*

Indeed, on what can we ground our intuitive judgments of the truth values of counterfactuals, if not on the assumption that the counterfactual world in which

---

[65] «The big miracle required for perfect reconvergence consists of a multitude of little miracles, spread out and diverse»: LEWIS, *Philosophical Papers...*, 47.

[66] See REICHENBACH, *The Direction of Time...*

[67] LEWIS, *Philosophical Papers...*, 170.

the antecedent is true, is governed by the same laws as the actual world? Traditionally[68], one of the main functions a knowledge of laws is thought to have is to allow us to judge which course events would have taken in counterfactual circumstances. Our capacity to make such judgments is necessary in order to make decisions: we evaluate the consequences of each possible action we might choose. We can only hypothesize these consequences on the condition of assuming that the laws of nature do not change after such a possible act. It seems therefore necessary to ground the evaluation of counterfactuals on the requirement that the possible world to be considered for that evaluation must be closest to ours with respect to the facts existing at the time of the antecedent, and must obey the same laws as the actual world during the temporal interval between the antecedent and the consequent. The requirement of holding the actual laws constant at least for that time span, forces us to situate the miracle *before* the consequent (the event where N. pushes the button). In this case, there are no principled obstacles any more to backwards counterfactual reasoning. The counterfactual is true if the consequent is true in the world that is closest to the actual world with respect to the facts at the time of the antecedent (posterior in time) and in which, during the interval between antecedent and consequent, the laws are the same as the actual laws[69]. Even if not all laws applying to a given situation are reversible, it is sufficient for the truth of a given backwards counterfactual that the consequent event *c* is, in the circumstances, necessary for the antecedent event *e*. The backwards counterfactual about Nixon is true if, in the circumstances, other possible causes that could have triggered the holocaust without the intervention of a miracle are situated in possible worlds more distant from the actual world than the world where N. triggers the holocaust by pushing the button.

To the extent that this argument refutes Lewis' thesis that no backwards counterfactual is true, his defence of the counterfactual theory of causation, threatened by the counterexamples mentioned above, is deprived of its main conceptual resource. If backwards counterfactuals can be true, the counterfactual theory is defeated by the problems of effects and of epiphenomena.

So far, I have agreed to reason in terms of "miraculous" worlds. As Lewis is

[68] Cfr. W. KNEALE, *Probability and Induction*, Oxford University Press, Oxford 1949; N. GOODMAN, *Fact, Fiction and Forecast* [1955], Bobbs-Merrill, Indianapolis 1973³; E. NAGEL, *The Structure of Science*, Routledge and Kegan Paul, London 1961; F. JACKSON, *A Causal Theory of Conterfactuals*, «Australasian Journal of Philosophy», 55 (1977), 3-21; D.M. ARMSTRONG, *What is a Law of Nature?*, Cambridge University Press, Cambridge 1983; D.W. ZIMMERMANN, *Shoemaker's Argument for his Theory of Properties*, «Facta philosophica», 2 (2000), 271-290.

[69] See JACKSON, *A Causal Theory…*, 9ff.

careful to point out[70], miraculous worlds are not inconsistent: they do not contain any violations of their own laws. Rather, they appear miraculous from the point of view of the actual world. Their laws (those of w_1, which are strictly followed in w_1) differ from ours in such a way as to permit that the course of events differs from the actual course only in a tiny portion of space-time but nowhere else. However, the conception of such a world seems to contain a contradiction. By definition, a law is a constraint determining a *global* regularity[71]; hence, even a tiny difference between two possible worlds as to the laws valid in them, must have as a consequence a vast divergence in facts and particular events throughout all past and future. Furthermore, any change with respect to a particular fact, even as tiny as a few extra neurons firing in N.'s brain, requires the violation of several laws, beginning with the laws of conservation. The counterfactual analysis faces the following dilemma.

Either the worlds that are candidates for being the basis of evaluation of a counterfactual because they are very similar to ours are coherent worlds in the sense that their laws are strictly valid. But then these laws must be the same as our actual laws, otherwise their difference from actuality would necessarily be vast with respect to particular facts, and we would have no means to make judgments about what happens in them because we cannot reason on the basis of laws that differ from ours in an unknown way.

Or else these worlds are truly miraculous. However, a contradictory world whose laws are no laws (because they are violated) cannot serve as a basis for the evaluation of a counterfactual.

It is possible to avoid the consideration of miraculous possible worlds, by grounding the evaluation of counterfactuals in possible worlds that strictly share the actual laws but that lack the counterpart of a part of the history of the actual world[72]. To evaluate a "forward" counterfactual where the antecedent concerns an event at t, one considers a possible world containing the counterpart of a temporal part of the actual world, which extends from t to infinity in future, but which does not contain any counterpart of the actual past, before t. This solution requires rejecting Lewis' realist conception which has it that all possible worlds are as real as the actual world. It presupposes conceiving them as devices for reasoning. As such they must be internally consistent but need not be possible *complete* worlds. A complex world beginning at t without having a past is a possible world in this sense, although it could not possibly be a complete world.

---

[70] Lewis, *Philosophical Papers…*, 44-5.

[71] According to regularity theories of laws, a law simply *is* a global regularity.

[72] See M. Kistler, *L'identité des propriétés et la nécessité des lois de la nature*, «Cahiers de philosophie de l'Université de Caen, Numéro spécial: Le réalisme des universaux», 38-39 (2002).

3. The interventionist or manipulability analysis of causation

As we have seen, the asymmetry of causation raises important difficulties for both the nomological and the counterfactual approach to causation. The interventionist analysis[73] tries to overcome them by grounding that asymmetry in the asymmetry between the past, which we cannot influence, and the future, which presents itself to us agents as an «open 'space' of possibilities»[74]. If A is the occurrence of an avalanche, and B and C are the collapses of two houses caused by A, on what ground can we reject the following problematic counterfactuals: the first, which raises the "problem of effects", explicitly expresses backwards reasoning: "if B had not occurred, then A would not have occurred". The second, which raises the "problem of epiphenomena", also presupposes, although only implicitly, backwards reasoning: "if B had not occurred, then C would not have occurred". The interventionist analysis proposes judging their truth value on the basis of a possible world differing from ours simply by a human intervention, which plays the role of Lewis' miracle. The worlds we compare in order to evaluate counterfactuals share our actual laws, but differ with respect to a free action. In the case of the example, the minimal intervention that prevents B consists in directly protecting the house in B. But if this had been done, A as well as C would still have occurred, which is what grounds our judgment that the two controversial counterfactuals are false[75].

More generally, the interventionist or manipulability account of causation[76] reduces the existence of a causal relation between two distinct events A and B to the fact that «bringing about the occurrence of A would be an effective means by which a free agent could bring about the occurrence of B»[77]. However, this proposal encounters two major problems[78]. First, it is anthropocentric, in that it postulates a conceptual dependency between causation and our human capacity to act. Not content to note that our actions are causes, it holds that generally,

---

[73] D. GASKING, *Causation and Recipes*, «Mind», 64 (1955), 479-487; G.H. VON WRIGHT, *Explanation and Understanding*, Cornell University Press, Ithaca, N.Y. 1971; CARTWRIGHT, *How the Laws…*; H. PRICE, *Agency and Causal Asymmetry*, «Mind», 101 (1992), 501-520; P. MENZIES AND H. PRICE, *Causation as a Secondary Quality*, «British J. Phil. of Science», 44 (1993), 187-203; KEIL, *Handeln…*

[74] KEIL, *Handeln…*, 440.

[75] This is Keil's way of reasoning. Cfr. KEIL, *Handeln…*, 442.

[76] Menzies and Price call it the «agency theory» (MENZIES AND PRICE, *Causation as a Secondary…*, 187).

[77] MENZIES AND PRICE, *Causation as a Secondary…*, 187.

[78] Hausman mentions three difficulties: «The manipulability theory is narrowly circular, unjustifiably anthropomorphic, and it makes causal inferences on the basis of passive observation needlessly problematic»: D. HAUSMAN, *Causation and Experimentation*, «American Philosophical Quarterly», 23 (1986), 143-154, 145.

any event *c* is a cause of another event *e* if and only if a human agent could obtain *e* by bringing about *c*. This approach is condemned to consider causal relations taking place without the reach of possible human actions to be causal only by analogy. It is in principle impossible for a human agent to bring about the explosion of a supernova. Hence, the interventionist analysis doesn't apply to that process and cannot explain, at least not directly, why it is causal. Furthermore, decisions to consider one relation or process to be causal because analogous, or similar enough, to an action, and another relation or process not to be causal because not similar enough, seem arbitrary. Hence, the theory seems to have no clear application to causal relations taking place without the reach of possible human actions.

The second problem stems from the circularity of the account. The interventionist account uses our capacity to act as a conceptual tool to analyse causation in general. However, when we act, we cause events. Hence, the concept of action presupposes the concept of causation. The circularity can be avoided if one explicitly takes the concept of action to be primitive and unanalysable, in particular to be incapable of an analysis in causal terms[79]. However, in that case we don't seem to have any analysis of the concept of causation at all, but only the thesis that a certain species of causal interactions, namely human actions, are paradigmatic and can serve as a model for the whole genus. In the absence of a direct analysis of what characterizes that genus, i.e. causation itself, the theory leaves us in the dark as to what, among the properties of human actions, can be generalized to all causal relations.

This critique is not meant to deny the fact, well established by von Wright[80], that all *knowledge* of objective causal relations must rely on experimental interventions[81]. Our capacity to intervene in the natural course of events is our only epistemic handle on causation. In von Wright's own words,

> «it is *established* that there is a causal connection between p and q when we have satisfied ourselves that, by manipulating the one factor, we can achieve or bring it about that the other is, or is not, there. We usually satisfy ourselves as to this by making experiments»[82].

---

[79] The dependence of the concept of intervention on the concept of causation makes it impossible to use the interventionist theory in the causal theory of action. According to the latter, actions differ from other bodily movements by their causal origin in an appropriate mental event. However, it is obviously circular to analyse both action in terms of causation, and causation in terms of action. Cfr. KEIL, *Handeln…*, chap. III, 3.

[80] VON WRIGHT, *Explanation…*

[81] These methods of investigation of causal relations are the object of John Stuart Mill's (J.S. MILL, *A System of Logic, Ratiocinative and Inductive* [1843], Parker, London 1848) famous classification.

[82] VON WRIGHT, *Explanation…*, 72; italics von Wright's.

However, we can acknowledge this important epistemic fact about our knowledge of causation, without obscuring the fact that natural causal relations (which are not actions) do not, for their objective existence, depend on human actions.

4. The probabilistic theory of causation

The fundamental thesis of the probabilistic approach to causation is that A is cause of B if and only if the occurrence of A raises the chance of the occurrence of B. The theory has a twofold origin. It can be considered as a weakened version of the deductive-nomological theory of causation, which is obtained by substituting probabilistic or statistical laws for the deterministic laws in the explanans[83]. However, the theory can also be seen to be motivated by the theory of rational choice[84]. Here is the link. It is rational for an agent to do A in order to obtain B if and only if the agent believes that A causes B. According to the normative theory of choice, it is rational to do A in order to obtain B if and only if the agent believes that $P(B|A) > P(B|\neg A)$. If both equivalences are correct, one can conclude that A causes B if and only if $P(B|A) > P(B|\neg A)$[85].

The major problem with this approach concerns its application to the analysis of singular causation, i.e. to causal relations between particular pairs of events. To take an example well known in the debate on probabilistic causation, let us assume that the chance of developing lung cancer is higher among smokers than non-smokers, in the sense that the frequency of cancer is higher in the population of smokers than in that of non-smokers. The probabilistic theory of causation, in its simplest version, concludes that smoking is a cause of cancer. However, the statistical correlation between two populations is not sufficient to justify a causal judgment in any individual case[86]. It seems perfectly compati-

---

[83] Frederick Suppes introduces the theory in this way. Cfr. F. SUPPES, *A Probabilistic Theory of Causality*, North Holland, Amsterdam 1970.

[84] In this respect, it can be considered as a variant of the manipulability theory, discussed in the preceding section.

[85] Papineau (D. PAPINEAU, *Pure, Mixed, and Spurious Probabilites and Their Significance for a Reductionist Theory of Caustion*, in *Minnesota Studies in the Philosophy of Science*, vol XII: *Scientific Explanation*, ed. by P. Kitcher, University of Minnesota Press, Minneapolis 1989, 307-356, 321 and 328) and Mellor (D.H. MELLOR, *The Facts of Causation*, Routledge, London and New York 1995, chap. 7) present this argument as a major motivation for the probabilistic theory. It is certainly not conclusive, in that it only allows one to draw the conclusion that the *beliefs* are equivalent: it shows that it is equivalent to *believe* that A causes B and to *believe* that $P(B|A) > P(B|-A)$. It is of course invalid to conclude from this, without further argument, that the objects of these beliefs are equivalent too.

[86] «Given the conceptual independence of token-level causal facts from type-level causal facts, it should not be surprising that what is true at the type level is […] conceptually independent of what is true of the token level, and that token-level causation cannot be straightforwardly understood in terms of type-

ble with our assumption that John's smoking does not cause cancer in him, either because he doesn't get cancer at all or because his cancer is causally due to some other factor, e.g. to the inhalation of asbestos dust. Hence, higher chance of occurrence of an event of type B, within a population that also satisfies A, in comparison with the chance of B in a population that does not satisfy A, is not *sufficient* for concluding that the occurrence of event a (of type A) causes event b (of type B), even if both have occurred. The fact that A raises the odds of B is not a *necessary* condition for a's causing b either, for an individual event a (of type A) can well cause the individual event b (of type B) while in the general case, the occurrence of events of type A *lowers* the probability of occurrence of an event of type B. This is illustrated in the following case[87]: Moriarty, Watson, and an unstable rock are at the edge of a cliff. Holmes stands at the foot of the cliff. Watson must leave, but he knows that if he doesn't act before leaving, Moriarty will push the rock over the edge of the cliff, with the intention of killing Holmes. The only thing Watson can do in order to try to save Holmes, is to push the rock himself hard enough for it to fall down beyond Holmes. Watson manages to get hold of the rock and pushes it as hard as he can. Unfortunately, and improbably, he fails and Holmes is killed. The cause of Holmes' death is an event of a type which diminishes the chance of his death. The event of pushing the rock «had a tendency to prevent»[88] Holmes' death, but has nevertheless caused it.

The probabilistic theory seems ill suited for the analysis of causal relations between particular events for two other reasons. The terms of probabilistic correlations are abstract entities, such as properties, factors, types of events or properties of events, e.g. smoking or developing cancer. These entities are situated neither in space nor in time. They cannot therefore be causally related strictly speaking, to the extent that causality links entities situated in space and time, generally conceived of as events. This suggests that the probabilistic theory might really be a theory not so much of causation, but of explanation or of the nomological link between properties, in other words, of laws of nature. In the D-N model of explanation, it could play both of these roles.

The second reason for the inadequacy of the probabilistic theory for the

level causal relations»: E. EELLS, *Probabilistic Causality*, Cambridge University Press, Cambridge 1991, 16.

[87] The example is due to I.J. GOOD, *A Causal Calculus I-II* [1961-62], in I.J. GOOD, *Good Thinking*, University of Minnesota Press, Minneapolis 1983. Other examples can be found in E. EELLS and E. SOBER, *Probabilistic Causality and the Question of Transitivity*, «Phil. of Science», 50 (1983), 35-57; EELLS, *Probabilistic Causality*, 281ff.; M. HEIDELBERGER, *Kausalität: Eine Problemübersicht*, «Neue Hefte für Philosophie», 32-33 (1992), 130-153, 151; D. EDGINGTON, *Mellor on Chance and Causation*, «British Journal for the Philisophy of Science», 48 (1997), 411-433, 420.

[88] GOOD, *A Causal Calculus...*, 216-217.

analysis of causation between spatio-temporal events is this. The statistical correlations that constitute the point of departure of the probabilistic analysis are often spurious, in the sense that the correlation exists between two (types of) effects of a common (type of) cause, but where these effects are not linked as cause and effect. Hence, the theory needs to be completed by the following requirement: the fact that A raises the probability of B justifies belief in the existence of a causal relation between (events of type) A and B, only if there is no third factor that is the real cause of both. Such a factor is called a 'screening factor'. The correlation between barometer fallings and storms is spurious; the theory predicts that it is not causal because the fall in air pressure constitutes a screening factor: Within a set of situations in which air pressure is constant, the correlation between barometer fallings and storms disappears. However, this strategy cannot be generalised. Insofar as the sets, within which the conditional probabilities are evaluated, contain more than one element, it is always possible to partition the set in such a way that the conditional probabilities give a different result, while of course such a change in the calculation of the probability doesn't affect the reality of the causal relations. This fact is known as 'Simpson's paradox'[89]. As an illustration, take the famous case of the alleged sexist discrimination in the admissions policy of the University of Berkeley. Cartwright[90] cites a study showing that the frequency of admission at Berkeley is higher among male than among female applicants. However, the existence of this statistical correlation does not allow inferring the existence of a "causal"[91] relation between being a male applicant and the fact of being favoured in the selection. For it can be shown that there is another partition of the set of candidates that inverts the result, if the relevant conditional probabilities are evaluated separately within the sets of this partition: The partition with respect to the different departments selecting the candidates. It has turned out that the correlation between the gender of the candidate and her admission is not grounded in the existence of a "causal" relation between these two factors, but on a common "causal" factor: men have the tendency to candidate in departments in which it is easier to be accepted, while women candidate more often in the more selective departments. To simplify, let us consider a case in which there are only two departments and where the relevant frequencies are these[92]:

---

[89] After E.H. SIMPSON, *The Interpretation of Interaction in Contingency Tables*, «J. of the Royal Statistical Society», Ser. B, 13 (1951), 238-241. Cfr. EELLS, *Probabilistic Causality*, chap. 2.

[90] CARTWRIGHT, *How the Laws...*, 37.

[91] As I have already noted, calling abstract entities, such as properties or factors, 'causes', seems to be a category mistake. With respect to these entities, one should rather speak of nomological dependency, which is indeed a domain where the probabilistic theory can fruitfully be deployed.

[92] The numbers are taken from EELLS, *Probabilistic Causality*, 63.

	Department 1	Department 2	Total
Male	81/90	2/10	83/100
Female	9/10	18/90	27/100

In this particular situation, global evaluation of the probabilities – without distinguishing between the two departments – yields the result that the chance to be accepted is more than three times higher for male than for female candidates. However, the correlation between the gender of the candidates and their admission disappears if the probabilities are evaluated within the subsets obtained by partitioning the set of candidates according to departments. The first department accepts 90% of the male but also 90% of the female candidates, whereas the second department accepts only 20% of men but also only 20% of female candidates. Eells[93] shows that the existence of a global positive correlation between being male and being accepted is not only compatible with the absence of such a positive correlation (under a different partition), but also with the existence of a negative correlation (i.e. the fact that the departments in fact admit women more often than men) in the case of an evaluation by departments. It can also happen[94] that the evaluation on the basis of a partition in departments shows that there really is discrimination against women, but that it is smaller than the global evaluation would have made it appear. In short, it is impossible to draw any conclusive causal information whatsoever from an analysis of conditional probabilities calculated in an inhomogeneous set. To the extent that it is possible to introduce new partitions by paying attention to new factors, it can always happen that the conditional probabilities change radically. The fact that judgments of statistical correlation always remain relative to the choice of partition, and that the direction of correlation may depend on that choice, makes them inappropriate to ground absolute causal judgments.

The only possible way out seems to be to require, as Cartwright does, that the probabilities be evaluated within absolutely homogeneous sets, so that no causally relevant factor is left hidden from the analysis[95]. However, this solution tends to deprive the theory of its probabilistic elements, and to make it equiva-

---

[93] EELLS, *Probabilistic Causality*, 66f.

[94] See EELLS, *Probabilistic Causality*, 68.

[95] According to Cartwright, conditional probabilities must be evaluated «holding fixed other causal factors» (CARTWRIGHT, *How the Laws…*, 28), and one is only justified in inferring the existence of a causal relation on the basis of a rise in probability, if the set of situations over which the probabilities are taken are «homogeneous with respect to all other causal factors» (CARTWRIGHT, *How the Laws…*, 25). However, when it comes to considering real events, there are always other causal factors; no pair of concrete situations is homogeneous with respect to all causal factors.

lent to the nomological theory. For the case of deterministic causal relations, the relation of the conditional probability that B, given that A, reduces to the nomic relation according to which B nomically depends on A. As Eells shows, in order to evaluate correctly the dependence of B on A, all situations must be excluded in which there exists a correlation between A and B but where A is not causally related to B, and all situations in which the correlation between A and B is enhanced or diminished by factors influencing B, but which are not caused by A. Hence, to judge whether A causes B, one evaluates the probability of B, given that A, in all populations that are of types A and B but are perfectly homogeneous with respect to all potentially interfering factors. This undermines the ambition of the probabilistic theory to deliver an analysis of the causal relation, for the following reason.

In a deterministic world, if all potentially interfering external factors are held constant (let us assume that they are systematically absent), the probability that Y, given that X, is 1 if X causes Y, and 0 if X does not cause Y. The intermediate values are the result of the variability of the external interfering factors, which are here excluded. This undermines the fundamental idea of the probabilistic theory of causation. The intermediate probabilities do not measure X's impact on Y (which can only exist or not exist) but the probability of interferences by external factors. However, that probability is independent of the relation between X and Y itself. Apart from the case of irreducibly indeterministic processes, the content of the attribution of intermediary probabilities does therefore not bear on the causal relation itself, but on our ignorance of the interfering factors. Therefore, the probabilistic theory does not give a new analysis of the causal relation itself: with respect to objective causation, the probabilistic theory reduces to the nomological theory.

The probabilistic theory draws its plausibility from the analogy between the evaluation of conditional probabilities and the evaluation of a series of experiences, following Mill's methods. It is practically impossible to exclude all interfering factors, and the probability that B, given an experience where A has been observed and where the interfering factors have been eliminated to the extent that this is practically possible, will be intermediary. But the reason does not lie with the relation between A and B itself.

Eells[96] proposes a probabilistic theory of particular causal relations, which tries to reduce what I shall call 'causal responsibility' of a fact for another fact, to a certain evolution of the probability of the effect in the course of time, beginning with the time of occurrence of the cause, depending on this cause and

---

[96] EELLS, *Probabilistic Causality*, chap. 6.

on the external interfering factors. However, the probabilistic theory of particular causal relations encounters the same difficulties as the probabilistic theory of the causal dependence between factors.

In the case of particular causal relations, the probability of the effect given the cause is conceived as the limit of the frequencies of the effect in an infinite series of populations in which the cause is present. This series contains an infinite number of copies of the situation in its totality, with all interfering factors included. However, in a deterministic universe, to the extent that the situations in this series do not differ at all with respect to all factors that affect directly or indirectly the effect, if fact X is causally responsible for fact Y, then the probability that Y, given X, is 1. In this case, the resulting analysis of the relation of causal responsibility is not probabilistic any more. The probabilistic theory of causation, at least in its most elaborate form given by Eells, fails to solve the problem Russell's second argument raised against the regularity theory of causation. The attribution of probabilities below 1 is only justified to the extent that part of the context is ignored.

5. Transference theories of causation

The general conclusion of the analysis of the alternatives to the deductive-nomological theory of causation examined so far is that two of these, the counterfactual and the probabilistic theory, are better suited as theories of nomic dependence than as theories of causation. This seems particularly clear in the case of the probabilistic theory: its object is the analysis of systematic correlations between different properties (or factors, or between the sets that are their extensions). These correlations can reveal the existence of nomic dependencies existing between these properties[97]. However, they do not give any conclusive evidence about the existence of a causal relation between the events exemplifying these properties. As we have seen in our critical analysis of the deductive-nomological conception, nomic dependence is not equivalent to causation. Two properties can be nomically related even where two events exemplifying these properties are not related as cause and effect. However, I have also tried to show (what is more controversial) that the theory of causation in terms of counterfactual dependence can also more easily be interpreted as a theory of nomic dependence. The antecedent of a counterfactual is a proposition asserting the exemplification of a property (expressed by the predicate of the proposition). In order to deter-

---

[97] On the major contemporary conceptions of the laws of nature, see KISTLER, *Causalité...*, chap. 2, and BARBEROUSSE, KISTLER, LUDWIG, *La philosophie...*, chap. 4.

mine whether the proposition expressed by the consequent is true in the nearest possible world in which the antecedent is true, what one has to do is to examine whether there exists a nomic dependence between the property expressed by the predicate of the antecedent and the property expressed by the predicate of the consequent. To answer this question, it is necessary to contemplate possible worlds sharing our laws of nature.

The fact that the proper object of these theories is nomic dependency between properties, rather than the causal relations between particular events, is the deep reason for their failure to avoid counterexamples presented by situations in which A is nomically linked to B, but where an event *a* exemplifying A is nevertheless no cause of an event *b* exemplifying B: either because *b* is cause of *a*, or because *a* and *b* are effects of a common cause, or because *b* has been caused by a preempting cause *c*. Hence, the key to the solution to these problems lies in the search for an account that is no longer at the level of properties but at the level of the spatio-temporal events that are the relata of particular causal relations[98].

The most promising approach in this respect begins with the following hypothesis. What makes it the case that event *a* is a cause of event *b*, is the fact that something is transferred between these events. It has been proposed[99] that the fundamental characteristic of causal processes is their capacity to transmit a mark, where a mark is meant to be «the result of an intervention by means of an irreversible process»[100]. A paradigmatic case of the introduction of a mark, which then propagates all along a causal process is the passage of a light ray through a red filter. The colour is transferred by the ray as long as that process does not undergo any new interaction. In this respect, causal processes contrast with pseudo-processes, a clear example of which has been given by Salmon[101]. A beacon rotates at the centre of a hollow cylinder and projects a spot of white light on the internal wall. The world line that consists of the series of illuminated portions of the wall is a pseudo-process, rather than a causal process, because, if a red filter is posed on one of the portions of the wall traversed by the light spot, the spot will be marked red but this mark will not be transmitted along the pseudo-process which will consist of white light just as it did before that moment.

[98] This is also Ehring's conclusion. Cfr. D. Ehring, *Causation and persistence: A Theory of Causation*. Oxford University Press, New York 1997.

[99] See Reichenbach, *The Direction of Time*, 198f.; W. Salmon, *Scientific Explanation and the Causal Structure of the World*, Princeton University Press, Princeton 1984.

[100] Reichenbach, *The Direction of Time*, 198. The similarity with the manipulability account is only apparent. The interactions of transference theory do not necessarily result from human interventions.

[101] Salmon, *Scientific Explanation…*, 141-142.

However, this version of a theory of causation in terms of transference faces a decisive objection[102]. The problem is that the definitions of the "ability to transmit a mark" and of "causal process" form a vicious circle[103]. The main thesis of the theory says that a causal process is a process that is able to transmit a mark. However, the notion of a mark can only be defined by using the concept of a causal process. For it is necessary to require that only *causal* interactions be able to introduce a mark in a process. A mark is an alteration of a property due to a local singular interaction. The requirement that this interaction be causal is necessary, for otherwise the intersection between two pseudo-processes could introduce a mark, and this would contradict the very idea of a pseudo-process.

To escape this circle, Dowe[104] has proposed that two events are linked as cause and effect if and only if they manifest the same amount of a conserved quantity (from now on "CQ"). Dowe's account is explicitly opposed to the idea that such amounts of CQ can be transferred, and cannot therefore be considered as a version of a transference theory of causation. But it seems natural to use the concept of amount of a conserved quantity to generalize a traditional variant of the transference theory according to which causal relations are transmissions of energy[105]. In the context of the contemporary conception of conserved quantities grounded on symmetries, it seems arbitrary to ground causation in the transmission of energy exclusively, as it would be to ground it in any other particular set of quantities, such as momentum[106], force[107], or information[108]. The idea of generalizing to any amount of any CQ as the material ground of causal relations can be justified by reflection on the concept of transference itself. Only what is con-

---

[102] Even the theory's major advocate, Wesley Salmon, has abandoned it for this reason. Cfr. W. Salmon, *Causality Without Counterfactuals*, «Philosophy of Science», 61 (1994), 297-312.

[103] Cfr. Kitcher, P. Dowe, *Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory*, «Phil. of Science», 59 (1992), 195-216, 200-201; Kistler, *Causalité...*, 89.

[104] Dowe, *Wesley Salmon's....*

[105] This idea has been put forward since the discovery of the law of the conservation of energy in the 19th century, in particular by Julius Robert Mayer, and by the school of "energetism", in particular W. Ostwald. Cfr. W. Krajewski, *Energetic, informational, and triggering causes*, «Erkenntnis», 46 (1997), 193-202. It has more recently been taken up by K. Lorenz, *Kants Lehre vom Apriorischen im Lichte gegenwärtiger Biologie*, «Blätter für deutsche Philosophie», 15 (1941), and in the literature in English by Aronson (J. J. Aronson, *The Legacy of Hume's Analysis of Causation*, «Studies in the History and Philosophy of Science», 2 (1971), 135-165; J. J. Aronson, *On the Grammar of 'Cause'*, «Synthese», 22 (1971), 414-430) and Fair (D. Fair, *Causation and the Flow of Energy*, «Erkenntnis», 14 (1979), 219-250).

[106] Cfr. Fair, *Causation...*

[107] Cfr. Aronson, *The Legacy...*

[108] Cfr. W. Krajewski, *Four Conceptions of Causation*, in W. Krajewski (ed. by), *Polish Essays in the Philosophy of the Natural Sciences*, Reidel, Dordrecht 1982.

served during transmission can be transmitted. This reasoning leads to the proposal to reduce the causal relation between two particular events to the transmission between them of some amount of a CQ[109]. It is up to science to establish the explicit list of all the CQ that exist in nature and that can therefore ground causal relations.

This idea can be elaborated so as to provide a satisfactory account of the situations that raise insuperable problems for the theories we examined earlier. The first qualification concerns the asymmetry of causation. The concept of transference is symmetrical: the individual amount of CQ X is transferred from a to b if it is present both in event a and in event b. In a possible world in which all processes are reversible, there is transmission and therefore causation, but causal relations (and with them, time, if the direction of time is reduced to the direction of causation) are symmetric. In such a world, there is no intrinsic difference between the two terms of a causal relation, which makes one the cause rather than the other. However, in the actual world, most processes are irreversible. This irreversibility is at the origin of the asymmetry of both causation and time. Borrowing from Reichenbach[110] the concept of a causal net in which all events are causally linked to other events, so that no event is causally absolutely isolated, it is possible to consider that insofar as a large majority of the processes in the net possess the same intrinsic direction, which is the direction of growing entropy, that direction is the direction of the entire net. In this way, it is possible to attribute a direction even to intrinsically symmetrical processes, i.e. to reversible processes.

The problem of effects can then be solved by appealing to the empirical fact that the majority of actual causal processes have the same intrinsic direction: a is cause of b because first, an individual amount of a CQ is transmitted between them and second, because either the transmission process between a and b is physically asymmetrical and irreversible or it is reversible but part of a network of causal processes possessing a global direction and imposing it on the transmission process between a and b.

The transference theory overcomes the problems of epiphenomena, preemption and causal overdetermination rather simply. There is no transmission whatsoever between the events of Jones and Smith watching the same TV program. Although these events are perfectly correlated qualitatively, special relativity forbids transmission between them, which would need to have superluminal

[109] Cfr. M. KISTLER, *Reducing Causality to Transmission*, «Erkenntnis», 48 (1998), 1-24; KISTLER, *Causalité…*, chap. 1.

[110] REICHENBACH, *The Direction of Time.*

speed[111]. This provides a simple explanation for the absence of a causal relation between them. In the case of preemption considered above, the excess weight has transmitted kinetic energy to the bridge, which is transformed into deformation energy during the collapse. The explosion of the bomb has transmitted energy, but not to the bridge, for when the shock wave transmitting this energy arrives, the bridge has already (started to) collapse.

However, transference theory faces its own problems. Sosa and Tooley judge that «the prospects for a physicalistic reduction of causation do not [...] appear very bright»[112], for three reasons. The first two concern only Fair's version of the theory[113], but not the version sketched above. Sosa and Tooley's third argument, however, raises a challenge with respect to clarifying the conceptual status of the theory. Their first objection to transference theories is that the causal relation can only be reduced to a type of physical relation if the latter has a direction that is not itself causally grounded. It is indeed plausible to think that the *common sense* concept of transmission is only asymmetrical because its meaning has a causal and probably even anthropomorphic aspect, its paradigmatic application being to *actions* of transference of an object from one place to another. To avoid both circularity and anthropomorphism, Fair proposes to ground causal asymmetry on temporal asymmetry. Of two events linked by transmission, the earlier causes the later. However, we have already noted that first, this strategy makes the reduction of the asymmetry of time to the asymmetry of causation impossible and second, that is stipulates for a priori reasons that there is no backwards causation, whereas this seems to be an empirical issue. We have already indicated the solution to this difficulty. It is to conceive of transference of an amount of a CQ as the presence of the individual amount in two different events. This conception is neither implicitly causal nor anthropomorphic. However, it is symmetrical. Within this conception, the asymmetries both of causation and of time (i.e. time's having a privileged direction) are not a priori consequences of the concept of causation. Rather, their asymmetry is a contingent property of the causal relations of the actual world, grounded in the fact that a large majority of actual causal processes are intrinsically asymmetrical and irreversible: they are processes with growing entropy[114].

The second objection against transference theories that Sosa and Tooley men-

---

[111] The distance between the events of reception is called "space-like", whereas only "time-like" distances can be bridged by causal processes, in particular by signals.

[112] E. Sosa and M. Tooley, *Introduction*, in E. Sosa and M. Tooley (ed. by), *Causation*, Oxford University Press, Oxford 1993, 4.

[113] D. Fair, *Causation and the Flow of Energy*, «Erkenntnis», 14 (1979), 219-250.

[114] There exist also microphysical processes that are intrinsically asymmetrical, in particular the disintegration of K-mesons, also called "kaons". See Dowe, *Wesley Salmon's...*, 189.

tion concerns the causal theory of persistence. There are two major conceptions of the persistence of macroscopic objects in time. According to "three-dimensionalism" the identity of ordinary objects is determined only in the three spatial dimensions but not in the temporal dimension. Such a three-dimensional object, traditionally called a 'substance', 'endures'[115] over time without having any temporal parts, simply by being entirely present at every moment of its existence. The idea of something being entirely present at many instants of time has its major application in the traditional conception of universals, according to which a universal, say red, is entirely present in every one of its instances, in red objects. Both substances and universals can be entirely present at more than one place and time. However, this is inconceivable for four-dimensional entities, such as events, whose identity is determined according to both space and time. An event taking place within a space-time region cannot be present at a different space-time region. Only its *parts* occupy different space-time regions. According to "four-dimensionalism" ordinary objects are extended both in space and time, and have temporal parts. It conceives of the persistence of an object over time as of "perdurance", based on a relation between its temporal parts.

For three-dimensionalism, the identity of an object though time, its «genidentity»[116], is a true primitive and irreducible identity. For four-dimensionalism, however, it is no identity in the strict logical sense; it is rather based on a relation between the different temporal parts of the object, where the relation may be conceived as causal. However – and this is Sosa and Tooley's second objection against Fair's version of the transference theory – insofar as the causal relation is conceived on the model of transference between different *objects*, the relation between different temporal parts of *one* object cannot be conceived as being causal; hence, the four-dimensional (or "perdurance") view of persistence through time loses its most plausible foundation. However, the version of transference theory sketched above, which takes the terms of causal relations to be *events*, and not *objects*, avoids this objection: it can explain the persistence of objects on the basis of a causal relation between the temporal parts of the object, which are events, thanks to the transmission of amounts of different CQ, first of all mass-energy.

The third objection concerns the status of the theory. Transference theory implies a radical change in the very conception of the problem of the analysis of the concept of causation, insofar as it does not try to provide a purely conceptu-

---

[115] The terms 'enduring' and 'perduring' (see below) have been introduced by Mark Johnston. Cfr. D. LEWIS, *On the Plurality of Worlds*, Blackwell, Oxford 1986.

[116] On the notion of genidentity, see M. KISTLER, *Le concept de génidentité*, in S. LAUGIER (éd. par), *Carnap et la construction logique du monde*, Vrin, Paris 2001, 163-188.

al and a priori analysis, as the alternative theories do. Rather, its fundamental hypothesis is that the causal relation is a natural kind of relation whose nature needs to be discovered a posteriori, by empirical and in principle testable speculation. In traditional terminology, it aims to discover the "real essence" of the causal relation[117]. The discovery of telepathy or other cases of (simultaneous) action at a distance would, e.g., constitute an empirical refutation of transference theory, whereas the alternative conceptions can only be refuted by a priori conceptual arguments. The instantaneous correlations between distant but "entangled" physical systems predicted by quantum mechanics and experimentally observed, do not refute it because they are not causal. They cannot be used to serve actions at a distance (instantaneously bringing about their effect), such as signalling or producing changes.

Sosa and Tooley's thesis that «causation possesses an intrinsic nature, so that causation must be one and the same relation in all possible worlds»[118], can be accepted, on condition of adding the qualification that causation has this essential nature in all possible worlds *in which it exists*. However, the sketched version of transference theory no longer conceives this essence, as Sosa and Tooley do following tradition, as a nominal essence that constitutes the a priori content of the concept of causation. Rather it conceives it as a real essence to be discovered empirically, a posteriori. This implies that causation does not exist necessarily. There are possible worlds in which no relation belongs to the natural kind causation, just as some possible worlds do not contain any object belonging to the natural kind gold. However, this thesis does not force us to conclude, as Sosa and Tooley claim it does, that

> «one can appeal to the possibility of worlds that involve causation, but that do not contain the physicalistic relations in question [such as the relation of transference of an amount of a CQ; M.K.] – or, more radically, that contain no physicalistic states at all – in order to draw the conclusion that causation cannot, even in this world, be identical with any physicalistic relation»[119].

If causation is a natural kind, such worlds that would refute the conception we have proposed, do not exist any more than there exist possible worlds containing gold that does not have the atomic number 79[120]. In such worlds, there might

---

[117] For the idea of construing causation as a "natural kind" of relation, I am indebted to E.J. Lowe (personal communication).

[118] Sosa and Tooley, *Introduction*, 3.

[119] Sosa and Tooley, *Introduction*, 3.

[120] Kripke (S.A. Kripke, *Naming and Necessity* [1972], Harvard University Press, Cambridge, MA 1980) has shown that natural kinds, such as gold, possess a real essence, which is however discovered a

be a different source of correlations, but it would not be causation: it would not share the real essence of the causal relation of the actual world.

6. Causal responsibility

We have seen that the nomological, counterfactual and probabilistic accounts are not up to the task of analyzing the singular causal relation between events. We have also found that they are instead adequate to the analysis of nomological dependence, not causation. However, it is possible to acknowledge the usefulness of these approaches within the framework of a richer concept of causation, which takes nomological dependence into account and which deals with a relation between facts rather than events. If a white billiard ball A moves a red billiard ball B (by a frontal elastic shock, where B was at rest before the shock and A and B have equal masses), we have a paradigmatic case of causation. However, we can characterize this relation more finely, because certain properties of the cause are responsible for certain properties of the effect, whereas others are not. The fact that A has momentum M at the moment of its impact on B is causally responsible for the fact that B has momentum M after the shock. However, the fact that A is white before the shock is not causally responsible for the fact that B is red after the shock. The dependence of the effect's possession of certain properties on the possession of certain properties by the cause is nomological and gives rise, first to counterfactual dependence between facts about cause and effect involving these properties, and second to a relation of probability raising. If A had not had M before the shock, B would not have had M after the shock. However, if A had not been white, B would nevertheless have been red. The probability that B has M at t is greater if it lies on the trajectory of A approaching with M than if B doesn't lie on it. However, the probability that B is red after t is the same whether or not A is present.

In this way, the relation of *causal responsibility*[121] allows us to integrate the contribution of the nomological, counterfactual and probabilistic approaches to

posteriori. There is no possible world in which gold does not have its essential properties, such as its atomic number, 79.

[121] Similar relations have been studied by other authors (T. HORGAN, *Mental Quausation*, in J.E. TOMBERLIN (ed. by), *Philosophical Perspectives 3: Philosophy of Mind and Action Theory*, Ridgeview, Atascadero, CA 1989; A. MARRAS, *Kim's Principle of Explanatory Exclusion*, «Australasian Journal of Philosophy», 76 (1998), 439-451; B. MCLAUGHLIN, *On Davidson's Response to the Charge of Epiphenomenalism*, in J. HEIL and A. MELE (eds.), *Mental Causation*, Clarendon Press, Oxford 1993, 27-40; J. KIM, *The Non-Reductivist's Troubles with Mental Causation*, in HEIL and MELE, *Mental Causation*, 189-210) under the name 'qua-causation' or 'quausation'. Still others consider the relation between facts to be the causal relation itself. «Token causal claims are about the causal role of the actual *exemplification of one type* for

causation within the framework of the transference theory. The relation of causal responsibility of the fact that the cause event (*c*) has property F for the fact that the effect event (*e*) has property G can be written as a two place relation between the facts Fc and Ge: CR (Fc, Ge).

To a first approximation, the relation of causal responsibility can be analyzed as a complex relation grounded on the existence of a causal relation between events *c* and *e*[122]. Building upon a nomological conception of the dependency between properties[123], the relation of causal responsibility CR (Fc, Ge) of fact Fc about the cause for fact Ge about the effect may be analyzed into a conjunction C(*c,e*) ∧ Fc ∧ Ge ∧ L(F,G). C(*c, e*) means that the events *c* and *e* are linked as cause and effect in virtue of transference; Fc and Ge mean that *c* and *e* possess properties F and G, respectively; L(F, G) means that the exemplification of F by *e* depends nomologically on the exemplification of F by *c*.

7. Conclusion

Contemporary analytic philosophy of science is in the midst of a vigorous debate on the nature of causation. Each of the main proposals discussed in this chapter faces important problems. Among the most actively pursued lines of research aimed at overcoming these problems, are several attempts to combine the insights of these different approaches. For lack of space, I have sketched only one of them[124] in the last section[125].

the actual *exemplification of another type*. [...] 'y's being Y was *because of* (or *token caused by*) x's being X'» (EELLS, *Probabilistic Causality*, 288; Eells' italics).

[122] I have defended this conception in KISTLER, *Causalité…*, chap. 5, and M. KISTLER, *Causation as transference and responsibility*, in W. SPOHN, M. LEDWIG and M. ESFELD (eds.), *Current Issues in Causation*, Mentis, Paderborn 2001, 115-133. In M. KISTLER, *Causes as events and facts*, «Dialectica», 53 (1999), 25-46, I defend it against arguments by Mellor (MELLOR, *The Facts of Causation*) and others, to the effect that *only* facts, but not events, can be causally related.

[123] Alternatively, this dependency could be conceived according to the models of counterfactual dependency or probability raising.

[124] Other proposals include P. DOWE, *A Counterfactual Theory of Prevention and "Causation" by Omission*, «Australasian Journal of Philosophy», 79 (2001), 216-226, and J. SCHAFFER, *Causes as probability raisers of processes*, «Journal of Philosophy», 98 (2001), 75-92.

[125] I am indebted to Richard Carter for improving my English.