PAST AND CONTEMPORARY PERSPECTIVES ON EXPLANATION

Stathis Psillos

> The word *explanation* occurs so continually and holds so important a place in philosophy, that a little time spent in fixing the meaning of it will be profitably employed.
>
> John Stuart Mill

0. INTRODUCTION

In spite of Mill's guarded optimism that fixing the meaning of *explanation* is a task that requires "a little time", the truth is that the time and energy spent on this task in the history of philosophy have been enormous. Though this time and energy have been profitably spent, we do not yet know what exactly explanation is. There is no single and definite meaning attached to the word *explanation.* There is no fully adequate model of explanation that covers everything we think, intuitively, an explanation consists in. We are not even clear on what are the core platitudes that the concept of explanation must satisfy. Yet, we know that the more we think of it, the more the concept of explanation is shown to be indispensable to the ways in which we think and act.

This is just a small sample of the questions and issues that are involved in explaining explanation. Explanation is intimately linked with causation and laws, but how exactly? Does all explanation have to be causal? Does all causal explanation have to be nomological? Do explanations have to be arguments whose conclusion is the fact/event to be explained (the *explanandum*) and whose premises (the *explanans*) have to cite laws of nature? If they do, does explanation consist in the deductive demonstration of the *explanandum* or is an inductive link between the *explanans* and the *explanandum* enough? If explanations are not arguments, what is their logical form? Is it enough to say that they are causal stories linking the purported cause with the effect? Shouldn't these causal stories be backed up with laws? And how are laws themselves explained? Is it enough to say that more fundamental laws explain the less fundamental ones by subsuming the latter under them? What, in the end, does this idea of fundamental laws amount to? Is part of the intuitive meaning of explanation that to explain a number of apparently disparate facts (or laws) is to unify them under a small set of unexplained

Handbook of the Philosophy of Science: Focal Issues
Volume editor: Theo Kuipers. Handbook editors: Dov M. Gabbay, Paul Thagard and John Woods.

explainers? Are there special patterns of explanation that are suitable for purposeful and intentional behaviour? Or are teleological and intentional patterns of explanations just variants of the causal/nomological patterns? Might it be more profitable or appropriate to focus on the *act* or the *process* of explaining, instead of the *product* of explanation. If an explanation is, ultimately, an answer to a why-question, shouldn't it be the case that the *relevant* answers will depend on the presuppositions or the interests of the questioner, on the space of alternatives, and, in general, on the context of the why-question? If this is so, can there be an objective conception of what an explanation is?

Without aspiring to address all of the above issues and questions, the present essay aims to show how our thinking about explanation has evolved and where it stands now. Its first part presents how some major thinkers, from Aristotle to Mill, conceived of explanation. The second part offers a systematic examination of the most significant and controversial contemporary models of explanation.

Despite the more than two millennia that separate Aristotle's thinking from ours, Aristotle's conception — the thought that explanation consists in finding out *why* something happened and that answering why-questions requires finding causes — set the agenda for almost all subsequent thinking about explanation. For better or worse, causal explanation had been taken to be the model of explanation. The rivalry had been between those who thought that all causal explanation must proceed in terms of efficient causation and those who (following closely on Aristotle's footsteps) thought that there is room (and need for) teleological explanation (that is, for explanation that cites final causes). Most modern philosophers revolted against all but efficient causation. The latter was taken to be the *only* type of causation by (almost) all those who advocated, in one form or another, the mechanical philosophy: in their hands, efficient causation became tantamount to pushings and pullings. Final causes, in particular, were cast to the winds. Where Aristotle saw goals and purposes in nature, mechanical philosophers either excised all purpose from nature (Hobbes, Hume) or placed it firmly in the hands of God (Descartes). It was mainly Gottfried Leibniz who tried to reconcile efficient (mechanical) causation with final causation.

For Aristotle, causal explanation is captured by *demonstrative arguments* of certain sorts: those that respect the asymmetry between cause and effect. In his hands, causal explanation and demonstration became one. What is more, Aristotle embedded his account of explanation into a rich ontological framework that included essences, substantial forms, powers, activities and so on. Most moderns revolted, to varying degrees, against this rich ontological landscape. From René Descartes onwards, the idea gained momentum that causal explanation proceeds by subsuming the events to be explained under general laws. Causation was intertwined with the presence of laws and explanation was taken to consist in a law-based demonstration of the *explanandum.* However, two key Aristotelian ideas, that there is necessity in nature and that this necessity is the same as the logical necessity of a demonstrative argument, remained part of the mainstream philosophical thinking about causation and explanation until David Hume subjected them to severe criticism and undermined them. In a sense, Hume was the first to remove the efficiency from efficient causation: causation just is regular succession — one thing following another. In doing this,

he was the first to free *causation* and *explanation* from the metaphysical fetters that his predecessors had used to pin them down. Immanuel Kant reacted to Hume by trying to secure the metaphysical foundations of the fundamental laws of nature, but it was Mill who pushed the Humean project to its extreme by offering a well worked out model of scientific explanation based on the idea that there is no necessity in nature and that, ultimately, explanation amounts to unification into a comprehensive deductive system, whose axioms capture the fundamental laws of nature.

Most of contemporary thinking about explanation has taken place against the backdrop of the views of the Logical Positivists. The key to understanding the Logical Positivist agenda on explanation is this. They thought that by sufficiently explicating the concept of explanation, they could thereby legitimise *causation*. The deep problems they saw in causation stemmed from Hume's critique of it. Ever since Hume's work, philosophers of empiricist persuasions thought that the concept of causation is too mysterious or metaphysical to be taken for granted without any further analysis. They thought that the main culprit was the idea that causation implies the existence of *necessary connections* in nature.

Explanation, the Logical Positivists thought, is different. It can be as transparent as the notions on the basis of which it can analysed, viz., deductive argument and laws of nature. This is an idea they took, almost straightaway, from Mill. Given, as they thought, that both notions (deductive argument and laws of nature) are scientifically respectable, *explanation* becomes a legitimate concept, too. Besides, if whatever is valid in the concept of causation can be captured on the basis of the concept of explanation, a valid residue of causation is preserved and demystified. The project of demystifying causation culminated in the attempts made by Carl Hempel and his followers to articulate the *Deductive-Nomological* model of explanation, the basic kernel of which is that explanation is fully understood as a species of deductive argument, with one of its premises stating a universal law of nature. Later on, Hempel and his followers advanced their project further by enlarging the kind of arguments that can be explanations so as to include *inductive* arguments (and statistical, as opposed to universal, laws), thereby hoping to capture the thrust of probabilistic (or stochastic) causation.

The irony for the Hempelian project is that what came out of the front door seemed to be re-entering from the rear window. For, as Aristotle noted, it seems that one cannot distinguish between good and bad explanations of some phenomena, *unless* one first distinguishes between causal and non-causal explanations, or better between those explanations that reflect the causal connections between the *explanans* and the *explanandum* and those that do not. It appears, then, that we need first to sort out the concept of causation and then talk about causal explanation. If this is right, the empiricist project outlined above gets things the wrong way around. Many alternative models of causal explanation that have seen the light of day since Hempel's rely on this last thought: they start with a preferred account of causation and then try to tailor causal explanation to it. A distinctive trend in this alternative approach is the rejection of the view that explanations are arguments (deductive or inductive). More recent mechanistic or interventionist approaches to causal explanation take the latter to consist in revealing underlying causal mechanisms or relations of invariance among magnitudes.

There have been some plausible ways for modern empiricists to defend the view that explanations are arguments, and in particular, the view that it is the explanatory relations that are primary and not the causal ones. Following Mill, again, it has been argued that explanation amounts to *unification*. This thought ties in well with the best empiricist view on what laws of nature are. As Mill, Frank Ramsey and David Lewis have argued, laws of nature are those regularities that are captured by the axioms and theorem of the best, in terms of simplicity and strength, deductive systematisation of our knowledge of the world. But even this view is not problem-free; and, in any case, the fertile concept of unification resists a fully adequate formulation.

Given that most of this essay is about other people's thinking about explanation, I would like to make a general point about the moral I have drawn from my own thinking about other people's thinking about explanation. Most of the thoughts and arguments in currency during the twentieth century had been put forward, in one form or another, by some past thinker from Aristotle to Mill. Perhaps with very few exceptions, the twentieth century did not add new powerful ideas. But, thanks to the unprecedented level of technical sophistication of the philosophers of the twentieth century, old and powerful ideas got a new lease of life by being more precisely formulated and more carefully worked out. This is by no means a small feat. Most of the time, the devil is in the details. And unless an idea is made precise and sharp, its strengths, limitations and possible flaws do not become visible.

PART I: A (SELECTIVE) HISTORY OF EXPLANATION

1 ARISTOTLE: EXPLANATION AS DEMONSTRATION

Aristotle thought that causal knowledge is a superior type of knowledge, the type that characterises science. He took it that there is a sharp distinction between understanding the fact and understanding the *reason* why. The latter type of understanding, which characterises explanation, is tied to finding the causes (*aitia*) of the phenomena. Though both types of understanding proceed via deductive syllogism, only the latter is characteristic of science because only the latter is tied to the knowledge of causes. He illustrated the difference between these two types of understanding by contrasting the following two instances of deductive syllogism:

> (A): Planets do not twinkle; what does not twinkle is near; therefore, planets are near. (B): Planets are near; what is near does not twinkle;

therefore, planets do not twinkle.

(A), Aristotle says, demonstrates the fact that planets are near, but does *not* explain it, because it does not state its causes. In contrast, syllogism (B) is explanatory because it gives the *reason why* planets do not twinkle: *because* they are near. Explanatory syllogisms like (B) are formally similar to non-explanatory syllogisms like (A). Both are demonstrative arguments of the form: All Fs are Gs; All Gs are Hs; therefore, all Fs are Hs. The difference between them lies in the "middle term" G. In (B), but not in (A), the middle term states a *cause.* As Aristotle says:

> The middle term is the cause, and in all cases it is the cause that is being sought (90a5-10).

To ask why F is H is to look for a causal link joining F and H. More specifically, the search for causes, which for Aristotle is constitutive of science, is the search for middle terms which will link, like a chain, the major premise of an argument with its conclusion: why is F H? Because F is G and G is H. What Aristotle observed was that, besides being demonstrative, explanatory arguments should also be *asymmetric*: the asymmetric relation between causes and effects should be reflected in the relation between the premises and the conclusion of the explanatory arguments — the premises should explain the conclusion and not the other way around.

How is explanatory (that is, causal) knowledge possible? For Aristotle, scientific knowledge forms a tight deductive-axiomatic system whose axioms are *first principles*, being "true and primary and immediate, and more known than and prior to and causes of the conclusion" (71b19-25). Being an empiricist, Aristotle thought that knowledge of causes has experience as its source. But experience on its own cannot lead, through induction, to the first principles: these are universal and necessary and state the ultimate causes. On pain of either circularity or infinite regress, the first principles themselves cannot be demonstrated either. Something besides experience and demonstration is necessary for the knowledge of first principles. This is a process of abstraction based on intuition, a process that reveals the essences of things, that is the properties by virtue of which the thing is *what it is* (cf. 1140b31-1141a8).

Aristotle calls the first principles "definitions". Yet, they are not verbal: they do not just state what words mean; they also state the essences of things. In the example (B) above, it is of the essence of something's being near that it does not twinkle. In the rich Aristotelian ontology, causes, i.e., middle terms of explanatory arguments, are essential properties of their subjects and necessitate their effects. Accordingly, causal explanation is explanation in terms of essences and essential properties, where "the essence of a thing is what it is said to be in respect of itself" (1029b14). He thought that the logical necessity by which the conclusion follows from the premises of an explanatory argument mirrors the physical necessity by which causes produce their effects.

Though Aristotelian explanations are arguments, that is, ultimately, linguistic constructions, Aristotle favoured an *ontic* conception of explanation. This is because he tied explanation to causation: it is the causes that do the explaining. He distinguishes between four types of causes. The material cause is "the constituent from which something comes to be"; the formal cause is "the formula of its essence"' the efficient cause is "the source of the first principle of change or rest"; and the final cause is "that for the sake of which" something happens (194b23-195a3). For instance, the material cause of a statue is its material (e.g., bronze); its formal cause is its form or shape; its efficient cause is its maker; and its final cause is the purpose for which the statue was made. These different types of a cause correspond to different answers to why-questions. But Aristotle thought that, *ceteris paribus*, a complete causal explanation has to cite all four causes (that is, to answer all four why-questions): the efficient cause is the active agent that puts the form on matter for a purpose. The four causes do not explain the same *feature* of the object (e.g., the material cause of the statue — bronze — explains why it is solid, while its formal cause explains why it is only a bust), yet they all contribute to the explanation of the features of the very same object. All four types of cause can be cast as middle terms in proper causal explanations (cf. 94a20-25).

2 DESCARTES: MECHANICAL EXPLANATION

In *Principles of Philosophy* (1644), Descartes expanded on the Aristotelian idea that explanation consists in demonstrations from first principles. But he gave this idea two important twists. The first is that the basic principles are the fundamental rules or laws of nature. The second was the idea that all explanations of natural phenomena is mechanical. Like Aristotle, Descartes thought that explanation amounts to the search of causes. But unlike Aristotle, he thought that all causation is efficient causation and, in particular mechanical. Though Descartes did not fully abandon the rich Aristotelian philosophical framework, (for instance, he too conceived of the world in terms of substances, natures, essences and necessary connections, the latter being, by and large, a priori demonstrable), he thought that the explanation of natural phenomena proceeds by means of mechanical interactions, and not by reference to violent and natural motions; nor in teleological terms. To be sure, he took God to be "the efficient cause of all things" [1985, 202]. But in line with the scholastic distinction between primary cause (God) and secondary causes (worldly things), he claimed that the secondary and particular causes — "which produce in an individual piece of matter some motion which it previously lacked" [1985, 240] — are the laws of nature and the initial conditions viz., the shapes, sizes and speeds of material corpuscles.

Descartes was not a pure rationalist who thought that *all* science could be done a priori. But he was not, obviously, an empiricist either. He did not think that all knowledge stemmed from experience. As he claimed in *Principia*, the human mind, by the light of reason alone, can arrive at substantive truths about the world, concerning mainly the fundamental laws of nature. These, for instance that the total quantity of motion in the world is conserved, are discovered and justified a priori, as they are supposed to stem directly from the immutability of God. Accordingly, the basic structure of the world is discovered independently of experience, is metaphysically necessary and known with metaphysical certainty; for instance, that the world is a plenum with no vacuum (or atoms) in it, that all bodies are composed of one and the same matter, that the essence of matter is extension etc. It is on the basis of these fundamental laws and principles that all natural phenomena are explained, by being deduced from them. Accordingly, Cartesian causal explanation is nomological explanation. More precisely, causal explanation consists in finding nomologically sufficient causes of the effects. In this sense, causal explanations are demonstrative arguments whose premises include reference to laws of nature.

How is then empirical science possible? Descartes thought that once the basic nomological structure of the world has been discovered by the lights of reason, science must use hypotheses and experiments to fill in the details. This is partly because the basic laws of nature place constraints on whatever else there is and happens in the world, without determining it uniquely. The initial conditions (the shapes, sizes and speeds of corpuscles) can only be determined empirically. That is, among the countless initial conditions that God might have instituted, only experience can tell us which he has actually chosen for the actual world. Besides, though grounded in the fundamental laws, the less fundamental laws of physics are not immediately deducible from them. Further hypotheses are needed to flesh them out. Hence, Descartes thought that the less fundamental laws could be known only with moral certainty.

Indeed, Descartes allowed for the possibility that there are competing systems of hypotheses which, though compatible with the fundamental laws, offer different explanations of the phenomena. He illustrated this possibility by reference to an artisan who produced two clocks that indicate the hours equally well, are externally similar and yet work with different internal mechanisms. In light of this possibility, Descartes wavered between two thoughts, which were to become the two standard responses to the argument from underdetermination of theories by evidence. The first (cf. [1985, §44]) is that it does not really matter which of the two competing systems of hypotheses is true, provided that they are both empirically adequate, that is, they correspond accurately to all the phenomena of nature. The other (cf. [1985 §§44 and 205]) is that theoretical virtues such as simplicity, coherence, unity, naturalness etc. are marks of truth in the sense that it would be very unlikely that a theory possesses them and be false. Interestingly, Descartes put a premium on novel predictions: when postulated causes explain phenomena not previously thought of, there is good reason to think they are their true causes.

Explanatory hypotheses, Descartes claimed, must be mechanical, that is cast in terms of “the shape, size, position and motion of particles of matter” [1985, 279], and that the selfsame mechanical principles should deductively explain the whole of nature, both in the heavens and on the earth. It wouldn’t be an exaggeration to claim that Descartes advanced an unificationist account of explanation, where the unifiers are the fundamental laws of nature.

Famously, Descartes distinguished all substances into two sorts: thinking things (*res cogitans*) and extended things (*res extensa*). He took the essence of mind to be thought and of matter extension. Unlike Aristotle, he thought that matter was inert (since its essence is that it occupies space). Yet, there are causal connections between bodies (bits of matter) and between minds and bodies (that is, between different substances). Two big questions, then, emerge within Cartesianism. The first is: how is body-body interaction possible? The second is: how is mind-matter interaction possible? Briefly put, Descartes' answer to the first question is the so-called *transference* model of causation: when x causes y a property of x is communicated to y. He thought that this view is an obvious consequence of the principle "Nothing comes from nothing". As he put it:

> For if we admit that there is something in the effect that was not previously present in the cause, we shall also have to admit that this something was produced by nothing. [1985, Vol. 1, 97]

But Descartes failed to explain how this communication is possible. Indeed, by taking matter to be an inert extended substance, he had to retreat to some external cause of motion and change and ultimately to God himself. This retreat to God cannot save the transference model. Besides, the transference model makes an answer to the second question above (how do mind and matter interact?) metaphysically impossible. Being distinct substances, they have nothing in common which can be communicated between them. In a sense, Descartes was a failed interactionist: there is matter-matter and mind-matter causal interaction but there is no clear idea of how it works.

Descartes' successors were divided into two groups: the occasionalists and those who, following Leibniz reintroduced *activity* into nature. Occasionalism is the view that the only real cause of everything is God and that all causal talk which refers to finite substances is a sham. Nicholas Malebranche drew a distinction between real causes and natural causes (or occasions). As he put it:

> A true cause as I understand it is one such that the mind perceives a necessary connection between it and its effect. Now the mind perceives a necessary connection between the will of an infinite being and its effect. Therefore, it is only God who is the true cause and who truly has the power to move bodies. [1674-5/1997, 450]

Natural causes are then merely the occasions on which God causes something to happen. Malebranche pushed Cartesianism to its extremes: since a body's nature is exhausted by its extension, bodies cannot have the power to move anything, and hence to cause anything to happen. What Malebranche also added was that since causation involves a necessary connection between the cause and the effect (a view that Descartes accepted too), and since no such necessary connection is perceived in cases of alleged worldly causation (where, for instance, it is said that a billiard ball causes another one to move), there is no worldly causation: all there is in the world is regular sequences of events, which strictly speaking are not causal. For Malebranche, all causal explanation must ultimately refer to God and his powers.

3 LEIBNIZ: POWERS AND TELEOLOGY

As noted already, Leibniz tried to reintroduce forces and powers into nature. He thought that the rejection by his contemporaries of the scholastic philosophy had gone too far. Though he too favoured mechanical explanations of natural phenomena and denounced occult qualities and virtues as non-explanatory, he thought wrong the key Cartesian thought that the essence of matter was extension. Extension cannot account for the presence of activity in nature. In *Discourse on Metaphysics* (1686), he argued that the essence of substance is activity. He then found it compelling to appeal to substantial forms as the individuating principles that explain the unity of each substance and the variation among different substances. These substantial forms (which he took them to be vital principles analogous to the souls of human bodies), Leibniz thought, were indispensable in metaphysics (especially in teleological explanations of the phenomena) yet they must not be employed in the explanation of particular events [1686, §X]. The latter should proceed by mechanical demonstrations. But Leibniz was not content with the prevailing mechanistic explanations of phenomena. He thought that the mechanical principles of nature need metaphysical grounding and that they should be supplemented by dynamical explanations in terms of forces and powers. Leibniz insisted that every substance is essentially active; since whatever acts is force, every thing is force or a compound of forces. Indeed, Leibnizian substances (what he called "the monads") are sustained by internal "primitive active forces" which cause their subsequent states.

Like Descartes before him, Leibniz too thought that explanation consists in demonstrations from premises that comprise (descriptions of) the fundamental laws of nature. And like Descartes, he thought that the fundamental laws of nature stemmed directly from God. Yet he drew a distinction between the most fundamental law of nature, viz., that nature is orderly and regular, and "subordinate regularities" such as his three laws of motion. "The universal law of the general order", as he put it, is metaphysically necessary, since in whatever way God might have created the world, "it would always have been regular and in a certain order" [1686, §VI]. It is important to note that the three basic Leibnizian laws of motion are conservation laws.[1] Hence, being invariant, they preserve the fundamental order of nature. The subordinate laws are metaphysically *contingent*, since they might well differ in other possible worlds. Yet, Leibniz thought that among all possible worlds, God has created the most perfect one: the actual world

[1] Translated into modern idiom, these laws state: (i) the conservation of *vis viva* in every impact; (ii) the conservation of the directed quantity of motion in every impact; and (iii) the conservation of relative velocity before and after impact.

is the most perfect of all possible worlds, and it is such that it is the simplest in laws and the richest in phenomena. Hence we may reasonably conclude that Leibniz took laws to be the simplest and strongest set of principles that allow the deduction of all phenomena. Though metaphysically contingent, the subordinate laws are physically necessary (since, as Leibniz put it, denying them would imply an imperfection on the part of God [1973, 139]). Under these laws fall others of an even lower level (cf. [1973, 99])

It was a key Leibnizian thought that all (mechanical) laws of nature need metaphysical grounding. As he put it:

> Nature must always be explained mathematically and mechanically, provided it be kept in mind that the principles or the laws of mechanics and of force do not depend upon mathematical extension alone but have certain metaphysical causes. [Letter to Arnauld, 14 July, 1686]

This point is also made in [1686 §XVIII], where he adds that the general principles of corporeal nature and of mechanics are metaphysical rather than mathematical and "belong rather to certain indivisible forms or natures as the causes of the appearances, than to corporeal mass or to extension". Apart from the universal law of the general order, these principles include "the laws of cause, power and activity" [1973, 140]. They are established a priori and, interestingly, their grounding is, ultimately, teleological: they are grounded in the wisdom of God and in particular in his choice of the best possible plan in creating the actual world. The subordinate laws are then the "most fitted to abstract and metaphysical reasons" [1973, 200]. Ultimately all natural laws are explained by means of two central Leibnizian principles: *the principle of sufficient reason* and *the principle of fitness*. According to the principle of sufficient reason, for everything that happens, there must be a reason, sufficient to bring this about instead of anything else. According to the principle of fitness, the actual world is the fittest or most perfect among all possible worlds that God could have created, a fact that "the wisdom of God permits him to know; his goodness causes him to choose it and his power enables him to produce it" [1698, 55].

Leibniz did admit teleological explanations alongside mechanical ones. Apart from the need of teleological explanations (in terms of God's purposes) in metaphysics, he argued that physical phenomena can be explained by mechanical as well as teleological principles. For instance, he claimed that anatomical phenomena can be best explained in terms of goals and that many other physical phenomena (e.g., Snell's law of reflection) can be explained by teleological principles of least effort or least action [1686, §XXII]. Most interestingly, he thought that, though possible, mechanical explanations in terms of matter in motion are useless in historical explanation, where the aims, desires and intentions of historical actors are most relevant to the explanation of their actions (cf. [1686, §XIX]). Indeed, Leibniz wholeheartedly accepted the Aristotelian final causes alongside efficient causes.

He argued that these two distinct kinds of explanation — efficient and final — are reconcilable (cf. [1686, §XXII]). In the end, all things have efficient and final causes. Things have efficient causes when considered as parts of the material world and final causes when considered as substantial forms. Leibniz's reconciliation is effected by means of a third principle he enunciated, the *principle of pre-established harmony* (cf. [1973, 196]). In all its generality, this principle states that when God created this world as the best among an infinity of possible worlds, he put everything in harmony (the monads and the phenomenal world, the mind and the body, the final and the efficient causes). This principle played an important role in explaining how efficient and final causes, or the body and the mind, are co-ordinated with each other. Each domain (the domain of efficient causes and the domain of final causes; the body and the soul) obeys its own laws and does not interact with the other. Hence, they are independent and yet in accord with each other: it is *as if* they influence each other, though they do not.

It is noteworthy that Leibniz rejected the transference account of causation (what he called the "real influx" model, arguing that no impetus or qualities are transferred from one body to another (or between matter and soul). Instead, each body is moved by an innate force. For Leibniz causes are required to explain why objects exist and change and they do this by providing a reason for this existence and change (cf. [1973, 79]). But the reason (and hence the cause) is to be found in the "primitive active force" of each body, viz., in its power of acting and be acted upon (cf. [1973, 81]). Here again, the *principle of pre-established harmony* plays a key role. For it guarantees that there will be a perfect agreement between all bodies (substances), thereby "producing the same effect as would occur if these communicated with one another by means of a transmission of species or qualities (...)" [1973, 123]. Thus, the principle makes sure that there is something like causal order in the world without the existence of real influences among bodies (or among bodies and souls). In light of this, there is a sense in which Leibniz thought that there is no real causation in nature, since Leibnizian substances do *not* interact. Rather, they are co-ordinated with each other by God's act of pre-established harmony, which confers on them the natural agreement of two very exact clocks. So the only real causation admitted by Leibniz is *within* each finite substance (by means of its primitive active force) and in God who pre-establishes the harmony among substances.[2]

[2] A chief difference between Leibniz and occasionalism concerns the role and nature of miracles. Leibniz claimed that occasionalism was unsatisfactory because, by making God the real cause of every event, it introduced a continuous miracle in nature. This move, Leibniz thought, fails to explain anything since God's will is not sufficient to explain anything: Leibniz's God always has a sufficient reason to act. Leibniz's God would obey the laws of nature even when he intervened in nature. If he did not, what reason would he have to impose these laws? Miracles, for Leibniz, are quite rare. Actually, he thought that, strictly speaking, only one miracle ever happened: the creation of the world by God, and his subsequent imposition of the laws of nature. In his argument with the Newtonians, Leibniz attacked Newton with the claim that the Newtonian gravitational force comes down to a perpetual miracle.

4 NEWTON: DYNAMICAL EXPLANATION

The real break with the Aristotelian philosophical and scientific tradition occurred with the consolidation of empiricism in the seventeenth century. Empiricists attacked the metaphysics of essences and the epistemology of rational intuition, innate ideas and infallible knowledge. Sir Isaac Newton's own influence on empiricism was two-fold.

On the one hand, his own scientific achievements, presented in his monumental *Philosophiae Naturalis Principia Mathematica* (*Mathematical Principles of Natural Philosophy*, 1687), created a new scientific paradigm. The previous paradigm, Cartesianism, was overcome. Down with it went the views that space is a plenum, that there are no atoms, that the planets are carried around by vortices, that the quantity of motion (as distinct from momentum) is conserved etc. Newton extended the mechanical view of nature by systematically using the category of *force* alongside the two traditional mechanical categories, *matter* and *motion*. Force was set in a mechanical framework in which it is measured by the *change* in the quantity of motion it could generate. But Newton insisted that his concept of force was mathematical (cf. *Principia*, Book I, Definition VIII). Mechanical interactions were enriched to include attractive and repulsive forces between particles (where again, these forces were considered not physically but mathematically). The concept *mass* was clearly defined for the first time, by being distinguished from weight. Motion and rest were united: they are relative states of a body. Space became the infinite container in which motion of corpuscles takes place. The mechanics of the earth and the heavens were united: a single, mathematically simple, law of gravity governs all phenomena in the universe.

On the other hand, Newton's methodological reflections became the standard reference point for all subsequent discussion concerning the nature and aim of science and its method. Newton demanded certainty of knowledge but rejected the Cartesian route to it. By placing restrictions on what can be known and on what method should be followed, he thought he secured certainty in knowledge. As he explained, he used the term "hypothesis" "to signify only such a proposition as is not a phenomenon nor deduced from any phenomena, but assumed or supposed — without any experimental proof" (cf. [Letters to Cotes, 1713 in Thayer, 1953, 6]). And he proceeded with his famous dictum *Hypotheses non fingo* (I do not feign hypotheses), which was supposed to act as a constraint on what can be known: it rules out all those metaphysical, speculative and non-mathematical hypotheses that aim to explain, or to provide the ultimate ground of, the phenomena. As he said in the *General Scholium*, [*Principia*, Book III]

> For whatever is not deduced from the phenomena is to be called a hypothesis, and hypotheses, whether metaphysical or physical, whether of occult qualities or mechanical, have no place in experimental philosophy.

Newton took Descartes to be the chief advocate of hypotheses of the sort he was keen to deny. His official suggestion for the method of science was that it is deduction from the phenomena. This was contrasted to the hypothetico-deductive method endorsed by Descartes. Newton's approach was fundamentally mathematical-quantitative. He did not subscribe to the idea that knowledge begins with a painstaking experimental natural history of the sort suggested by Francis Bacon in his *Novum Organum* (1620). The basic laws of motion do, in a sense, stem from experience. They are not a priori true; nor metaphysically necessary. The empirically given phenomena that Newton starts with are laws (e.g. Kelper's laws). Then, by means of mathematical reasoning and the basic laws of motion further conclusions can be drawn, e.g., that the inverse square law of gravity applies to all planets.

Undoubtedly, Newton thought that the explanation of natural phenomena consists in finding the most general principles that account for them, where this relation of 'accounting for' is deductive. These general principles are the fundamental laws of nature. As he stated:

> Natural Philosophy consists in discovering the frame and operations of Nature, and reducing them, as far as may be, to general Rules or Laws — establishing these rules by observations and experiments, and thence deducing the causes and effects of things (. . .).[3]

But his views led to considerable controversy in connection, in particular, with his account of gravity. Leibniz, for instance, denounced Newtonian gravity as being an occult quality. Indeed, as Newton himself claimed: "But hitherto I have not been able to discover the cause of those properties of gravity from phenomena, and I frame no hypotheses" (ibid.).

Newton's thought was that an explanation cannot be faulted on the grounds that it does not unveil the ultimate causes of the phenomena. On the contrary, since explanations must have empirical content, they must be independently testable. Consequently, the employment of general explanatory hypotheses that transcend the limits of what is observed and inductively generalised in laws is futile and has nothing to do with the mathematical principles of natural philosophy. Newton's defence against Leibniz was that, though he had not explained the cause of gravity, he had established that gravity *is* causal (and hence that it can offer adequate causal explanations of the phenomena). As he stressed (*General Scholium*, *Principia*, Book III):

And to us it is enough that gravity does really exist and act according to the laws which we have explained, and abundantly serves to account for all the motions of the celestial bodies and of our sea.

And in Query 31 of *Optics*, he noted:

> To tell us that every species of things is endowed with an occult specific quality by which it acts and produces manifest effects is to tell us

[3] Quoted by Richard Westfall, *Never at Rest* (Cambridge: Cambridge University Press, 1980, 632).

> nothing, but to derive two or three general principles of motion from the phenomena, and afterward to tell us how the properties and actions of all corporeal things follow from those manifest principles, would be a very great step in philosophy, though the causes of those principles were not yet discovered.

Consequently, it suffices for explanation to subsume the phenomena under universal laws, even if the underlying causal mechanisms are not known. In a recent piece, McMullin [2001] has claimed that Newton offered a dynamical account of explanation placed between an agent-causal account (in terms of the powers of agents to produce effects) and a simple law-based account (in terms of subsumption under a law). Though Newton did emphasise the role of laws in explanation, he also stressed that nomological explanation should be unifying: it should subsume all phenomena under a "single sort of underlying causal agency" [2001, 298] — even if, I should add, this underlying causal agency (e.g., gravity) is not further causally explainable.

5 HUME: AGAINST THE METAPHYSICS OF EXPLANATION

All empiricists of the seventeenth century accepted nominalism and denied the existence of universals.[4] This led them to face squarely the problem of induction. Realists about universals, including Aristotle himself who thought that universals can only exist *in* things, could easily accommodate induction. They thought that after a survey of a relatively limited number of instances, the thought ascended to the universal (what is shared in common by these instances) and thus arrived at truths which are certain and unrevisable. This kind of route was closed for nominalists. They had to rely on experience through and through and inductive generalisations based on experience could not yield certain knowledge. This problem came in sharp focus in Hume's work.

The subtitle of Hume's ground-breaking *A Treatise of Human Nature* (1739-40) was: *Being an attempt to introduce the experimental mode of reasoning into moral*

[4] From Plato and Aristotle on, many philosophers thought that a number of philosophical problems (the general applicability of predicates, the unity of the propositions, the existence of similarity among particulars, the generality of knowledge and others) required positing a separate type of entity — the *universal* — along side the particulars. Universals are the features that several distinct particulars share in common (e.g. the colour red or the triangular shape). They are the properties and relations in virtue of which particulars are what they are and resemble other particulars. They are also the referents of predicates. Philosophers who are realists about universals take universals to be really there in the world, as constituents of states-of-affairs. Universals are taken to be the repeatable and recurring features of nature. When we say, for instance, that two apples are both red, we should mean, that the very same property (redness) is instantiated by the two particulars (the apples). Redness is a repeatable constituent of things in the sense that the very same redness — *qua* universal — is instantiated in different particulars. Some realists (like Plato) think that there can be uninstantiated universals (the Platonic forms) while others (like Aristotle) argue that universals can only exist when instantiated *in* particulars. Though there are many varieties of nominalism, they all unite in denying that universals are self-subsistence things. For nominalism, only particulars exist.

subjects. This was a clear allusion to Newton's achievement and method. Hume thought that the moral sciences had yet to undergo their own Newtonian revolution. He took it upon himself to show how the Newtonian rules of philosophising were applicable to the moral sciences. All ideas should come from impressions. Experience must be the arbiter of everything. Hypotheses should be looked at with contempt. His own principles of association by which the mind works, resemblance, contiguity and causation, were the psychological analogue of Newton's laws: "they are really *to us* the cement of the universe" [1740, 662].

Hume focused on causation and aimed to dissolve the issue of its metaphysical nature. Before Hume, here is how Malebranche had characterised the metaphysical state of play:

> There are some philosophers who assert that secondary [i.e., worldly] causes act through their matter, figure, motion (...) others assert that they do so through a substantial form; others through accidents or qualities, and some through matter and form; of these some through form and accidents, others through certain virtues or faculties different from the above. (...) Philosophers do not even agree about the action by which secondary causes produce their effects. Some of them claim that causation must not be produced, for it is what produces. Others would have them truly act through their action; but they find such great difficulty in explaining precisely what this action is, and there are so many different views on the matter that I cannot bring myself to relate to them. [1674-5/1997, 659]

As already noted, Malebranche found in this situation a reason to advocate occasionalism. Hume, on the other hand, presenting the situation in a way strikingly similar to the above, found in it a reason to bury the metaphysical issue altogether, to secularise causation completely and to challenge the distinction between causes and occasions. As he put it, there is "no foundation for that distinction (...) betwixt cause and occasion" [1739, 171]. In effect, Hume made the scientific hunt for causes possible, by freeing the concept of causation from the metaphysical chains that his predecessors had used to pin it down. For Hume, causation, as it is in the world, is regular succession of event-types: one thing invariably following another. His *first* definition of causation runs as follows:

> We may define a CAUSE to be 'An object precedent and contiguous to another, and where all the objects resembling the former are plac'd in like relations of precedency and contiguity to those objects, that resemble the latter'. [1739, 170]

Taking a cue from Malebranche, Hume argued that there was no perception of the supposed necessary connection between the cause and the effect. When a sequence of events that is considered causal is observed (e.g., two billiard balls hitting each other and flying apart), there are impressions of the two balls, of their motions, of their collision and of their flying apart, but there is *no* impression of any alleged necessity by which the cause brings about the effect. Hume went one step further. He found totally worthless his predecessors' appeals to the power of God to cause things to happen, since, as he characteristically said, such claims give us "no insight into the nature of this power or connection" [1739, 249]. So Hume secularised completely the notion of causation.

But Hume faced a puzzle. According to his empiricist theory of ideas, there were no ideas in the mind unless there were prior impressions (perceptions) (cf. [1739, 4]). Yet, he [1739, 77] did recognise that the ordinary concept of causation involved the idea of *necessary connection.* Where does this idea come from, if there is no perception of necessity in causal sequences? Hume argued that the *source* of this idea is the perception of "a new relation betwixt cause and effect": a "constant conjunction" such that "like objects have always been plac'd in like relations of contiguity and succession" [1739, 88]. The perception of this constant conjunction leads the mind to form a certain habit or custom. As he put it:

> after frequent repetition I find, that upon the appearance of one of the objects, the mind is *determin'd* by custom to consider its usual attendant, and to consider it in a stronger light upon account of its relation to the first object [1739, 156].

And he adds:

> 'Tis this impression, then, or *determination*, which affords me the idea of necessity.

So Hume *does* explain the idea of necessary connection in a way consistent with his empiricism. But instead of ascribing it to a feature of the natural world, he takes it to arise from *within* the human mind, when the latter is conditioned by the observation of a regularity in nature to form an expectation of the effect, when the cause is present. Indeed, Hume went on to offer a *second* definition of causation:

> A CAUSE is an object precedent and contiguous to another, and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other. [1739, 170]

Hume took the two definitions to present "a different view of the same object" [1739, 170]. The idea of necessary connection features in none of them. In fact, he thought that he had unpacked the "essence of necessity": it "is something that exists in the mind, not in the objects" [1739, 165]. He went as far as to claim that the supposed objective necessity in nature is *spread* by mind onto the world [1739, 167].

Hume placed causation firmly within the realm of experience: all causal knowledge should stem from experience. He revolted against the traditional view that the necessity which links cause and effect is the same as the logical necessity of a demonstrative argument. He argued [1739, 86-7] that there can be *no* a priori demonstration of any causal connection, since the cause can be conceived without its effect and conversely. But his far-reaching observation was that the alleged necessity of causal connection cannot be proved empirically either. As he [1739, 89-90] argued, any attempt to show, based on experience, that a regularity that has held in the past *will* or *must* continue to hold in the future too will be circular and question-begging. It will presuppose a *principle of uniformity of nature*, viz., a principle that "instances, of which we have had no experience, must resemble those, of which we have had experience, and that the course of nature continues always uniformly the same" [1739, 89]. But this principle is *not* a priori true. Nor can it be proved empirically without circularity. For any attempt to prove it empirically will have to assume what needs to be proved, viz., that since nature has been uniform in the past it *will* or *must* continue to be uniform in the future. This Humean challenge to any attempt to establish the necessity of causal connections on empirical grounds has become known as his *scepticism* about induction. But it should be noted that Hume never doubted that people think and reason inductively. He just took this to be a fundamental psychological fact about human beings (as well as higher animals) which cannot be accommodated within the confines of the traditional conception of Reason, according to which all beliefs should be justified in order to be rational.

A central target of Hume's criticism is the view that causal action (and interaction) is based on the powers that things have. As we have already seen, this view was resuscitated by Leibniz and was, partly, criticised by Newton. Hume spends quite some time trying to dismiss the view that we can meaningfully talk of powers. His *first* move is that an appeal to "powers" in order to understand the idea of necessary connection would be no good because terms such as "*efficacy*, *force*, *energy*, *necessity*, *connexion*, and *productive quality*, are all nearly synonimous" [1739, 157]. Hence, an appeal to "powers" would offer no genuine explanation of necessary connection. His *second* move is to look at his opponents' theories: Locke's, Descartes', Malebranche's and others'. The main theme of his reaction is that all these theories have failed to show that there are such things as "powers" or "productive forces". In the end, however, Hume's argument was that we "never have any impression, that contains any power or efficacy. We never therefore have any idea of power" [1739, 161]. He endorsed what might be called the *Manifestation Thesis*: there cannot be unmanifestable "powers", i.e., powers which exist, even though there are no impressions of their manifestations. This thesis should be seen as an instance of *Ockham's Razor*: do not multiply entities beyond necessity. For Hume, positing unmanifestable powers would be a gratuitous multiplication of entities, especially in light of the fact that he can explain the origin of our idea of necessity without any appeal to powers and the like.

Hume articulated the principles on which causal explanation should be based. These are his well-known "rules by which to judge of causes and effects" [1739, 173]. These principles include:

1. The same cause always produces the same effect, and the same effect never

arises but from the same cause.

2. Where several different causes produce the same effect, it must be by means of some quality, which we discover to be common amongst them.

3. The difference in the effects of two resembling causes must proceed from that particular, in which they differ.

4. An object, which exists for any time in its full perfection without any effect, is not the sole cause of that effect, but requires to be assisted by some other principle, which may forward its influence and operation.

These principles are grounded in the first one noted above, viz., *same cause, same effect*. This, Hume thought, is an empirical principle derived from experience. The second and the third principles are early versions of Mill's methods of agreement and difference. Hume's point is that causal explanation (and causal knowledge) does not require the backing of a metaphysical theory of causation. It can proceed by means of principles such as the above. He is adamant that these principles are extremely difficult in their application. But this does not imply that they are inapplicable; nor that they do not yield causal knowledge. After all, Hume denied that knowledge requires certainty.

In Hume then we see the first important philosophical step away from the metaphysics of causal explanation and towards the epistemology or methodology of causal explanation. But Hume made possible what has come to be known as the *Humean* view of causation, viz. the *Regularity View of Causation*. According to this, whether or not a sequence of events is causal depends on things that happen elsewhere and elsewhen in the universe, and in particular on whether or not this particular sequence instantiates a regularity.

6 KANT: THE METAPHYSICAL GROUNDS OF EXPLANATION

It was Hume's critique of necessity in nature that awoke Kant from his "dogmatic slumber", as he famously stated. Kant thought that Hume questioned the very possibility of science and took it upon himself to show how science was possible.

Kant rejected strict empiricism (which denied the active role of the mind in understanding and representing the world of experience) and uncritical rationalism (which did acknowledge the active role of the mind but gave it an almost unlimited power to arrive at substantive knowledge of the world based only on the lights of Reason). He famously claimed that although all knowledge starts with experience it does not arise from it: it is actively shaped by the categories of the understanding and the forms of pure intuition (space and time). The mind, as it were, imposes some conceptual structure onto the world, without which no experience could be possible. There was a notorious drawback, however. Kant thought there could be no knowledge of things as they were in themselves (*noumena*) and only knowledge of things as they appeared to us (*phenomena*). This odd combination, Kant thought, might well be an inevitable price one has to pay in order to defeat empiricist scepticism and to forgo traditional idealism. Be that as it may, his master thought was that some synthetic a priori principles should be in place for experience to be possible. And not just that! These synthetic a priori principles (e.g., that space is Euclidean, that every event has a cause, that nature is law-governed, that substance is conserved, the laws of arithmetic) were necessary for the very possibility of science and of Newtonian mechanics in particular.

Like Hume before him, Kant does not claim that reason alone can discover the connection between any specific cause and any specific effect, nor understand its necessity (cf. [1787, A195; B240-41]). He agrees with Hume that particular causal connections can be discovered only empirically. But unlike Hume, Kant *denies* that the concept of causation arises from experience and in particular that it arises in the same way as the knowledge of the causes of particular events. In his *Second Analogy of Experience*, Kant tried to demonstrate that the principle of causation, viz., "everything that happens, that is, begins to be, presupposes something upon which it follows by rule", is a precondition for the very possibility of objective experience. He took the principle of causation to be required for the mind to make sense of the temporal irreversibility that there is in certain sequences of impressions. This temporal *order* by which certain impressions appear can be taken to constitute an objective happening *only if* the later event is taken to be necessarily determined by the earlier one (i.e., to follow by rule from its cause). For Kant, objective events are not 'given': they are constituted by the organising activity of the mind and in particular by the imposition of the principle of causation on the phenomena. Consequently, the principle of causation is, for Kant, a synthetic a priori principle.

In *Metaphysical Foundations of Natural Science* (1786), Kant claimed that

> Only that whose certainty is apodeictic can be called science proper; cognition that can contain merely empirical certainty is only improperly called science.

Besides, natural science proper relies on laws that are known a priori and hold with necessity (they are not merely laws of experience). Kant thought all natural science should derive its legitimacy from its pure part, i.e., the part that contains "the a priori principles of all remaining natural explications". He took as his task to show that these a priori principles of pure natural science are certain and necessary for the very possibility of science and experience. This, he thought, was the task of the metaphysics of nature. Unlike Newton, Kant thought that there could not be proper science without metaphysics. Yet, his own understanding of metaphysics was in sharp contrast with that of his predecessors (Leibniz's in particular). Metaphysics, Kant thought, was a science, and in particular *the science of synthetic a priori judgements*. Mathematics was taken to be the key element in the construction of natural science proper: without mathematics no doctrine concerning determinate natural things was possible. The irony, Kant thought, was that though many past thinkers (and Newton in particular) repudiated metaphysics and had relied on mathematics in order to understand nature, they failed to see that this very reliance on mathematics made them unable to dispense with metaphysics. For, in the end, they had to treat matter in abstraction from any particular experiences. They postulated universal laws without inquiring into their a priori sources.

As Kant argued in *Critique of Pure Reason* (1781), the a priori source of the universal laws of nature was the transcendental principles of pure understanding. These constitute the object of knowledge in general. Thought (that is, the understanding) imposes upon objects in general certain characteristics in virtue of which objects become knowable. The phenomenal objects are constituted as objects of experience by the schematised categories of quantity, quality, substance, causation and community. If an object is to be an object of experience, it must have certain necessary characteristics: it must be extended; its qualities must admit of degrees; it must be a substance in causal interaction with other substances. In his three Analogies of Experience, Kant tried to prove that three general principles hold for all objects of experience: that substance is permanent; that all changes take place in conformity with the law of cause and effect; that all substances are in thoroughgoing interaction. These are synthetic a priori principles that make experience possible. They are imposed a priori by the mind on objects.

Yet, these transcendental principles make no reference to any experienceable objects in particular. It was then Kant's aim in *Metaphysical Foundations of Natural Science* to show how these principles could be concretised in the form of laws of matter in motion. These were metaphysical laws in that they determined the possible behaviour of matter in accordance with mathematical rules. Kant thus enunciated the law of conservation of the quantity of matter, the law of inertia and the law of equality of action and reaction and thought that these laws were the mechanical analogues (cases *in concreto*) of his general transcendental principles. They determine the pure and formal structure of motion, where motion is treated purely mathematically *in abstracto*. It is no accident, of course, that the last two of these laws are akin to Newton's law and that the first law was presupposed by Newton too. Kant's metaphysical foundations of (the possibility of) matter in motion were precisely meant to show how Newtonian mechanics was possible. But Kant also thought that there are physical laws that are discovered empirically. Though he held as a priori true that matter and motion arise out of repulsive and attractive forces, he claimed that the particular force-laws, even the law of universal attraction, can only be discovered empirically.

His predecessors, Kant thought, had failed to see this hierarchy of laws that make natural science possible: transcendental laws that determine the object of possible experience in general; metaphysical laws that determine matter in general and physical laws that fill in the actual concrete details of motion. Unlike the third kind, the first two kinds of law require a priori justification and they are necessarily true.

Overall, then, Kant was mostly concerned with the metaphysical foundations of causal explanation, viz., that causal explanation presupposes necessary connections. Given that causation is nomological, Kant's thought amounted to the claim that all causal explanation is nomological explanation. But, especially towards the end of *Critique*, he highlighted another important dimension of explanation, viz., unification. He claimed it to be a "regulative idea" that nature is unified and uniform. He took it that reason aims to systematise its body of knowledge, i.e., "to exhibit the connection of its parts in conformity with a single principle" [A645/B673]. It is this systematic unity of knowledge that shifts it from being "a mere contingent aggregate" to being "a system connected according to necessary laws". This "systematic unity of knowledge" is "*the criterion of the truth* of its rules" [A647/675]. As an example of this, he offered the subsumption of more specific (causal) powers under more fundamental powers. This subsumption, he thought, "claims to have an objective reality, as postulating the systematic unity of the various powers of a substance (...) [A650/B678]. This, to be sure, is a regulative idea (an idea of Reason) and not a principle constitutive of experience. Still, as he put it:

> In all such cases reason presupposes the systematic unity of the various powers, on the ground that special natural laws fall under more general laws, and that parsimony in principles is not only an economical requirement of reason, but is one of nature's own laws. [A650/B678]

By calling "regulative" the idea that nature has an objectively valid and necessary systematic unity (cf. [A651/B679]), he wanted to stress that it is indemonstrable. Yet, without it, Kant thought, there would be no criterion of empirical truth. Besides, it can be confirmed in view of their empirical success in science (cf. [A661/B689]). Then, unification of all phenomena under universal laws of nature emerges as both the ultimate goal of the explanation of natural phenonema and as the criterion for truth. Besides, for Kant, unification confers necessity on certain principles, thereby rendering laws of nature (cf. [Kitcher, 1986]).

Though philosophically impeccable, Kant's architectonic suffered severe blows in the nineteenth and the early twentieth centuries. The blows came, by and large, by science itself. The crisis of the Newtonian mechanics and the emergence of the special and the general theories of relativity, the emergence of non-Euclidean geometries and their application to physics, Gottlob Frege's claim that arithmetic, far from being synthetic a priori, was a body of analytic truths and David Hilbert's arithmetisation of geometry which proved that no intuition was necessary created an explosive mixture that, in the end of a long process, led to the collapse of the Kantian synthetic a priori principles.

7 MILL: EXPLANATION AS UNIFICATION

In his monumental *A System of Logic Ratiocinative and Inductive* (1843), Mill defended the Regularity View of Causation, with the sophisticated addition that in claiming that an effect invariably follows from the cause, the cause should not be taken to be a single factor, but rather the whole conjunction of the conditions that are sufficient and necessary for the effect. For Mill, regular association (or invariable succession) is not sufficient for causation. An invariable succession of events is causal only if it is "unconditional", that is only if its occurrence is *not* contingent on the presence of further factors which are such that, given their presence, the effect would occur even if its putative cause was not present. A clear case in which unconditionality fails is when the events that are invariably conjoined are, in fact, effects of a common cause.

The problem that Mill faced was that there are regularities that are not causal and do not constitute laws. For instance, as Thomas Reid noted, the night always follows the day, but it is not caused by the day. They are both caused by the rotation of the earth around the sun. A similar problem arose in connection with Kant's theory of causation. Arthur Schopenhauer charged Kant with showing the absurd result that all sequence is consequence. As he noted, the tones of a musical composition follow each other in a certain objective order and yet it would be absurd to say that they follow each other according to the law of causation. The problem was that both Hume and Kant seemed to have ended up with a *loose* notion of causation. It was in order to strengthen the concept of causation that Mill introduced the idea of unconditionality. Ultimately, Mill took to be causal those invariable successions that are unconditional. It is these regularities that constitute laws of nature.

Considering how to answer the central problem of "how to ascertain the laws of nature", Mill [1843, 207] noted:

> According to one mode of expression, the question, What are the laws of nature? may be stated thus: What are the fewest and simplest assumptions, which being granted, the whole existing order of nature would result? Another mode of stating it would be thus: What are the fewest general propositions from which all the uniformities which exist in the universe might be deductively inferred?

Mill [1843, 208] was adamant that he was defending a view of laws as regularities:

> for the expression, Laws of Nature, *means* nothing but the uniformities which exist among natural phenomena [...] when reduced to their simpler expression.

Mill's breakthrough (prefigured by Kant, as we have seen) was that the issue of characterising what the laws of nature are cannot be dealt by looking at individual regularities and by trying to identify when an individual regularity is a law. Rather, it should be dealt with by looking at how the laws form a "web composed of individual threads" (ibid.). "The study of nature", Mill suggested, "is the study of laws, not *a* law; of uniformities in the plural number" (ibid.)

Borrowing Mill's expression, we may call his view 'the web of laws' approach.

What Mill perceived was that there could be no adequate characterisation of the distinction between laws of nature and merely accidentally true generalisations, unless we adopted a holistic view of lawhood. Laws are those *regularities* which are members of a coherent system of regularities, in particular, a system which can be represented as a deductive axiomatic system striking a good balance between *simplicity* and *strength*. As we shall see in section 12, Mill's approach resurfaced in the twentieth century in the writings of Frank Ramsey and David Lewis. But we have already seen that it is was a common approach in the age that Mill wrote. Leibniz and Kant held versions of it. Mill's radical twist was that he did not thereby thought that laws are rendered necessary. Nor that, at bottom, laws are anything other than regularities.

Mill was a thoroughgoing inductivist, who took all knowledge to arise from experience through induction. He even held that the law of universal causation, viz., that for every event there is a set of circumstances upon which it is invariably and unconditionally consequent, is an inductively established — and true — principle. Hence, Mill denied that there could be any certain and necessary knowledge.

He should also be credited with the first attempt to articulate the *deductive-nomological model* of explanation, which became prominent in the twentieth century. As he put it:

> An individual fact is said to be explained by pointing out its cause, that is, by stating the law or laws of causation of which its production is an instance. [1843, 305]

Similarly,

> a law of uniformity in nature is said to be explained when another law or laws are pointed out, of which that law is but a case, and from which it could be deduced. (ibid.)

The explanatory pattern that Mill identified is deductive, since the explananda (be they individual events or regularities) must be deduced from the *explanans*. And it is nomological, since the *explanans* must include reference to laws of nature. Mill went on to distinguish three patterns within this broad deductive-nomological framework. *First*, an explanation consists in the isolation of the several laws that contribute to the production of a complex effect; more accurately, that the law governing a certain effect is explained by being analysed into separate laws that govern its causes. We can call this the *de-compositional pattern* of scientific explanation. Mill frames this pattern of explanation is terms of tendencies. He notes:

> The first mode, then, of explanation of Laws of Causation, is when the law of an effect is resolved into the various tendencies of which it is the result, together with the laws of those tendencies. [1843, 306]

It may then appear that Mill offers a dispositional account of de-compositional explanation. But this is misleading. Mill introduced tendencies to save the universality of laws. He observed that "[a]ll laws of causation are liable to be (. . .) counteracted, and seemingly frustrated, by coming into conflict with other laws, the separate results of which is opposite to theirs, or more or less inconsistent with it" [1911, 292]. He then restored the universality of laws by claiming that laws describe tendencies: "All laws of causation, in consequence of their liability to be counteracted, require to be stated in words affirmative of tendencies only, and not of actual results" [1843, 293]. So, "all heavy bodies *tend* to fall; and to this there is no exception (. . .)" [1843, 294]. But Millian tendencies are occurrent qualities.[5] These tendencies are present and manifested even when the laws that govern them are counteracted by other laws. So Mill spoke as if the full effects of two separate causes actually occur and are fused into the resultant.

The *second* explanatory pattern consists in finding the complete causal history of the *explanandum*, as when intermediary causal links between the cause and the effect are found out. Mill thought that this pattern amounts to resolving the law that connects the distal cause and the effect into laws that connect the distal causes with the proximate ones and the proximate causes with the effect. It might be plausibly claimed that this second pattern amounts to *mechanistic* explanation, viz., explanation in terms of the mechanisms through which a cause brings about the effect.

The third pattern of explanation is *unification*:

> The subsumption (. . .) of one law under another, or (what comes to the same thing) the gathering up of several laws into one more general law which includes them all. [1843, 309]

Unification, according to Mill, is the hallmark of explanation and of laws. Unification is explanatory not because it renders the explananda less mysterious than they were before they were subsumed under a law, but because it minimises the number of laws that we take as ultimately mysterious, that is, as inexplicable. This very process of unification, Mill thought, brings us nearer to solving the problem of what the laws of nature are. And, as we have seen, this is no other than the problem of

> What are the fewest general propositions from which all the uniformities existing in nature could be deduced? [1843, 311]

Unification underpins the distinction between fundamental laws of nature (the basic unifiers) and derivative laws. Derivative are those laws that can be resolved into more fundamental ones either by the first or by the second pattern of scientific explanation (cf. [1843, 343]). So de-compositional and mechanistic patterns

[5] Though Mill talks of capacities (dispositions), he takes it that they are not real things existing in objects. He takes dispositional predicates to be 'names' for the claim that objects "will act in a particular manner when certain new circumstances arise" [1843, 220-1]. Mill takes it that dispositions are reducible to the categorical properties of objects.

of explanation are means for the unification pattern, which is the ultimate pattern of explanation. Though, Mill thought, all patterns of explanation are deductive, only the first two are, strictly speaking, patterns of *causal* explanation. Unification is not causal in the sense that the more fundamental laws do *not* cause the less fundamental ones. Rather, they subsume them under them, which amounts to saying that the less fundamental laws are instances, or cases, of the most fundamental ones (cf. [1843, 311]). But, of course, the unified nomological structure of the world captures its causal structure in the sense that the causal structure of the world (viz., the structure of causal laws) is exhausted by its nomological structure.

Like Kant, then, Mill put a premium on unification. It is via unification that regularities are rendered laws of nature. It is via unification that the causal structure of the world is known. But, unlike Kant, he denied that unification confers necessity on laws of nature.

Mill is also famous for his methods by which causes can be discovered. These are known as the *Method of Agreement* and the *Method of Difference*. Briefly put, according to the first, the cause is the common factor in a number of otherwise different cases in which the effect occurs. According to the second, the cause is the factor that is different in two cases, which are similar except that in the one the effect occurs, while in the other doesn't. In effect, Mill's methods encapsulate what is going on in a controlled experiment: we find causes by creating circumstances in which the presence (or the absence) of a factor makes the only difference to the production (or the absence) of an effect. Mill, however, was adamant that his methods (and the scientific method in general) work *only if* certain metaphysical assumptions are already in place. It must be the case that: a) events have causes; b) events have a *limited* number of possible causes; c) same causes have same effects, and conversely; and d) the presence or absence of causes makes a difference to the presence or absence of their effects.

PART II: UNDERSTANDING EXPLANATION

8 THE LOGICAL POSITIVIST LEGACY

The Logical Positivists took Hume to have offered a *reductive* account of causation: one that frees talk about causation from any commitments to a necessary link between cause and effect. Within science, Carnap stressed, "causality means nothing but a functional dependency of a certain sort" [1928, 264]. The functional dependency is between two states of a system, and it can be called a "causal law" if the two states are in temporal proximity and one precedes the other in time. Schlick [1932, 516] expressed this idea succinctly by pointing out that

> the difference between a mere temporal sequence and a causal sequence is the regularity, the uniformity of the latter. If C is *regularly* followed

> by E, then C is the cause of E; if E only 'happens' to follow C now and then, the sequence is called mere chance.

Any further attempt to show that there was a necessary "tie" between two causally connected events, or a "kind of glue" that holds them together, was taken to have been proved futile by Hume, who maintained that "it was impossible to discover any 'impression' of the causal nexus" [Schlick 1932, 522]. The twist that Logical Empiricists gave to this Humean argument was based on their verifiability criterion of meaning: attributing, and looking for, a "linkage" between two events would be tantamount to "committing a kind of nonsense" since all attempts to verify it would be necessarily futile (cf. [Schlick, ibid.]).

For the Logical Positivists, the concept of causation is intimately linked with the concept of law. And the latter is connected with the concept of regular (exceptionless) succession. Given that the reducing concept of regular succession is scientifically legitimate, the reduced concept of causation becomes legitimate, too. Yet, regular succession (or correlation) does not imply causation. How could Schlick and Carnap have missed this point?

The following thought is available on their behalf. The operationalisation of the concept of causation they were after was not merely an attempt to legitimise the concept of causation. It was part and parcel of their view that science aims at *prediction*. If prediction is what *really* matters, then the fact that there can be regularities, which are not causal in the ordinary sense of the word, appears to be irrelevant. A regularity can be used to predict a future occurrence of an event irrespective of whether it is deemed to be causal or not. Correlations can serve prediction, even though they leave untouched some intuitive aspect of causation, according to which not all regularities are causal.

This idea is explicitly present in Schlick [1932]. Carnap too noted that "causal relation means predictability" [1974, 192]. But he was much more careful than Schlick in linking the notion of predictability — and hence, of causation — with the notion of the law of nature. For not all predictions are equally good. Some predictions rely on laws of nature, and hence are more reliable than others which rely on "accidental universals" [1974, 214]. For Carnap, causation is not *just* predictability. It is more akin to subsumption under a universal regularity, i.e., a law of nature. As he [1974, 204] stressed:

> When someone asserts that A caused B, he is really saying that this is a particular instance of a general law that is universal with respect to space and time. It has been observed to hold for similar pairs of events, at other times and places, so it is assumed to hold for any time and place.

It seems reasonable to argue that what Carnap was really after was the connection between causation and *explanation*. When we look for explanations, as opposed to predictions, we look for something more than regularity, and relations of causal dependence might well be what we look for. The thought suggests itself that what distinguishes between a causal regularity and a mere predictive one is their different roles in explanation. It appears, then, that the concept of explanation, and in particular of nomic explanation, is the main tool for an empiricist account of causal dependence.

9 NOMIC EXPECTABILITY

What is it to explain a singular event e, e.g., the explosion of a beer-keg in the pub's basement? The intuitive answer would be to provide the cause of this event: whatever brought about its occurrence. But is it enough to just cite another event c, e.g., the rapid increase of temperature in the basement, in order to offer an adequate explanation of e? Explanation has to do with *understanding*. An adequate explanation of event e (that is, of why e happened) should offer an adequate understanding of this happening. Just citing a cause would not offer an adequate understanding, unless it was accompanied by the citation of a law that connects the two events. According to Hempel [1965] the concept of explanation is primarily *epistemic*: to explain an event is to show how this event would have been *expected* to happen, had one taken into account the laws that govern its occurrence, as well as certain initial conditions. If one expects something to happen, then one is not surprised when it happens. Hence, an explanation amounts to the removal of the initial surprise that accompanied the occurrence of the event e. *Nomic expectability* is the slogan under which Hempel's account of explanation can be placed.

Hempel systematised a long philosophical tradition, going back at least to Mill, by explicating the concept of explanation in terms of his *Deductive-Nomological* model (henceforth, *DN*-model). A singular event e (the *explanandum*) is explained if and only if a description of e is the conclusion of a valid deductive argument, whose premises, the *explanans*, involve essentially a lawlike statement L, and a set C of initial or antecedent conditions. The occurrence of the *explanandum* is thereby subsumed under a natural law. Schematically, to offer an explanation of an event e is to construct a valid deductive argument of the following form:

(*DN*)
Antecedent/Initial Conditions $C_1, \ldots, C_i$
Lawlike Statements $L_1, \ldots, L_j$

Therefore, e event/fact to be explained (*explanandum*)

A *DN*-explanation is a special sort of a valid deductive argument — whose logical form is both transparent and objective — and conversely, the species of valid deductive arguments that can be *DN*-explanations can be readily circumscribed, given only their form: the presence of lawlike statements in the premises is the characteristic that marks off an explanation from other deductive arguments. Hempel codified all this by offering 4 conditions of adequacy for an explanation.

Conditions of adequacy:

1. The argument must be deductively valid.

2. The *explanans* must contain essentially a lawlike statement.

3. The *explanans* must have empirical content, i.e., they must be confirmable.

4. The *explanans* must be true.

The first three conditions are called “logical” by Hempel [1965, 247], because they pertain to the *form* of the explanation. The fourth condition is “empirical”. Hempel rightly thought that it was an empirical matter whether the premises of an explanation were true or false. He called a *DN*-argument that satisfies the first three conditions a “potential explanation”, viz., a valid argument such that, if it were also sound, it would explain the *explanandum.* He contrasted it with an “actual explanation”, which is a sound *DN*-argument. The fourth condition is what separates a potential from an actual explanation. The latter is the correct, or the true, explanation of an event. With the fourth condition, Hempel separated what he took to be the issue of “the logical structure of explanatory arguments” [1965, 249, note 3] from the empirical issue of what is the correct explanation of an event. But the structure of an explanatory argument cannot be purely logical. Indeed, if the issue of whether an argument was a potential explanation of an event was purely logical, it would be an *a priori* matter to decide that it was a potential explanation. But condition three shows that this cannot be a purely *a priori* decision. Without empirical information about the kinds of predicates involved in a lawlike statement, we cannot decide whether the *explanans* have empirical content.

Sometimes, the reference to laws in an explanation is elliptical and should be made explicit. Or, the relevant covering laws are too obvious to be stated. Hempel thought that a proper explanation of an event should use laws, and that unless it uses laws it is, in some sense, defective. It’s no accident that the *Deductive-Nomological* model became known as “the covering law model” of explanation. Subsumption under laws is the hallmark of Hempelian explanation. So, a lot turns on what exactly the laws of nature are and this has proved to be a very sticky issue. Without a robust distinction between laws and accidents, the *DN*-model loses most of its putative force as a correct account of explanation.

Hempel took his model to provide the correct account of *causal explanation.* As he put it: “causal explanation is a special type of deductive nomological explanation” [1965, 300]. Let us call this the *Basic Thesis* (*BT*):

> (*BT*)
>
> All causal explanations of singular events can be captured by the Deductive-Nomological model.

10 ENTER CAUSATION

It has been a standard criticism of the *Deductive-Nomological* model that, insofar as it aims to offer sufficient and necessary conditions for an argument to count as a *bona fide* explanation, it fails. There are arguments that satisfy the structure of the *DN*-model, and yet fail to be *bona fide* explanations of a certain singular event. Conversely, there are *bona fide* explanations that fail to instantiate the *DN*-model. In what follows, we shall examine the relevant counterexamples and try to see how a Hempelian can escape from them. To get a clear idea of what they try to show, let me state their intended moral in advance: the *DN*-model fails precisely because it leaves out of the explication of the concept of explanation important considerations about the role of *causation* in explanation.

The first class of counter-examples, which aim to show that the *DN*-model is insufficient as an account of explanation, are summarised by the famous flagpole-and-shadow case. Suppose that we construct a *DN*-explanation of why the shadow of a flagpole at noon has a certain length. Using the height of the pole as the initial condition, and employing the relevant nomological statements of geometrical optics (together with elementary trigonometry), we can construct a deductively valid argument with a statement of the length of the shadow as its conclusion. But as Sylvain Bromberger [1966] observed, we can reverse the order of explanation: we can 'explain' the height of the flagpole, using the very same nomological statements, but (this time) the length of the shadow as the initial condition. Surely, this is not a *bona fide* explanation of the height of the pole, although it satisfies the *DN*-model. It is not a *causal* explanation of the height of the pole: although the height of the pole is the *cause* of its shadow at noon, the shadow does not cause the flagpole to have the height it does.

This counter-example can be generalised by exploiting the functional character of some lawlike statements in science: in a functional law, we can calculate the values of each of the magnitudes involved in the equation that expresses the law by means of the others. Given some initial values for the 'known' magnitudes, we can calculate, and hence '*DN*-explain', the value of the 'unknown' magnitude. Suppose, for instance, that we want to explain the period T of a pendulum. This relates to its length $l$by the functional law: $T = 2\pi\sqrt{l/g}$. We can construct a *DN*-argument whose conclusion is some value of the period T and whose premises are the above law-statement together with some value l of the length as our antecedent condition. Suppose, instead, that we wanted to explain the length of the pendulum. We could construct a *DN* argument similar to the above, with the length l as its conclusion, using the very same law-statement but, this time, conjoined with a value of the period T as our antecedent condition. If, in the former case, it is straightforward to say that the length of the pendulum *causes* it to have a period of a certain value, in the latter case, it seems problematic to say that the period causes the pendulum to have the length it does.

Put in more abstract terms, the *Deductive-Nomological* model allows explanation to be a symmetric relation between two statements, viz., the statement that expresses the cause and the statement that expresses the effect. So given the relevant nomological statements, an effect can *DN*-explain the cause as well as conversely. If we take causation to be an asymmetric relation, the *DN*-model seems unable fully to capture the nature of causal explanation, despite Hempel's contentions to the contrary.

If we wanted to stick to the *DN*-account of explanation and its concomitant claim to cover *all* causal explanation, if, that is, we wanted to defend the *Basic Thesis* (*BT*), what sort of moves would be available?

The counterexamples we have seen so far do not contradict the *Basic Thesis.* They contradict the converse of *BT*, a thesis that might be called (+):

> (+)
>
> All Deductive-Nomological explanations of singular events are causal explanations.

But neither Hempel nor his followers endorse (+). He fully accepted the existence of non-causal *DN*-explanations of singular events (cf. [1965, 353]).[6] The counterexamples do not dispute that causal explanation is a subset of *DN*-explanation. What they claim is that the *DN*-model licenses apparently inappropriate applications of the *DN*-pattern. This claim does *not* contradict *BT*. Still, the above counterexamples do show something important, viz., that unless causal considerations are imported into *DN*-explanatory arguments, they fail to distinguish between legitimate (because causal) and illegitimate (because non-causal) explanations. The task faced by the defender of the *DN*-model is to show what could be added to a *DN*-argument to issue in legitimate (causal) explanations. Schematically put, we should look for an extra X such that *DN*-model + X = causal explanation. What could this X be?

One move, made by Hempel [1965, 352] is to take X to be supplied by the law-statements that feature in a *DN*-explanation. To this end, Hempel relied on a distinction, already drawn by Mill, between *laws of coexistence* and *laws of succession.* A *law of co-existence* is the type of law in which an equation links two or more magnitudes by showing how their values are related to one another. Laws of co-existence are *synchronic*: they make no essential reference to time (i.e., to how a system or a state evolves over time); they state how the relevant magnitudes relate to each other at any given time. The law of the pendulum, Ohm's law and the laws of ideal gases are relevant examples. A *law of succession* describes how the state of a physical system changes over time. Galileo's law and Newton's second law would be relevant examples. In general, laws of succession are described by differential equations. Given such an equation, and some initial conditions, one can calculate the values of a magnitude over time. Laws of co-existence display a kind of symmetry in the dependence of the magnitudes involved in them, but

[6] Mathematical explanation is a clear case of non-causal explanation; as is the case in which one explains why an event happened by appealing to conservation laws, or to general non-causal principles (such as Pauli's exclusion principle).

laws of succession do not. Or, at least, they are not symmetric given that *earlier* values of the magnitude determine, via the law, *later* values.

Hempel [1965, 352] argued that only laws of succession could be deemed causal. Laws of co-existence cannot. They do not display the time-asymmetry characteristic of causal laws. Note that the first type of counter-examples to the *DN*-model, where there is explanatory symmetry but causal asymmetry, involves laws of co-existence. In such cases, the explanatory order can be reversed. If these laws are *not* causal, then there is no problem: there is no causally relevant feature of these laws, which is not captured by relations of explanatory dependence. So, the extra X that should be added to a *DN*-argument in order to ensure that it is a causal explanation has to do with the asymmetric character of some laws. Only asymmetric laws are causal, and can issue in causal explanations. *DN*-explanation + asymmetric laws (of succession) = causal explanation.

There seems to be something unsatisfactory in Hempel's reply. For, the thought will be, we do make causal ascriptions, even when laws of co-existence are involved. It was, after all, the *compression* of the gas that caused its pressure to rise, even though pressure and volume are two functionally dependent variables related by a law of co-existence. This seems to be a valid objection. However, the following answer is available to someone who wants to remain Hempelian, due basically to von Wright [1973]. Strictly speaking, when laws of co-existence are referred to in a *DN* explanatory argument, the explanation can be symmetric: we can explain the values of magnitude A by reference to the values of magnitude B, and conversely. But, Hempel's defender might go on, in particular *instances* of a *DN* explanatory argument with a law of co-existence, this symmetry can be (and is) broken. How the symmetry is broken — and, hence how the direction of explanation is determined — depends on which of the functionally dependent variables is actually *manipulated*.

When laws of co-existence are involved, the symmetry that *DN* explanations display can be broken in different ways in order to capture what causes what on the particular occasion. An appeal to manipulability can also show how Bromberger-type counter-examples can be avoided. A *DN*-model which cites the length of the shadow as the explanation of the height of the flagpole should not count as a *bona fide* explanation. Although the length of the shadow and the height of the pole are functionally inter-dependent, only the height of the pole is really manipulable. One can create shadows of any desirable length by manipulating the heights of flagpoles, but the converse is absurd. Manipulability can then be seen as the sought after supplement X to the *DN*-model which determines what the causal order is across different symmetric contexts in which a *DN*-argument is employed. *DN*-explanation (with functional laws) + manipulability = causal explanation.

Yet, the concept of manipulation is causal. This means that an advocate of *DN*-explanation who summons von Wright's help can at best have a Pyrrhic victory. For she is forced to employ irreducible causal concepts in her attempt to show how a *DN*-model of explanation can accommodate the intuitive asymmetry that explanatory arguments can possess.

The popular philosophical claim that the *DN*-model leaves important causal considerations out of the picture is supported by a second class of counter-examples. These aim to show that satisfaction of the *DN*-model is not a necessary condition for *bona fide* causal explanations. In fact, these counterexamples aim directly to discredit *BT*. Remember that *BT* says, in effect, that the claim that *c* causes *e* will be elliptical, unless it is offered as an abbreviation for a full-blown *DN*-argument. This view has been challenged by Michael Scriven. He made this point by the famous example of the explanation of the ink-stain on the carpet. Citing the fact that the stain on the carpet was *caused* by inadvertently knocking over an ink-bottle from the table, Scriven [1962, 90] argues,

> is the explanation of the state of affairs in question, and there is no nonsense about it being in doubt because you cannot quote the laws that are involved, Newton's and all the others.

His point is that there can be fully legitimate *causal* explanations that are not *DN*. Instead, they are causal stories, i.e., stories that give causally relevant information about how an effect was brought about, without referring to any laws, and without having the form of a deductive argument. Collaterally, it has been a standard criticism of the Hempelian model that it wrongly makes all explanations to be arguments. A main criticism is that citing a causal mechanism can be a legitimate explanation of an event without having the form of a Hempelian deductive-nomological argument (cf. [Salmon 1989, 24].

One can accept Scriven's objection without abandoning the *Deductive-Nomological* model of explanation; nor the *Basic Thesis*. The fact that the relevant nomological connections may not be fully expressible in a way that engenders a proper deductive explanation of the *explanandum* merely shows that, on some occasions, we shall have to make do with what Hempel called "explanation sketches" instead of full explanations. Explanation-sketches can well be ordinary causal stories that, as they stand, constitute incomplete explanations of an event E. But these stories can be completed by taking account of the relevant laws that govern the occurrence of the event E. Scriven's point, however, seems to be more pressing. It is that a causal explanation can be *complete*, without referring to laws (cf. [1962, 94]). So, he directly challenges Hempel's assumption that all causal explanation *has to* be nomological. Scriven insists that explanation is related to understanding and that the latter might, but won't necessarily, involve reference to laws. He proposes [1962, 95]:

> a causal explanation of an event of type [E], in circumstances [R] is exemplified by claims of the following type: there is a comprehensible cause [C] of [E] and it is understood that [C]s can cause [E]s.

But, a Hempelian might argue, it is precisely when we move to the nomological connection between Cs and Es that we understand how Cs can cause Es.

One important implication of the *Deductive-Nomological* model is that *there is no genuine singular causal explanation* (cf. [Hempel 1965, 350 & 361-2]). Scriven's own objection can be taken to imply that a singular causal explanation of an event-token (e.g., the staining of the carpet by ink) is a complete and fully adequate explanation of its occurrence. Since the *DN*-model denies that there can be legitimate singular *causal* explanations of events, what is really at stake is whether causal stories that are not nomological can offer legitimate explanation of singular events. What, then, is at stake is the *Basic Thesis*.

Note that there is an ambiguity in the singularist approach. What does it mean to say that there is *no* nomological connection between two event-tokens c and e that are nonetheless such that c causally explains e? It might mean one of the following two things: (i) there are no relevant event-types under which event-tokens c and e fall such that they are nomologically connected to each other; (ii) even if there is a relevant law, we don't (can't) know it; nor do we have to state it explicitly in order to claim that the occurrence of event-token c causally explains the occurrence of event-token e.

The first option is vulnerable to the following objection. One reason why we are interested in identifying causal facts of the form 'c causes e' (e.g., heating a gas at constant pressure causes its expansion) is that we can then *manipulate* event-type C in order to bring about the event-type E. But the possibility of manipulation requires that *there is* a nomological connection between types C and E. It is this nomological connection that makes possible bringing about the effect e, by manipulating its causes. Hence, if causation is to have any bite, it had better instantiate laws.[7] So the singularist's assertion should be interpreted to mean the second claim above, viz., that even if there is a law connecting event-types C and E, we don't know it; nor do we have to state it explicitly in order to claim that the occurrence of event-token c causally explains the occurrence of event-token e. Given this understanding, it might seem possible to reconcile the singularist approach with a Hempelian one. This is precisely the line taken by Davidson [1967]. On his view, all causation is nomological, *but* stating the law explicitly is not required for causal explanation.

Considering this idea, Hempel noted that when the law is not explicitly offered in a causal explanation, the statement 'c causes e' is incomplete. In making such a statement, one is at least committed to the view that "*there are* certain further unspecified background conditions whose explicit mention in the given statement would yield a truly general law connecting the 'cause' and the 'effect' in question" [1965, 348]. But this purely existential claim does not amount to much. As Hempel carried on to say, the foregoing claim is comparable to having "a note saying that there is a treasure hidden somewhere" [1965, 349]. Such a note would be useless, unless "the location of the treasure is more narrowly circumscribed". So, the alleged reconciliation of the singularist approach with Hempel's will not work, unless there is an attempt to make the covering law explicit. But this will

[7] Even if one were to take the currently popular view that manipulation requires only *invariant relations* among magnitudes or variables, and even if it was admitted that these invariant relations do not hold universally but only for a certain range of interventions/manipulations, one would still be short of a genuinely singularist account of causation.

inevitably take us back to forging a close link between stating causal dependencies and stating laws.

To sum up: if Scriven's counterexample were correct, it would establish the thesis that the *Deductive-Nomological* model is not necessary for causal explanation. Let's call this thesis *UNT*. *UNT* says: if Y is a causal explanation of a singular event, then Y is not necessarily a *DN*-explanation of this event. *UNT*, if true, would contradict the *Basic Thesis.* But we haven't yet found good reasons to accept *UNT.*

11 CAUSAL HISTORIES

In his [1986], Lewis takes causal explanation of a singular event to consist in providing some information about its causal history. In most typical cases, it is hard to say of an effect e that its cause was *the* event c. Lots of things contribute to bringing about a certain effect. All these factors, Lewis says, comprise the *causal history* of the effect. This history is a huge causal net in which the effect is located. To explain why this event happened, we need to offer some information about this causal net. This is "explanatory information" [1986, 185]. A *full* explanation consists in offering the whole causal net. But hardly ever this full explanation is possible. Nor, Lewis thinks, is it necessary. Often, some chunk of the net will be enough to offer an adequate causal explanation of why a certain singular event took place.

Lewis [1986, 221-4] thinks there is no such thing as non-causal explanation of singular events. That is, he endorses the following thesis:

> (*CE*)
>
> All explanation of singular events is causal explanation.

Recall that the *Basic Thesis* (*BT*) says:

> All causal explanation can be captured by the Deductive-Nomological model.

If we added *BT* to *CE,* then it would follow that

> (*CE**)
>
> All explanation of singular events can be captured by the Deductive-Nomological model.

Could a Lewisian accept *BT*, and hence also accept *CE**? That is, does Lewis's account of causal explanation violate the *Basic Thesis*? Or, is his view of causal-explanation-as-information-about-causal-histories compatible with *BT*? Lewis [1986, 235-6] asks:

> is it [...] true that any causal history can be characterised completely by means of the information that can be built into *DN* arguments?

Obviously, if the answer is positive, *BT* is safe. Lewis expresses some scepticism about a fully positive answer to the above question. He thinks that if his theory of causation, based on the notion of counterfactual dependence, is right, there can be genuinely singular causal explanation. Yet, he stresses that in light of the fact that the actual world seems to be governed by a "powerful system of (strict or probabilistic) laws, [...] the whole of a causal history could in principle be mapped by means of *DN*-arguments [...] of the explanatory sort" [1986, 236]. He adds:

> [...] if explanatory information is information about causal histories, as I say it is, then one way to provide it is by means of *DN* arguments. Moreover, under the hypothesis just advanced, [i.e., the hypothesis that the actual world is governed by a powerful system of laws], there is no explanatory information that could not in principle be provided in that way. To that extent the covering-law model is dead right. [ibid.]

So, the *Basic Thesis* is safe for a Lewisian, at least if it is considered as a thesis about causal explanation in the actual world. Then, what is Lewis's disagreement with the *DN*-model? There is a point of principle and a point of detail. The point of principle is this. The *Basic Thesis* has not been discredited. But, if I understand Lewis correctly, he thinks that it has been wounded. It may well be the case that if Y is a causal explanation of a singular event, then Y is also a *DN*-explanation of this event. Lewis does not deny this (cf. [1986, 239-40]). But, in light of the first set of counterexamples above, *BT* might have to be modified to *BT′*:

> (*BT′*)
>
> All causal explanation of singular events can be captured by *suitable instances* of the Deductive-Nomological model.

The modification is important. For it may well be the case that what instances of the *DN*-model are *suitable* to capture causal explanations might well be specifiable only "by means of explicitly causal constraints" [1986, 236]. And if this is so, then the empiricists' aspiration to capture causal concepts by the supposedly unproblematic explanatory concepts seems seriously impaired.

The point of detail is this. Take *BT′* to be unproblematic. It is still the case, Lewis argues, that the *Deductive-Nomological* model has wrongly searched for a "unit of explanation" [1986, 238]. But there is no such unit:

> It's not that explanations are things we may or may not have one of; rather, explanation is something we may have more or less of". [ibid.]

Although Lewis agrees that a full *DN*-explanation of an individual event's causal history is both possible and most complete, he argues that this ideal is chimerical. It is the "ideal serving of explanatory information" [1986, 236]. But, "other shapes and sizes of partial servings may be very much better — and perhaps also better within our reach" [1986, 238]. This is something that the advocate of the *DN*-model need *not* deny.

12 (A BRIEF NOTE ON) LAWS OF NATURE

The *Deductive-Nomological* model of explanation, as well as any attempt to tie causation to laws, faced a rather central conceptual difficulty: the problem of how to characterise the laws of nature. Most Humean-empiricists adopted the *Regularity View of Laws*: laws of nature are regularities. Yet, they have had a hurdle to jump: not all regularities are causal. Nor can all regularities be deemed laws of nature. So they were forced to draw a distinction between the good regularities (those that constitute the *laws of nature*) and the bad ones i.e., those that are, as John Stuart Mill put it, "conjunctions in some sense accidental". Only the former can underpin causation and play a role in explanation. The predicament that Humeans were caught in is this. Something (let's call it the property of lawlikeness) must be added to a regularity to make it a law of nature. But what can this be?

The first systematic attempt to characterise this elusive property of lawlikeness was broadly epistemic. The thought, advanced by A. J. Ayer, Richard Braithwaite and Nelson Goodman among others, was that inquirers have different *epistemic attitudes* towards laws and accidents. Lawlikeness was taken to be the property of those generalisations that play a certain *epistemic* role: they are believed to be true, and they are so believed because they are confirmed by their instances and are used in proper inductive reasoning. In a sense, 'natural law' was taken to be an honorific title that should be given to those regularities that are believed to hold on account of diverse evidence in their favour. But this purely epistemic account of lawlikeness fails to draw a robust line between laws and accidents. Couched in terms of belief, or in terms of a psychological willingness or unwillingness to extend the generalisation to unknown cases, the supposed difference between laws and accidents becomes spurious.

A much more promising attempt to characterise the property of lawlikeness is what may be called the *web of laws* view. According to this view, the regularities that constitute the laws of nature are those that are expressed by the axioms and theorems of an ideal deductive system of our knowledge of the world, and in particular, of a deductive system that strikes the *best* balance between simplicity and strength. Simplicity is required because it disallows extraneous elements from the system of laws. Strength is required because the deductive system should be as informative as possible about the laws that hold in the world. Whatever regularity is not part of this *best system* it is merely accidental: it fails to be a genuine law of nature. The gist of this approach, which, as we have seen, has been advocated by Mill, and in the twentieth century by Ramsey [1928] and Lewis [1973], is that no regularity, taken in isolation, can be deemed a law of nature. The regularities that constitute laws of nature are determined in a kind of holistic fashion by being parts of a structure.

The Mill-Ramsey-Lewis view has many attractions. It solves the problem of how to distinguish between laws and accidents. It shows, in a non-circular way, how laws can support counterfactuals. For, it identifies laws *independently* of their ability to support counterfactuals. It makes clear the difference between regarding a statement as lawlike and being lawlike. It respects the major empiricist thesis that laws of nature are contingent. For a regularity might be a law in the actual world without being a law in other possible worlds, since in these possible worlds it might not be part of the best system for these worlds. It solves the problem of uninstantiated laws. The latter might be taken to be proper laws insofar as their addition to the best system results in the enhancement of the strength of the best system, without detracting from its simplicity.

Yet, this view faces the charge that it cannot offer a fully *objective* account of laws of nature. For instance, it is commonly argued that how our knowledge of the world is organised into a simple and strong deductive system is, by and large, a subjective matter. Hence, what regularities will be deemed *laws* seems to be based on our subjective attitude towards regularities. But this kind of criticism is overstated. There is nothing in the *web-of-laws* approach that makes laws mind-dependent. The regularities that are laws are fully objective, and govern the world irrespective of our knowledge of them, and of our being able to identify them. In any case, as Ramsey, in effect, pointed out, it is a fact about the world that some regularities form, objectively, a system; that is, that *the world has an objective nomological structure*, in which regularities stand in certain relations to each other; relations that can be captured (or expressed) by relations of deductive entailment in an ideal deductive system of our knowledge of the world. Ramsey's suggestion grounds an objective distinction between laws and accidents in a *worldly* feature: that the world has a certain nomological structure.

In the 1970s, David Armstrong [1983], Fred Dretske [1977] and Michael Tooley [1977] put forward the view that lawhood cannot be reduced to regularity (not even to regularity-plus-something-that-distinguishes-between-laws-and-accidents). Lawhood, they claimed, is a certain contingent necessitating relation among properties (*universals*). Accordingly, it is a law that all Fs are Gs if and only if there is a relation of nomic necessitation $N(F, G)$ between the properties (universals) F-ness and G-ness such that all Fs are Gs. This approach has many attractions. It purports to explain why there are regularities in the world: because there are necessitating relations among properties. It thereby distinguishes between regularities and laws: the regularities that hold in the world do not constitute the laws that hold in the world. Rather, and at best, they are the *symptoms* of the instantiation of laws. It explains the difference between nomic regularities and accidents by claiming that the accidental regularities are not even symptoms of the instantiation of laws. It makes clear how laws can *cause* anything to happen:

they do so because they embody causal relations among properties. But the central concept of nomic necessitation is still not sufficiently clear. In particular, it is not clear how the necessitating relation between the property of *F-ness* and the property of *G-ness* makes it the case that *All Fs are Gs.* To say that there is a necessitating relation $N(F, G)$ is not yet to explain what this relation is. Nor does it say anything about how the corresponding regularity *All Fs are Gs* obtains. It might seem that $N(F,G)$ *entails* the corresponding regularity *All Fs are Gs*; but it is not clear at all how this entailment goes. If the regularity *All Fs are Gs* is contained in $N(F,G)$ as the sentence 'P' is contained in the sentence '$P\&Q$', the entailment is obvious. But then, there seems to be a mysterious extra 'Q' in $N(F,G)$ over the 'P' (= *All Fs are Gs*). And we are in the dark as to what this might be, and how it ensures that the regularity obtains.

Both the Humeans and the advocates of the Armstrong-Dretske-Tooley view agree that laws of nature are *contingent.* A growing rival thought has been that if laws did not hold with some kind of objective necessity, they could not be robust enough to support either causation or explanation. As a result of this, laws of nature are said to be metaphysically necessary. This amounts to a radical denial of the contingency of laws. Along with it came a resurgence of Aristotelianism in the philosophy of science. The advocates of the view the laws are *contingent* necessitating relations among properties took it to be the case that though an appeal to (natural) properties was indispensable for the explication of lawhood, the properties themselves are passive and freely recombinable. Consequently, there can be a possible world in which some properties are not related in the way they are related in the actual world. The advocates of metaphysical necessity took the stronger line that laws of nature flow from the essences of properties. In so far as properties have essences, and in so far as it is part of their essence to endow their bearers with a certain behaviour, it follows that the bearers of properties *must* obey certain laws, those that are issued by their properties. Essentialism was treated with suspicion in most of the twentieth century, partly because essences were taken to be discredited by the advent of modern science and partly because the admission of essences (and the concomitant distinction between essential and accidental properties) created logical difficulties. Essentialism required the existence of *de re* necessity, that is natural necessity, since if it is of the essence of an entity to be thus-and-so, it is *necessarily* thus-and-so. But before Kripke's [1972] work, the dominant view was that all necessity was *de dicto*, that is, it applies, if at all, to propositions and not to things in the world.

The thought that laws are metaphysically necessary gained support from the (neo-Aristotelian) claim that properties are active powers. The key idea here was introduced by Rom Harré and Edward H Madden [1975] and strengthened by Sidney Shoemaker [1980]. They argued that properties are best understood as powers since the only way to identify them is via their causal role. Two seemingly distinct properties that have exactly the same powers are, in fact, one and the same property. Similarly, one cannot ascribe different powers to a property without changing this property. It's a short step from these claims that properties are not freely recombinable: there cannot be worlds in which two properties are combined by a different law than the one that unites them in the actual world. On this view, it does not even make sense to say that properties are united by laws. Rather, properties — qua powers — *ground* the laws.

Many philosophers remain unconvinced. A popular claim is that the positing of irreducible powers is, in Mackie's [1977, 366] memorable phrase, the product of metaphysical double-vision. Far from explaining the causal character of certain processes (e.g., the dissolution of a sugar-cube in water), "they just *are* the causal processes which they are supposed to explain seen over again as somehow latent in the things that enter into these processes" [ibid.].[8]

13 UNIFICATION REVISITED

Scientific explanation is centrally concerned with explaining regularities — perhaps more centrally than with explaining particular facts. But when Hempel attempted to extend his *Deductive-Nomological* model to the explanation of laws, he encountered the following difficulty (cf. [1965, 273]). Suppose one wants to explain a low-level law L_1 in a DN-fashion. One can achieve this by simply subsuming L_1 under the more comprehensive regularity $L_1\&L_2$, where L_2 may be any other law one likes. For instance, one can DN-explain Boyle's law by deriving it from the conjunction of Boyle's law with the law of Adiabatic Change. Although such a construction would meet all the requirements of the DN-model, it wouldn't count as an explanation of Boyle's law. Saying that the conjunction $L_1\&L_2$ is not more fundamental than L_1 would not help. The issue at stake is precisely what makes a law more *fundamental* than another one. Intuitively, it is clear that the laws of the kinetic theory of gases are more fundamental than the laws of ideal gases. But if what makes them more fundamental *just* is that the latter are derived from the former, the conjunction $L_1\&L_2$ would also count as more fundamental than its components. Hempel admitted that he did not know how to deal with this difficulty. But this difficulty is very central to his project. The counter-example trivialises the idea that laws can be DN-explained by being deduced from other laws. Hence, the empiricist project should have to deal with 'the problem of conjunction'.

13.1 Reducing the Number of Brute Regularities

An intuitive idea is that a law is more fundamental than others, if it *unifies* them. But how exactly is *unification* to be understood? According to Friedman [1974], explanation is closely linked with understanding. Now, 'understanding' is a slippery notion. It relates, intuitively, to knowing the causes: how the phenomena are brought about. Friedman revived a long-standing empiricist tradition where

[8] For more on the issue of laws of nature, see my [2002, part II].

'understanding' is linked to conceptual economy.[9] The basic thought is that a phenomenon is understood, if it is made to fit within a coherent whole, which is constituted by some basic principles. If a number of seemingly independent regularities are shown to be subsumable under a more comprehensive law, then, the thought is, our understanding of nature is promoted. For, the number of regularities which have to be assumed as 'brute' is minimised. Some regularities, the fundamental ones, should still be accepted as brute. But the smaller the number of regularities that are accepted as brute, and the larger the number of regularities subsumed under them, the more we comprehend the workings of nature: not just what regularities there are, but also why they are and how they are linked to each other. After noting that in important cases of scientific explanation (e.g., the explanation of the laws of ideal gases by the kinetic theory of gases) "we have reduced a multiplicity of unexplained, independent phenomena to one", Friedman [1974, 15] added:

> I claim that this is the crucial property of scientific theories we are looking for; this is the essence of scientific explanation — sciences increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given. A world with fewer independent phenomena is, other things equal, more comprehensible than with more.

Explanation, then, proceeds via unification into a compact theoretical scheme. The basic 'unifiers' are the most fundamental laws of nature. The explanatory relation is still deductive entailment, but the hope is that, suitably supplemented with the idea of 'minimising the number of independently acceptable regularities', it will be able to deal with the conjunction problem.

In outline, Friedman's approach is the following. A lawlike sentence L_1 is acceptable independently of lawlike sentence L_2, if there are sufficient grounds for accepting L_1, which are not sufficient grounds for accepting L_2. This notion of 'sufficient grounds' is not entirely fixed. Friedman [1974, 16] states two conditions that it should satisfy:

i If L_1 implies L_2, then L_1 is not acceptable independently of L_2.

ii If L_1 is acceptable independently of L_2, and L_3 implies L_2, then L_1 is acceptable independently of L_3.

The basic idea is that lawlike sentence L_1 is not acceptable independently of its logical consequences, but it is independently acceptable of other statements logically independent from it. This is not very illuminating, as Friedman admits. But a further step shows how this idea can be put to work in solving 'the problem of conjunction'. Take a lawlike sentence L. Let us call a *partition* of L a set of sentences L_1,..., L_n such that

[9] This tradition goes back to Mach and Poincaré, but Friedman wants to dissociate the idea of unification from Mach's and Poincaré's phenomenalist or instrumentalist accounts of knowledge.

a their conjunction is logically equivalent to L; and

b each member L_i of the set is acceptable independently of L.

Let us call 'conjunctive' a sentence L which satisfies (a) and (b), and, following Friedman, let us call "atomic" a sentence L which *violates* them. Given this, a lawlike sentence L explains lawlike sentences $L_1,\ldots, L_n$, if L is "atomic". Conversely, a lawlike sentence L *fails* to explain lawlike sentences $L_1,\ldots, L_n$, if L is 'conjunctive'. We can now see how Friedman's account bars the mere conjunction L_1 &L_2 of Boyle's law (L_1) with the law of Adiabatic Change (L_2) from explaining Boyle's law: the conjunction of the two laws is not an atomic sentence; it is a 'conjunctive' sentence. It is partitioned into a (logically equivalent) set of independently acceptable sentences, viz. L_1 and L_2. Conversely, we can see why Newton's law of gravity offers a genuine explanation, via unification, of Galileo's law, Kepler's laws, the laws of the tides, etc. On Friedman's account, the difference between Newton's law and the mere conjunction L_1&L_2 is that the content of Newton's law *cannot* be partitioned into a (logically equivalent) set of independently acceptable laws: the sentence which express Newton's law is "atomic".

As Kitcher [1976] has shown, Friedman's account does *not* offer a necessary condition for the explanation-as-unification thesis. His general point is that if, ultimately, explanation of laws amounts to derivation of lawlike statements from other lawlike statements, then in mathematical physics at least, there will be many such derivations that utilise more than one lawlike statement as premises. Hence, ultimately, there are 'conjunctions' that are partitioned into independently acceptable lawlike statements which, nonetheless, explain other lawlike statements.

On the face of it, however, atomicity does offer a *sufficient* condition for genuine unifying, and hence explanatory, power. But can there be atomic lawlike sentences? At a purely syntactic level, there cannot be. *Any* sentence of the form 'All Fs are Gs' can be partitioned into a logically equivalent set of sentences such as Àll (Fs & Hs) are Gs' and 'All (Fs & *not-H*s) are Gs'}. So the predicate 'is a planet' can be partitioned into a set of logically equivalent predicates: 'is a planet and is between the earth and the sun' (F & H) and 'is a planet and is not between the earth and the sun' (F & *not-H*). Take, then, the statement that expresses Kepler's first law, viz., that all planets move in ellipses. It follows that this can be partitioned into two statements: 'All *inferior planets* move in ellipses'; and 'All *superior planets* move in ellipses'. A perfectly legitimate lawlike statement is partitioned into two other lawlike statements. Is it then "atomic"? Syntactic considerations alone suggest that it is not.

The advocates of "atomicity" might insist that not all syntactic partitions of a lawlike statement will undermine its atomicity, since not all syntactic partitions will correspond to 'natural kind' predicates. The thought might be that whereas F and G in 'All Fs are Gs' are 'natural kind' predicates, the predicates, F&H and F&*not-H*, which can be used to form the logically equivalent partition {All (Fs &Hs) are Gs; All (Fs & *not-H*s) are Gs}, are not necessarily 'natural kind' predicates. This admission reveals an important weakness of Friedman's approach.

In order to be viable, this approach requires a theory of what predicates pick out natural kinds. This cannot be a purely syntactic matter. One standard thought has been that the predicates that pick out natural kinds are the predicates that are constituents of genuine lawlike statements. But on Friedman's approach it seems that this thought would lead to circularity. In order to say what statements are genuinely atomic, and hence what statements express explanatory laws, we first need to show what syntactically possible partitions are *not* acceptable. If we do that by means of a theory of what predicates pick out natural kinds, we cannot, on pain of circularity, say that those predicates pick out natural kinds that are constituents of statements which express explanatory laws. This last objection, however, may not be as fatal as it first sounds. The genuine link that there is between delineating what laws of nature are and what the 'natural-kind' predicates are has led many philosophers to think that the two issues can only be sorted out together. The concept of a law of nature and the concept of a natural-kind predicate form a family: one cannot be delineated without the other.

The basic flaw in Friedman's approach is the following. He defines unification in a syntactic fashion. In this sense, he's very close to the original Hempelian attempt to characterise 'explanation' in a syntactic manner. Hempel run into the problem of how to distinguish between genuine laws and merely accidentally true generalisations. Purely syntactic considerations could not underwrite this distinction. Friedman attempted to solve this problem by appealing to unification. But the old problem re-appears in a new guise. Now it is the problem of how to distinguish between 'good' unifiers (such as Newton's laws) and 'bad' unifiers (such as mere conjunctions). A purely syntactic characterisation is doomed to fail, no less than it failed as a solution to Hempel's original problem.

13.2 Unified Explanatory Store

The failures of Friedman's approach to unification led Kitcher [1981] to advance an alternative view, which changes substantially the characterisation of unification. He calls us to envisage a set K of statements accepted by the scientific community. K is consistent and deductively closed. An "explanatory store $E(K)$" over K is "the best systematisation of K" [1981, 337]. The best systematisation, however, is not what Friedman took it to be. It is not couched in terms of the minimal set of lawlike statements that need to be assumed in order for the rest of the statements in K to follow from them. For Kitcher, the best systematisation is still couched in terms of the derivation of statements of K that best unifies K, but the unification of K is not taken to be a function of the size (cardinality) of its set of axioms. Rather, Kitcher takes unification to be a function of the number of explanatory patterns, or schemata, that are necessary to account for the statements of K. The smaller this number is, the more unified is $E(K)$. Given a small number of explanatory patterns, it may turn out that the number of facts that need to be accepted as brute in the derivations of statements of K might be small too. So, it may be that Kitcher's unification entails (the thrust of) Friedman's unification.

But it is important to stress that what bears the burden of unification for Kitcher is the explanatory pattern (schema). As he [1989, 432] put it:

> Science advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same pattern of derivation again and again, and in demonstrating this, it teaches us how to reduce the number of types of fact that we accept as ultimate.

Before we analyse further Kitcher's central idea, we need to understand his notion of an explanatory schema (or pattern). To fix our ideas, let us use an example (cf. [Kitcher 1989, 445-7]). Take one of the fundamental issues in the post-Daltonian chemistry, viz., the explanation of the fact that the compounds of X and Y always contains X and Y in the weight ratio $m : n$. Kitcher suggests that Dalton's approach can be seen as involving the following explanatory schema.

1. The compound Z between X and Y has an atomic formula of the form: $XpYq$.

2. The atomic weight of X is x and the atomic weight of Y is y.

3. The weight ratio of X to Y is $px : qy (= m : n)$.

This schema can be repeatedly (and successfully) applied to many cases of compounds. Take Z to be water. So, (1), X=H(yrdogen) and Y=O(xygen) and Z is H_2O_1. Then, (2) x=1 and y=16. Then, (3), 2X1:1X16=2:16=1:8 (=$m : n$). The structure of this explanatory schema (general argument-pattern) is an ordered triple: <schematic argument, filling instructions, classification>.

- The *schematic argument* is (1) to (3) above. It is schematic because it consists of schematic sentences. These are sentences in which some nonlogical expressions occurring in them (e.g., names of chemical elements) are replaced by dummy letters, (e.g., Z, X, Y) which can take several values.

- The *filling instructions* are directions for replacing the dummy letters of the schematic sentences with their appropriate values. In the example at hand, the dummy letters X and Y should be replaced by names of elements (e.g., Hydrogen and Oxygen), the dummy letters p and q should take natural numbers as values, and the dummy letters x, y should take real numbers are values.

- The *classification* is a set of statements that describe the inferential structure of the schema. In the case at hand, the classification dictates that (1) and (2) are the premises of the argument while (3) is the conclusion.

Explanatory schemata are the vehicles of explanation. The explanatory store E(K) is "a reserve of explanatory arguments" [1981, 332], whose repeated applications to many phenomena brings order — and hence unifies — K.

A central thought in Kitcher's account is that explanations are arguments, and in particular *deductive* arguments. The best systematisation is still a deductive systematisation, even if what effects the systematisation is the number of deductive patterns that are admissible, and not the number of axioms of the 'best system'. In this sense, Kitcher's approach is a descendant of Hempel's *Deductive-Nomological* model. It shares some of its most important features and consequences. The relation of explanatory dependence is a relation between sentences and it should be such that it instantiates a deductively valid argument with (a description of) the *explanandum* as its conclusion. Yet, we need to be careful here. Kitcher's account, as it now stands, does *not* demand that the premises of explanatory arguments be laws of nature. It does not even demand that they be universally quantified statements. They may be, and yet they may not. So, as it stands, Kitcher's account need not be a way to explicate what the laws of nature are. Nor does it demand that all explanation be *nomological*.

However, it seems that statements that express genuine laws of nature are uniquely apt to do the job that Kitcher demands of explanation. By being genuinely lawlike, these statements can underwrite the power that some schemata have to be repeatedly employed in explanations of singular events. Take the case, discussed also by Hempel, of trying to explain why John Jones is bald. Hempel rightly thought it inadmissible to explain this fact by constructing a *DN*-argument whose premises are the following: "John Jones is a member of the Greenbury School Board for 1964" and "All members of the Greenbury School Board for 1964 are bald". His reason was that the statement "All members of the Greenbury School Board for 1964 are bald" did not express a genuine law. Kitcher agrees with Hempel that this explanation is inadmissible: it rests on an accidentally true generalisation. But how is he to draw the distinction between laws and accidents within his own account? He says that an argument-pattern that aims to explain why certain individuals are bald by employing the sentence "All members of the Greenbury School Board for 1964 are bald" is not "generally applicable" [1981, 341]. On the contrary, an argument-pattern that would aim to explain why certain individuals are bald by reference to some principles of physiology would be generally applicable.

What, however, underwrites the difference in the applicability of argument-patterns such as the above is that the former rests on an accidental generalisation while the latter rests on genuine laws. It's not just that "All members of the Greenbury School Board for 1964 are bald" has a finite number of instances — a fact that would impair its applicability. Kepler's first law has only a finite number of instances, and yet we think that its presence in an argument-pattern would not impair its applicability. So, Kitcher needs to tie the explanatory applicability of an argument-pattern with the presence of genuine lawlike statements in it.

Is Kitcher's account of unification in terms of argument-patterns satisfactory? The notion of an argument-pattern is clear enough and does seem to capture some sense in which a system is unified. But when argument-patterns are applied to several cases, things seem to be more complicated than Kitcher thinks. Take one of his own examples: Newton's second law of motion. Once we are clear on the notion of 'force', Newton's law **F=ma** can be seen as specifying a Kitcher-like argument-pattern. The whole problem, however, is that none of the elements of the triple that specify an argument-pattern, viz., schematic argument, filling instructions, classification, can capture the all-important concept of a force-function. Each specific application of Newton's law requires, as Cartwright has repeatedly stressed, the prior specification of a suitable force-function. So, when we deal with a pendulum, we need to introduce a different force-function (e.g., F=-Kx) than when we are faced with a planet revolving around the sun. It's not part of the schematic argument what force-functions are applicable. Nor can this be added to the filling instructions, simply because the force-functions may be too diverse, or hitherto unspecified.

There is clearly something to the idea that, given a repertoire of force-functions, Newton's second law can be schematised à la Kitcher. But part of explaining a singular event is surely to figure out *what* force-function applies to this particular case. Besides, even when we have chosen the relevant force-function, we need to introduce further assumptions, related to the specific domain of application, which will typically rest on idealisations. All these cannot be part of the explanatory pattern. What really seems to matter in most (if not all) cases is that the phenomena to be explained are traced back to some kind of basic law, such as **F=ma**. It's not so much that we can repeatedly apply a certain argument-pattern to derive more specific cases. Instead, more typically, we show how specific cases can be reduced to being instances of some basic principles. That these basic principles will be applicable to many phenomena follows from their universal character. But it seems irrelevant whether or not the repertoire of the arguments from which (descriptions of) several phenomena derive is small or large. Unification consists in minimising the number of *types* of general principles, which are enough to account for the phenomena. Admittedly, this view is closer to Friedman's than to Kitcher's. But so be it.

14 MECHANISTIC EXPLANATION REVISITED

The thought that explanation amounts to identification of causal mechanisms reappeared in the work of Wesley Salmon. He distinguished between three approaches to scientific explanation, which he called the "epistemic conception", the "modal conception" and the "ontic conception".

The *epistemic conception* is the Hempelian approach. It makes the concept of explanation broadly epistemic, since, as we have seen, it takes explanation to be nomic expectability. The *modal conception* differs from the epistemic mostly in its account of necessity. The *explanandum* is said to follow necessarily from the *explanans*, in the sense that it was *not* possible for it not to occur, given the relevant laws. The *ontic conception* takes explanation to be intimately linked to *causation.* As Salmon [1984, 19] put it:

> To give scientific explanations is to show how events [. . .] fit into the causal structure of the world.

Salmon takes the world to have an already built-in causal structure. Explanation is then seen as the process in virtue of which the *explananda* are placed in their right position within this causal structure. On Salmon's ontic conception, causal relations are *prior* to relations of explanatory dependence. What explains what is parasitic on (or determined by) what causes what.

Kitcher [1985, 638] has rightly called Salmon's approach "bottom-up": we first discern causal relations among particular events, and then conceive of the task of explanation as identifying the causal mechanisms that produce the events for which we seek an explanation. To this approach, Kitcher contrasts a "top-down" one: we begin with a unified deductive systematisation of our beliefs; then, we proceed to make ascriptions of causal dependencies (i.e., of relations of cause and effect), which are parasitic on (or determined by) the relations of explanatory dependence that emerge within the best unified system.

A central motivation for Salmon's ontic conception is that causal explanation cannot be captured by the derivationist (top-down) models that have been very popular in the history of thinking about explanation. The commitment to explanation as a species of deductive derivation has been so pervasive that it can hardly be exaggerated. For Salmon, explanation is not a species of *deductive derivation*: explanations are not arguments. It is noteworthy that Salmon is as willing as anyone to adopt unification as the goal of scientific explanation. He [1985, 651] takes it that an advocate of a "mechanistic" view of explanation, (viz., of the view that explanation amounts to the identification of causal mechanisms) is perfectly happy with the idea that there is a small repertoire of causal mechanisms that work in widely different circumstances. He is also perfectly happy with the view that "the basic mechanisms conform to general laws" [ibid.]. Unification, Salmon stresses, promotes our understanding of the phenomena. Nonetheless, he takes it that the causal order of the world is (metaphysically) prior to the explanatory order. He thinks that the 'because' of explanation is always dependent on the 'because' of causation.

What is a causal mechanism? Salmon offers a *generic* account of causal mechanism, based on two key causal concepts: causal process and causal interaction. Though sometimes he talked as if causal processes and causal interactions are two distinct types of causal mechanism, they are really intertwined. According to Salmon, causal processes

> are the mechanisms that propagate structure and transmit causal influence in this dynamic and changing world. ('ldots) [T]hey provide the ties among the various spatiotemporal parts of our universe. [1997, 66]

Examples of causal processes include a light-wave travelling from the sun, or less exotically, the movement of a ball. Using the language of the Special Theory of Relativity, we can say that a process is represented by a world-line in a Minkowski diagram. An important aspect of Salmon's views is that processes are *continuous*. So, a process cannot be represented as a sequence of discrete events. The continuity of the process accounts for the direct link between cause and effect (cf. [1984, 156-7]).

Not all processes are causal. Borrowing an idea of Reichenbach's [1956], Salmon [1984, 142] characterised 'causal' those (and only those) processes that are capable of transmitting a mark. Consequently, non-causal (or "pseudo") processes are those (and only those) that cannot transmit a mark. Intuitively, to mark a process is to interact with it so that a *tag* is put on it. A moving white ball (that is, a process) can be marked by simply painting a red spot on the ball. But it is not enough that the process can be 'markable'. The process should be such that, after the mark has been put on it, by means of a single local interaction, the mark gets *transmitted*. Salmon insists on the *transmission* of the mark because without it, there cannot be an adequate characterisation of causal processes. *Any* process can be marked by means of a single local interaction. In order to avoid the trivialisation of the mark-method, Salmon insists that the mark should be *transmitted* by the process, after the interaction which marked it has taken place (cf. [1984, 142]).

Salmon [1984, 144] goes on to characterise the mark method a bit more formally. A process, be it causal or not, exhibits "a certain structure". A causal process is said to be a process capable of transmitting its own structure. But, Salmon adds, "if a process — a causal process — is transmitting its own structure, then it will be capable of transmitting certain modifications in the structure" [1984, 144]. A mark, then, is a modification of the structure of a process. And a process is causal if it is capable of transmitting the modification of its structure that occurs in a single local interaction. It should be noted, however, that Salmon offers *two* criteria for a process being causal. The *first* is that it is capable of transmitting its own structure, i.e., that it is, in some sense, self-maintaining or self-persisting, or self-determined. This criterion says nothing about marking, unless of course one thinks that the structure that characterises a process is its own mark. But even so, whether a process is causal will depend, on the *first* criterion, on whether the process is capable of *transmitting its own structure*. What pseudo-processes cannot do is *transmit* their structure, unless they are under the influence of some "external agency". The *second* criterion that Salmon offers is that a process is causal if it is capable of transmitting *modifications* of its structure. This modification is, clearly, a marking of the process. Hence, the second criterion is a genuine marking criterion. However, the two criteria are conceptually distinct. They are not even necessarily co-extensive. For instance, a photon might be rightly deemed as causal process according to the first criterion, but it seems that it cannot be a causal process on the second criterion, since it admits of no modification of its structure (assuming that it has one).

Isn't the ability to transmit a mark "a mysterious power"? Salmon's master thought is that there is no mystery in the view that a mark is transmitted from a point A of a process to a subsequent point B, if we take on board Russell's 'at-at' theory of motion. According to this theory — which Russell's developed as a reply to Zeno's paradox of the arrow — "to move from A to B is simply to occupy the intervening points at the intervening instants" [1984, 153]. That is, to move from A to B is to be *at* the intervening points, *at* the intervening times. Salmon (and Russell) argue that this is a *complete* explanation of the motion since there is no additional question (and hence no extra pressure to explain) why (or how) the object *gets* from point A to point B. Consequently, Salmon [1984, 148] defines mark-transmission (MT) as follows:

> (MT) Let P be a process that, in absence of interactions with other processes, would remain uniform with respect to a characteristic Q, which it would manifest consistently over an interval that includes both of the space-time points A and B ($A \neq B$). Then, a *mark* (consisting of a modification of Q into Q'), which has been introduced into process P by means of a single local interaction at point A, is *transmitted* to point B if P manifests the modification Q' at B and at all stages of the process between A and B without additional interventions.

Note that the first clause of MT strengthens the criteria for a process being causal by introducing a counterfactual characterisation, viz., that "the process P would have continued to manifest the characteristic Q if the specific marking interaction had not occurred" [1984, 148]. This is a considerable strengthening because the two criteria that we have encountered so far (viz., transmission of P's own structure, and transmission of a modification of P's own structure) make no references to counterfactuals. The strengthening, however, is necessary because there can be pseudo-processes which satisfy the second clause of MT.

MT makes extensive reference to the presence and absence of interactions. In his [1984, 171], Salmon defines *Causal Interaction* (*CI*) as follows:

> (CI): Let P_1 and P_2 be two processes that intersect with one another at the space-time point S, which belongs to the histories of both. Let Q be a characteristic that process P_1 would exhibit throughout an interval if the intersection with P_2 did not occur; let R be a characteristic that process P_2 would exhibit throughout an interval (which includes subintervals on both sides of S in the history of P_2) if the intersection with P_1 did not occur. Then, the intersection of P_1 with P_2 at S constitutes a causal interaction if:
>
> 1. P_1 exhibits the characteristic Q before S, but it exhibits a modified characteristic Q' throughout an interval immediately following S; and
> 2. P_2 exhibits the characteristic R before S, but it exhibits modified characteristic R' throughout an interval immediately following S.

The formulation of *CI* involves, once more, counterfactuals. This is to secure that intersections between pseudo-processes do not count as causal interactions.

Besides, the actual wording of *CI* is such that the concept of causal interaction is defined in terms of the geometric (i.e., non-causal) concept of *intersection* of two processes. It might then appear that Salmon offers an analysis of causal mechanism in non-causal terms. But this is not the case. *CI* makes an essential (if implicit) reference to *marks*, and hence to *causal* processes. Salmon thinks that his appeal to the non-causal concept of *intersection* is enough to ground his theory in non-causal terms. Here is the mature formulation of this theory [1997, 250]:

$S - I$	A process is something that displays consistency of characteristics.
S-II	A mark is an alteration to a characteristic that occurs in a single local intersection.
S-III	A mark is transmitted over an interval when it appears at each spacetime point of that interval, in the absence of interactions.
S-IV	A causal interaction is an intersection in which both processes are marked (altered) and the mark in each process is transmitted beyond the locus of the intersection.
$S - V$	In a causal interaction a mark is introduced into each of the intersecting processes.
S-VI	A causal process is a process that can transmit a mark.

Given this formulation, he is confident that his account is cast in non-causal terms. Yet, even if Salmon is right in this, it's not clear that his account in terms, ultimately, of intersections, is strong enough to characterise causal *interactions*.[10,11]

Suppose we were to leave aside the problems mentioned so far. The question to ask, then, would be the following: is Salmon's mark-method adequate as a theory of causation and hence of causal explanation? The key element of his theory is the idea of mark-transmission. Is, then, mark transmission necessary and sufficient for a process being causal? Kitcher [1985] has argued that it is neither. Take the case of a pseudo-process, e.g., the shadow of a moving car. This can be permanently marked by a single local interaction. The car crashes on a wall and a huge dent appears on its bonnet. The shadow of the car acquires, and transmits, a permanent mark: it is *the shadow of a crashed car*. So, the mark-transmission is not sufficient for a process being causal. Conversely, a process can be causal even if it does *not* transmit a mark. To see how this is possible, consider Salmon's requirement that

[10] For more on this see my [2002, 117-8].

[11] An important aspect of Salmon's theory that we shall not discuss, concerns the "production" of causal processes. Salmon's main idea is that the "production of structure and order" in the world is, at least partly, due to the existence of "conjunctive forks", which are exemplified in situations in which a common cause gives rise to two or more effects. The core of this idea goes back to Reichenbach's [1956], though Salmon also adds further cases of causal forks, such as "interactive forks" and "perfect forks", which correspond to different cases of common-cause situations. Salmon uses statistical relations among events to characterise causal forks. He also argues that it is the *de facto* direction of the causal forks from past events to future events that constitutes the direction of causation. For the details of Salmon's views, see [1984, chapter 6; 1997, chapter 18]. For criticisms, see [Dowe, 2000, 79-87].

a process should remain uniform with respect to a characteristic Q for some time. This is necessary in order to distinguish a process (be it causal or not) from what Kitcher has aptly called "spatiotemporal junk". This requirement however seems to exclude from being causal many genuine processes that are short-lived, e.g., the generation and annihilation of virtual (subatomic) particles (cf. [Dowe 2000, 74]).

A generic problem to which the above counterexamples point is the vagueness of the notion of 'characteristic Q', which gets either transmitted or modified in a causal process. Salmon could block the first of the counterexamples above by denying, for instance, that the modification of the shadow of the car after the crash is a modification of a genuine characteristic of the shadow. In specific cases, we seem to have a pretty clear idea of what this characteristic might be, e.g., the chemical structure of a molecule, or the energy-momentum of a system, or the genetic material of an organism. Once, however, we start thinking about all this in very abstract philosophical terms, it is not obvious that we can say anything other than this characteristic being a *property* of a process. Then again, new problems arise. For at this very abstract level, *any* property of *any* process might well be suitable for offering the markable characteristic of the process. We seem to be in need of an account of which properties are such that their presence or modification marks a causal process.

It has been repeatedly noted that Salmon's theory relies heavily on the truth of certain counterfactual conditionals. This has led some philosophers (e.g., [Kitcher 1989]) to argue that, in the end, Salmon has offered a variant of the counterfactual approach to causation. Such an approach would bring in its tow all the problems that counterfactual analyses face. In particular, it would seem to undermine Salmon's aim to offer an objective analysis of causation. For, it is an open issue whether or not there can be a fully objective theory of the truth-conditions of counterfactuals. In any case, Salmon has always been very sceptical about the objective character of counterfactual assertions. So, as he said, it was "with great philosophical regret", that he took counterfactuals on board in his account of causation (cf. [1997, 18]). The question then is whether his account could be formulated without appeal to counterfactuals.

The short answer to the above question is: *yes, but...* For the mark-method has to be abandoned altogether and be replaced by a variant theory, which seems to avoid the need for counterfactuals. The counterexamples mentioned above, as well as the need to avoid counterfactuals, led Salmon to argue that "the capacity to transmit a mark" is not constitutive of a causal process, but rather a "symptom" of its presence [1997, 253]. So, causal processes, i.e., the "the causal connections that Hume sought, but was unable to find" [1984, 147], should be identified in a different way. The best attempt so far to articulate this different way is Dowe's [2000] theory of *conserved quantities.* According to this theory:

> The central idea is that it is the possession of a conserved quantity, rather than the ability to transmit a mark, that makes a process a causal process. [2000, 89]

We shall not discuss this theory here. It shall only be noted that even if it is granted that it offers a neat account of physical causal mechanisms, it can be generalised as a theory of causal mechanisms *simpliciter* only if it is married to strong reductionistic views that all worldly phenomena (be they social or psychological or biological) are, ultimately, reducible to physical phenomena.

15 EXPLANATION AS MANIPULATION

James Woodward [2003] has put forward a 'manipulationist' account of causal explanation. Briefly put, c causally explains e if e causally depends on c, where the notion of causal dependence is understood in terms of relevant (interventionist) counterfactual, i.e., counterfactuals that describe the outcomes of interventions. A bit more accurately, c causally explains e if, were c to be (actually or counterfactually) manipulated, e would change too. This model ties causal explanation to actual and counterfactual experiments that show how manipulation of factors mentioned in the *explanans* would alter the *explanandum.* It also stresses the role of invariant relationships, as opposed to strict laws, in causal explanation. Explanation in this model consists in answering a network of "what-if-things-had-been-different questions", thereby placing the *explanandum* within a pattern of counterfactual dependencies (cf. [Woodward, 2003, 201]). The law of ideal gases, for instance, is said to be explanatory not because it renders a certain *explanandum* (e.g., that the pressure of a certain gas increased) nomically expected, but because it can tell us how the pressure of the gas would have changed, had the antecedent conditions (e.g., the volume of the gas) been different. The explanation proceeds by locating the *explanandum* "within a space of alternative possibilities" [Woodward 2003, 191]. The key idea, I take it, is that causal explanation shows how the *explanandum* depends on the *explanans* in a stable way.

Let us describe, somewhat sketchily, the two key notions of intervention and invariance. The gist of Woodward's characterisation of an *intervention* is this. A change of the value of X counts as an intervention I if it has the following characteristics:

a) the change of the value of X is entirely due to the intervention I;

b) the intervention changes the value of Y, if at all, only through changing the value of X.

The *first* characteristic makes sure that the change of X does not have causes other than the intervention I, while the *second* makes sure that the change of Y does not have causes other than the change of X (and its possible effects).[12] These characteristics are meant to ensure that Y-changes are exclusively due to X-changes, which, in turn, are exclusively due to the intervention I. As Woodward

[12] There is a *third* characteristic too, viz., that the intervention I is not correlated with other causes of Y besides X.

stresses, there is a close link between intervention and manipulation. Yet, his account makes no special reference to human beings and their (manipulative) activities. In so far as a process has the right characteristics, it counts as an intervention. So interventions can occur 'naturally'.

Woodward links the notion of intervention with the notion of *invariance*. A certain relation (or a generalisation) is invariant, Woodward says, "if it would continue to hold — would remain stable or unchanged — as various other conditions change" [2000, 205]. What really matters for the characterisation of invariance is that the generalisation remains stable under a set of actual and counterfactual *interventions*. So Woodward [2000, 235] notes:

> the notion of invariance is obviously a modal or counterfactual notion [since it has to do] with whether a relationship would remain stable if, perhaps contrary to actual fact, certain changes or interventions were to occur.

Let me highlight three important general elements of Woodward's approach. *First*, causal claims relate variables. Causes should be such that it makes sense to say of them that they could be changed or manipulated. Thinking of them as variables, which can take different values, is then quite natural. But as Woodward goes on to note, it is not difficult to translate talk in terms of changes in the values of variables into talk in terms of events and conversely. For instance, instead of saying that the hitting by the hammer (an event) caused the shattering of the vase (another event), we may say that the change of the value of a certain indicator variable from *not-hit* to *hit* caused the change of the value of another variable from *unshattered* to *shattered*. This strategy, however, will not work in cases in which putative causes cannot be understood as values of variables. But then again, this is fine for Woodward, as he claims that in those cases causal claims will be, to say the least, ambiguous (cf. [2003, 115ff]).

Second, generalisations need not be invariant under *all* possible interventions. Hooke's law, for instance, would 'break down' if one intervened to stretch the spring beyond its breaking point. Still, Hooke's law does remain invariant under some set of interventions. In so far as a generalisation is invariant under a certain range of interventions, it can be explanatorily useful, without being exceptionless (cf. [2000, 227-8]). Woodward [2000, 214] stresses: "[t]here are generalisations that are invariant and that can be used to answer a range of what-if-things-had-been-different questions and that hence are explanatory, even though we may not wish to regard them as laws and even though they lack many of the features traditionally assigned to laws by philosophers". In particular, a generalisation can be *causal* even if it is not universally invariant (cf. [2003, 15]).

Third, Woodward does not aim to offer a reductive account of causation or causal explanation. The notion of intervention is itself causal and, in any case, causal considerations are necessary to specify when a relationship among some variables is causal. For instance, an appropriate intervention I on variable X with respect to variable Y should be such that it is not correlated with other *causes* of

Y or does not directly *cause* a change of the value of Y. I think Woodward [2003, 104-7] is right in insisting that his account is not trapped in a vicious circle. In any case, an account of causation or causal explanation need not be reductive to be illuminating.

In light of the above, causal explanation proceeds by exploiting the manipulationist element of causation and the invariant element of generalisations. Explanatory information "is information that is potentially relevant to manipulation and control" [Woodward, 2003, 10]. Causal relations are explanatory because they provide information about counterfactual dependencies among causal variables. And invariant generalisations are explanatory because they exhibit stable patterns of counterfactual dependence among causal variables in virtue of which different values of the effect-variable counterfactually depend on different values of the cause-variable.

There have been many significant attempts to offer semantic for counterfactual conditionals. Perhaps the most well-developed, and certainly the most well-known, is Lewis's [1973] account in terms of possible worlds. I will not discuss this theory here.[13] The relevant point is that Woodward offers an account of counterfactuals that tries to avoid the metaphysical excesses of Lewis's theory.

Woodward is very careful in his use of counterfactuals. Not all of them are of the right sort for the evaluation of whether a relation is causal. Only counterfactuals that are related to *interventions* can be of help. An intervention gives rise to an "active counterfactual", that is, to a counterfactual whose antecedent is made true by interventions. In his [2003, 122] he stresses that

> the appropriate counterfactuals for elucidating causal claims are not just any counterfactuals but rather counterfactuals of a very special sort: those that have to do with the outcomes of hypothetical interventions. (...) it does seem plausible that counterfactuals that we do not know how to interpret as (or associate with) claims about the outcomes of well-defined interventions will often lack a clear meaning or truth value.

It follows that the truth-conditions of counterfactual statements (and their truth-values) are not specified by means of an abstract metaphysical theory, e.g., by means of abstract relations of similarity among possible worlds.

The main problem that I see in Woodward's theory relates to the question: what is it that makes a certain counterfactual conditional true? Woodward stresses that causal claims are irreducible:

> According to the manipulationist account, given that C causes E, which counterfactual claims involving C and E are true will always depend on which other *causal* claims involving other variables besides C and E are true in the situation under discussion. For example, it

[13] See my [2002, 92-101].

> will depend on whether other causes of E besides C are present. [2003, 136]

The idea here, I take it, is that the truth of counterfactuals depend on the truth of certain causal claims, most typically causal claims about the larger causal structure in which the variables that appear in the counterfactuals under examination are embedded. Intuitively, this is a cogent claim. Consider two variables X and Y and examine the counterfactual: if X had changed (that is, if an intervention I had changed the value of X), the value of Y would have changed. Whether this is true or false will depend on whether I causes the value of Y to change by a route independent of X, or on whether some other variable Z causes a direct change of the value of Y. Causal facts such as these are part of the truth-conditions of the foregoing counterfactual. It is clear that they may, or may not, obtain independently of any intervention on X. So whether or not an intervention I on X were to occur, it might be the case that were it to occur, it would not influence the value of Y by a route independent of X. The thought, then, may be that the truth-conditions of a counterfactual are specified by certain causal facts that involve the variables that appear in the counterfactual as well as the variables of the broader causal structure in which the variables of interest are embedded.

It appears, however, that this last thought leads to an unacceptable circle. Causal claims, we are told, should be understood in terms of counterfactual dependence (where the counterfactuals are interventionist). To fix our ideas, let us consider the causal claim

> B_0: X causes Y.

For B_0 to be true, the following counterfactual C_1 should be true.

> C_1: if X had changed (that is, if an intervention I had changed the value of X), the value of Y would have changed.

On the thought we are presently considering, the truth of C_1 will depend, among other things, on the truth of another causal claim:

> B_1: I does not cause a change to the value of Y directly, (that is, by a route independent of X).

How does the truth of B_1 depend on counterfactuals? Let us assume that relations of counterfactual dependence are *part* of the truth-conditions of causal claims. Then, at least *another* (interventionist) counterfactual C_2 would have to be true in order for B_1 to be true.

> C_2: if an(other) intervention I' had changed the value of I, the value of Y would not have changed (by a route independent of X).

But what makes C_2 true? Suppose it is another causal claim B_2.

B_2: I' does not cause a change to the value of Y directly.

For B_2 to be true, another counterfactual C_3 would have to be true, and so on. Either a regress is in the offing or the truth of some causal claims has to be accepted as a brute fact. In the former case, counterfactuals are part of the truth-conditions of other counterfactuals, with no independent account of what it is for a counterfactual to be true. In the latter case, we are left in the dark as to what causal claims capture brute facts. In particular, why should we not take it as a brute fact that B_0 or B_1 is true?

In any case, it turns out that there are sensible counterfactuals that fail Woodward's criterion of actual and hypothetical interventions. Some of them are discussed by Woodward himself [2003, 127-33]. Consider the true causal claim: Changes in the position of the moon with respect to the earth and corresponding changes in the gravitational attraction exerted by the moon on the earth's surface cause changes in the motion of the tides. As Woodward adamantly admits, this claim cannot be said to be true on the basis of interventionist (experimental) counterfactuals, simply because realising the antecedent of the relevant counterfactual is physically impossible. His response to this is an alternative way for assessing counterfactuals. This is that counterfactuals can be meaningful if there is some "basis for assessing the truth of counterfactual claims concerning what would happen if various interventions were to occur". Then, he adds, "it doesn't matter that it may not be physically possible for those interventions to occur" [2003, 130]. And he sums it up by saying that "an intervention on X with respect to Y will be 'possible' as long it is logically or conceptually possible for a process meeting the conditions for an intervention on X with respect to Y to occur" [2003, 132]. We now have a much more liberal criterion of meaningfulness at play, and it is not clear, to say the least, which counterfactuals end up meaningless by applying it.[14]

Perhaps, the foregoing worries do not affect causal explanation as a practical activity. In many practical cases, we may well have a lot of information about a particular causal structure and this may be enough to answer questions about which (interventionist) counterfactuals are true and what generalisations are invariant under interventions. When we deal with *stable causal or nomological structures* interventionist counterfactuals are meaningful and truth-valuable. The worries raised in this section concern the prospects of the manipulationist account as a philosophical theory of causal explanation. As it stands, Woodward's theory highlights and exploits the *symptoms* of a good causal explanation, without offering a fully-fledged theory of what causal explanation consists in. Invariance-under-interventions is a symptom of causal relations and laws. It is not what causation or lawhood consists in.

[14] For more on Woodward's account of causal explanation and the role of invariant generalisations in it, see my [2002, 182-187].

16 STATISTICAL EXPLANATION

Suppose we want to explain a statistical regularity, viz., the fact that in a large collection of atoms of the radioactive isotope of Carbon-14 ($C14$) approximately three-quarters of them will very probably decay within 11,460 years. This, Hempel [1965, 380-1] observed, can be explained *deductively* in the sense that its description can be the conclusion of a valid deductive argument, whose premises include a statistical nomological statement. The general claim above follows deductively from the statement that every $C14$ atom has a probability of 1/2 of disintegrating within any period of 5,730 year (provided that it is not exposed to external radiation). There is no big mystery here. A valid deductive argument can have a statistical generalisation as its conclusion provided that one of the premises also contains some suitable probabilistic statement. Hempel called this account the *Deductive-Statistical* (DS) model of explanation. Salmon [1989, 53] rightly observes that the DS-model is just a species of the *Deductive-Nomological* model, when the latter is applied to the explanation of statistical regularities.

But, there is more to statistical explanation than the DS-model can cover. We are also interested in explaining *singular events* whose probability of happening is less than unity. Suppose, to exploit one of Hempel's own examples (cf. [1965, 382]), that Jones has suffered from septic sore throat, which is an acute infection caused by bacteria known as *streptococcus hemolyticus*. He takes penicillin and recovers. There is no strict (deterministic) law, which says that whoever is infected by streptococcus and takes penicillin will recover quickly. Hence, we cannot apply the Deductive-Nomological model to account for Jones's recovery. Nor can we apply the DS-model, since what we want to explain is an individual event; not a statistical regularity. How are we to proceed?

Suppose that there is a statistical generalisation of the following form: whoever is infected by streptococcus and takes penicillin has a very high probability of recovery. Let's express this as follows:

$$\text{prob}(R/P\&S) \text{ is very high,}$$

where 'R' stands for quick recovery, 'P' stands for taking penicillin and 'S' stands for being infected by streptococcus germs. We can then say that given this statistical generalisation, and given that Jones was infected by streptococcus and took penicillin, the probability of Jones's quick recovery was high. Hence, we have *inductive* grounds to expect that Jones will recover. We can then construct an inductive argument that constitutes the basis of the explanation of an event whose occurrence is governed by a statistical generalisation. This is the birth of Hempel's *Inductive-Statistical* model (IS). Let a stand for Jones, and let 'R', 'P' and 'S' be as above. Applied to Jones's case, the IS-explanations can be stated thus:

$$\begin{array}{ll}
(1) & \\
Sa \text{ and } Pa & \\
\text{prob}(R/P\&S) \text{ is very high} & \\
\hline\hline
Ra & \text{[makes practically certain (very likely)]}
\end{array}$$

More generally, the logical form of an *Inductive-Statistical* explanation is this:

$$\begin{array}{ll}
(IS) & \\
Fa & \\
\text{prob}(G/F) = r\text{, where } r \text{ is high (close to 1)} & \\
\hline\hline
Ga. & [r]
\end{array}$$

The double line before the conclusion indicates that it is an *inductive* argument. The conclusion follows from the premises with high probability. The strength r of the inductive support that the premises lend to the conclusion is indicated in square brackets. As noted by Hempel, the fact an *IS*-explanation rests on an inductive argument does not imply that its premises cannot explain the conclusion. After all, *Ga* did occur and we can explain this by saying that, given the premises, we would have expected *Ga* to occur.

The *Inductive-Statistical* model inherits a number of important features of the *Deductive-Nomological* model. The *IS*-model makes explanations arguments, albeit inductive. It also understands explanation as nomic expectability. To explain an event is still to show how this event would have been *expected* (with high probability) to happen, had one taken into account the statistical laws that govern its occurrence, as well as certain initial conditions. The *IS*-model needs an essential occurrence of law-statements in the *explanans*, albeit expressing statistical laws.

Hempel's requirement of high probability is essential to his *Inductive-Statistical* model. It's this requirement that makes the *IS*-model resemble the *DN*-model, and it is also this requirement that underwrites the idea that an *IS*-explanation is a good inductive argument. Yet, this requirement is exactly one of the major problems that the *IS*-explanation faces. For we also need to explain events whose occurrence is *not* probable, but which, however, do occur. Richard Jeffrey [1969] highlighted this weakness of the *IS*-model by noting that the requirement of high probability is not a necessary condition for statistical explanation. We must look elsewhere for the hallmark of good statistical explanation. In particular, if the requirement of high probability is relaxed, then statistical explanations are no longer arguments.

Is the requirement of high probability sufficient for a good statistical explanation? The answer is also negative. To see why, we should look at some aspects of the statistical regularities that feature in the *Inductive-Statistical* model. Suppose we explain why Jones recovered from a common cold within a week by saying that he took a large quantity of vitamin C. We can then rely on a statistical law, which says that the probability of recovery from common colds within a week, given taking vitamin C, is very high. The formal conditions for an *IS*-explanation are met and yet the argument offered is not a good explanation of Jones's recovery from common cold. For, the statistical law is no good. It is *irrelevant* to the explanation of recovery since common colds, typically, clear up after a week, irrespective of the administration of vitamin C. This suggests that more stringent requirements should be in place if a statistical generalisation is to be explanatory. High probability is not enough.

It is noteworthy that the specific example brings to light a problem of *IS* that seems to be detrimental. The reason why we think that the foregoing statistical generalisation is not explanatory is that we, rightly, think that it fails to capture a *causal connection* between recovery from common colds and the administering of vitamin C. That two magnitudes (or variables) are connected with a high-probability statistical generalisation does not imply that they are connected causally. Even when the connection is not statistical but deterministic, it still does not follow that they are causally connected. Correlation does not imply causation. To say the least, two magnitudes (or variables) might be connected with a high-probability statistical generalisation (or by a deterministic one) because they are effects of a common cause. So, the causal arrow does not run from one to the other, but instead, from a common cause to both of them.

It might be thought that the *Inductive-Statistical* model is not aimed at causal explanation. Indeed, Hempel refrained from explicitly connecting *IS*-explanation with causal explanation (cf. [1965, sections 3.2 &3.3]). However, in his [1965, 393], he toyed with the idea that the *Inductive-Statistical* model offers

> a statistical-probabilistic concept of 'because' in contradistinction to a strictly deterministic one, which would correspond to deductive-nomological explanation.

But then, it's fair to say that insofar as the *IS*-model aims to capture a sense of statistical (or probabilistic) causation, it fails.

Enough has been said so far to bring to light the grave difficulties of the *Inductive-Statistical* model. But there is another one, which will pave the way for a better understanding of the nature of statistical explanation, and its relation to causation. Hempel [1965, 394] called this problem "the ambiguity of inductive-statistical explanation".

Valid deductive arguments have the property of *monotonicity*. If the conclusion Q follows deductively from a set of premises P, then it will also follow if further premises P^* are added to P. Inductive arguments, no matter how strong they may be, lack this property: they are *non-monotonic*. The addition of extra premises P^* to an inductive argument may even remove the support that the original set of premises P conferred on the conclusion Q. In fact, the addition of extra premises P^* to an inductive argument may be such that the *negation* of the original conclusion becomes probable. Take our stock example of Jones's recovery from streptococcal infection and refer to its *IS*-explanation (1) above. Suppose, now, that Jones was, in fact, infected by a germ of streptococcus that was resistant to penicillin. Then, Jones's taking penicillin cannot explain his recovery. What is now likely to happen is that Jones won't recover from the infection, despite the fact that he took penicillin, and despite the fact that it is a true statistical generalisation that most people who take penicillin recover from streptococcus infection. The addition of the extra premise that Jones was infected by a penicillin-resistant strain (Ta) will make it likely that Jones won't recover ($\textit{not-Ra}$). For now the probability $\text{prob}(\textit{not-R}/P\&S\&T)$ of non-recovery ($\textit{not-R}$) given penicillin (P), strept infection (S), and a penicillin-resistant germ (T) is very high. So:

(2)

$$\begin{array}{ll} Sa \text{ and } Pa & \\ Ta & \\ \text{prob}(\textit{not-R}/P\&S\&T) \text{ is close to } 1 & \\ \hline & [\text{makes practically certain (very likely)}] \\ \hline \textit{not-Ra} & \end{array}$$

The non-monotonic nature of *Inductive-Statistical* explanation makes all this possible. (1) and (2) are two arguments with mutually consistent premises and yet incompatible conclusions. It is this phenomenon that Hempel called the "ambiguity" of *IS*-explanation. What is ambiguous is to what reference class to include the *explanandum*. Given that it may belong to lots of different reference classes, which one shall we choose? The problem is precisely that different specifications of the reference class in which the *explanandum* might be put will lead to different estimations of the probability of its occurrence. Consider the following: what is the probability that an individual lives to be 80 years old? The answer will vary according to which reference class we place him/her.

The problem we are discussing is accentuated if we take into account the fact that, even if there was an objectively correct reference class to which an individual event belongs, in most realistic cases when we need to explain an individual event, we won't be able to know whether the correct identification of the reference class has been made. We will place the individual event in a reference class in order to *IS*-explain its occurrence. But will this be the *right* reference class, and how can we know of it? This is what Hempel [1965, 395] called the *epistemic version* of the ambiguity problem. The result of this is that an *IS*-explanation should always be relativised to a body K of currently accepted (presumed to be true) beliefs.

Note that the problem of ambiguity does not arise in the case of the *Deductive Nomological* explanation. The premises of a *DN*-argument are *maximally specific.* If it is the case that *All Fs are Gs*, then no further specification of *Fs* will change the fact that they are Gs. As the jargon goes, if all Fs are Gs, no further partition of the reference class F can change the probability of an instance of F to be also an instance of G, this probability being already equal to unity. On the contrary, in an *IS*-explanation, further partitions of the reference class F can change the probability that an instance of F is also an instance of G.

This suggests that we may introduce the *Requirement of Maximal Specificity* (*RMS*) to *Inductive-Statistical* explanation. Roughly, to say that the premises of an *IS*-explanation are maximally specific is to say that the reference class to which the *explanandum* is located should be the narrowest one. More formally, suppose that the set P of premises of an *IS*-explanation of an individual event Fa imply that prob(G/F)=r. The set of premises P is maximally specific if, given that background knowledge K tells us that a also belongs to a subclass F_1 of F, and given that prob(G/F_1)=r$_1$, then r=r$_1$.[15]

Let's call a reference class *homogeneous* if it cannot be further partitioned into subclasses which violate *RMS*. Clearly, there are two concepts of homogeneity. The *first* is objective: there is no partition of the reference class into subclasses which violate *RMS*.[16] The *second* is epistemic: we don't (currently) know of any partition that violates *RMS*. Hempel's version of *RMS* was the latter. Hence, *IS*-explanation is always relativised to a certain body of background knowledge K, which asserts what partitions of the reference classes are *known* to be relevant to an *IS*-explanation of an individual event. The fact that *IS*-explanations are always epistemically relative has made many philosophers think that the *IS*-model cannot be an adequate model of statistical explanation (cf. [Salmon 1989, 68ff]). What we would need of a statistical explanation is an identification of the relevant features of the world that are nomically connected (even in a statistical sense) with the *explanandum*. The *Inductive-Statistical* model is far from doing that, as the problem with the *Requirement of Maximal Specificity* makes vivid.

The friends of statistical explanation face a dilemma. They might take the view that all genuine explanation is *Deductive-Nomological* (*DN*) and hence treat statistical explanation as *incomplete* explanation. If, indeed, all explanation is *Deductive-Nomological*, then the problem of the reference class (and of *RMS*) does not even arise. On this view, an *IS*-explanation is a place-holder for a full *DN*-explanation of an individual event. The statistical generalisations are taken to express our *ignorance* of how to specify the correct reference class in which we should place the *explanandum*. This approach is natural, if one is committed to determinism. According to determinism, every event that occurs has a fully determinate and sufficient set of antecedent causes. Given this set of causes, its probability to happen is unity. If we knew this full set of causes of the *explanandum*, we could use this information to objectively fix its reference class and we would, thereby, establish a true universal generalisation under which the *explanandum* falls. If, for instance, the full set of causes of event-type E was the conjunction of event-types F, G and H, we could simply say that *All Fs &Gs &Hs are Es*. So, on the view presently discussed, statistical generalisations simply express our

[15] The exact definition, which is slightly more complicated and accurate than the one offered here, is given by Hempel [1965, 400]. An excellent detailed account of *RMS* is given in Salmon [1989, 55-7].

[16] The concept of objective homogeneity and its implications are discussed in Salmon [1984, chapter 3].

ignorance of the full set of causes of an event. They are by no means useless, but they are not the genuine article. This view is elaborated by Kitcher [1989].

Alternatively, the friends of statistical explanation could take the view that there is *genuine* statistical explanation, which is nonetheless captured by a model different to the *Inductive-Statistical*. In order to avoid the pitfalls of the *IS*-model, they would have to admit that there is a fact of the matter as to the *objectively homogeneous* reference class in which a certain *explanandum* belongs. But this is not enough for genuine statistical explanation, since, as we saw in the previous paragraph, the existence of an objectively homogeneous reference class is compatible with the presence of a universal law. So, the friends of genuine statistical explanation should also accept that even within an objectively homogeneous reference class, the probability of an individual event's occurring is not unity. They have to accept indeterminism: there are no further facts that, were they taken into account, would make this probability equal to unity. An example (cf. [Salmon 1989, 76]) will illustrate what is at issue here. Take a collection of radioactive carbon-14 atoms whose half-life time is 5730 years. This class is as close to being objectively homogeneous as it can be. No further partitions of this class can make a sub-class of carbon-14 atoms have a different half-life time. What is important here is that the law that governs the decay of carbon-14 atoms is *indeterministic*. The explanations that it licenses are genuinely statistical, because the probability that an atom of carbon-14 to decay within 5730 years is irreducibly 1/2. In genuine statistical explanation, there is no room to ask certain why-questions. Why did this *specific* carbon-14 atom decay? If indeterminism is true, there is simply no answer to this question.

17 STATISTICAL RELEVANCE

Take an event-type E whose probability to happen given the presence of a factor C (i.e., prob(E/C)) is r. In judging whether a further factor C_1 is relevant to the explanation of an individual event that falls under type E, we look at how taking C_1 into account affects the probability of E to happen. If prob$(E/C \; \& C_1)$ is different from prob(E/C), then the factor C_1 is *relevant* to the occurrence of E. Hence, it should be *relevant* to the explanation of the occurrence of an individual event that is E. Let's say that:

- C_1 is positively relevant to E, if prob$(E/C \; \& C_1) >$ prob(E/C);
- C_1 is negatively relevant to E, if prob$(E/C \; \& C_1) <$ prob(E/C);
- and C_1 is irrelevant to E, if prob$(E/C \; \& C_1) =$ prob(E/C).

Judgements such as the above seem to capture the intuitive idea of *causal relevance*. We rightly think, for instance, that the colour of one's eyes is causally irrelevant to one's recovery from streptococcus infection. We would expect that one's probability of recovery (R) given streptococcus infection (S) and penicillin

(P), i.e., prob($R/P\&S$), will be unaffected, if we take into account the colour of one's eyes (B). So, prob($R/P\&S$)=prob($R/P\&S\&B$). Analogously, we would think that the fact that one is infected by a penicillin-resistant strain of streptococcus (T) is causally relevant to his recovery (in particular, its lack). We would expect that prob($R/P\&S\&T$)¡prob($R/P\&S$). These thoughts, together with the fact that the requirement of high probability is neither necessary nor sufficient for a good statistical explanation, led Salmon [1971; 1984] to suggest a different conception of statistical explanation. The main idea is that we explain the occurrence of an individual event by citing certain statistical-relevance relations. In particular,

> a factor C explains the occurrence of an event E, if prob(E/C) $>$ prob(E) — which is equivalent to prob(E/C) $>$ prob($E/$*not*-C).

This came to be known as the *Statistical-Relevance* model (*SR*). Where an *Inductive-Statistical* explanation involves just one probability value, the *SR*-model suggests that explanation compares two probability values. As the jargon goes, we need to compare a *posterior probability* prob(E/C) with a *prior probability* prob(E). Note that the actual values of these probabilities do not matter. Nor is it required that the posterior probability be high. All that is required is that there is a *difference*, no matter how small, between the posterior probability and the prior. Suppose, for example, that the prior probability prob(R) of recovery from streptococcus infection is quite low, say .001. Suppose also that when one takes penicillin, the probability of recovery prob(R/P) is increased by only 10%. So, prob(R/P)=.01. We would not, on the *IS*-model, be entitled to explain Jones's recovery on the basis of the fact that he took penicillin. Yet, on the *SR*-model, Jones's taking penicillin is an explanatory factor of his recovery, since prob(R/P) $>$ prob(R). (Equivalently, prob(R/P) $>$ prob($R/$*not*-P))

An important feature of the *SR*-model, which paves the way to the entrance of *causation* in statistical explanation, is this. Suppose that taking penicillin is explanatory relevant to quick recovery from streptococcus infection. That is, prob(R/P) $>$ prob(R). Can we, without further ado, say that taking penicillin causes recovery from streptococcus infection? Not really. For one might be infected by a penicillin-resistant strain (T), thus rendering one's taking penicillin totally ineffective as a cure. So, if we take T into account, it is now the case that prob($R/P\&T$) = prob(R/T). The further fact of infection by penicillin-resistant germ renders *irrelevant* the fact that penicillin was administered. The probability of recovery given penicillin and infection by a penicillin-resistant germ is equal to the probability of recovery given infection by a penicillin-resistant germ. When a situation like this occurs, we say that factor T *screens off* R from P.

This relation of screening off is very important. Take the standard example in the literature. There is a perfect correlation between well-functioning barometers (B) and upcoming storms (S). The probability prob(S/B) that a storm is coming up given a drop in the barometer is higher than the probability prob(S) that a storm is coming up. So, prob(S/B)>prob(S). It is in virtue of this relationship that barometers can be used to predict storms. Can we then, using the *Statistical-Relevance* model, say that the drop of the barometer explains the storm? Worse, can we say that it causes the storm? No, because the correlation between a drop of the barometer and the storm is *screened off* by the fall of the atmospheric pressure. Let's call this A. It can be easily seen that prob($S/B\&A$)=prob(S/A). The presence of the barometer is rendered irrelevant to the storm, if we take the drop of the atmospheric pressure into account. Instead of establishing a causal relation between B and S, the fact that prob(S/B)>prob(S) points to the further fact that the correlation between B and S exists because of a *common cause*. It is typical of common causes that they screen off the probabilistic relation between their effects. But a factor can screen off a correlation between two others, even if it's not their common cause. Such was the case of infection by penicillin-resistant germ discussed above.

If the probabilistic relations endorsed by the *SR*-model are to establish genuine explanatory relations among some factors C and E, it's not enough to be the case that prob(E/C) >prob(E). It is also required that his relation not be screened off by further factors. Put more formally:

> C explains E if (i) prob(E/C) >prob(E) [equivalently, prob(E/C) > prob($E/$*not*-C)]; *and* (ii) there are no further factors H such that H screens off E from C, i.e., such that prob($E/C\&H$)=prob(E/H).

The moral of all this is that relations of statistical relevance do not imply the presence of causal relations. The converse seems also true, as the literature on the so-called Simpson paradox makes vivid. But we shall not go into this.[17] Correlations that can be screened off are called 'spurious'.

There should be no doubt that the *Statistical-Relevance* model is a definite improvement over the *Inductive-Statistical* model. Of course, if we go for the *SR*-model, we should abandon the Hempelian idea that explanations are arguments. We should also question the claim that statistical generalisations are really necessary for statistical explanation. For an *SR*-explanation is not an argument. Nor does it require citing statistical laws. Rather, as Salmon [1984, 45] put it, it is

> an *assembly of facts statistically relevant* to the explanandum, *regardless of the degree of probability* that results.

Besides, the *Statistical-Relevance* model makes clear how statistical explanation can be seen as a species of causal explanation. For if the relevant *SR*-relations are to be explanatory, they have to capture the right causal dependencies between the *explanandum* and the *explanans*. But it also paves the way for the view *that there is more to causation than relations of statistical dependence*. Salmon himself has moved from the claim that all there is to statistical explanation can be captured

[17] The 'Simpson paradox' suggests that C may cause E, even though C is not statistically co-related with E in the whole population. For more on this see Cartwright [1983, essay 1] and Suppes [1984, 55-7].

by specifying relations of statistical relevance to the claim that, even if we have all of them, we would still need to know something else in order to have genuine explanation, viz., facts about causal relationships. His latest view [1984, 34] is this:

> the statistical relationships specified in the S-R model constitute the *statistical basis* for a bone fide scientific explanation, but [. . .] this basis must be supplemented by certain *causal factors* in order to constitute a satisfactory scientific explanation.

So, according to Salmon [1984, 22], relations of statistical relevance must be explained by causal relations, and not the other way around. As we have already seen in section 14, his favoured account of causal relations is given in terms of unveiling the causal mechanisms, be they deterministic or stochastic, that connect the cause with its effect.

18 DEDUCTIVE-NOMOLOGICAL-PROBABILISTIC EXPLANATION

Does deductivism and indeterminism mix? Can, that is, one think that although all explanation is, in essence, deductive, there is still space to explain essentially chancy events? Railton's [1981] "Deductive-Nomological-Probabilistic" (*DNP*) model of probabilistic explanation is a very important attempt to show how this can happen.

Being dissatisfied with the epistemic ambiguity of the *Inductive-Statistical* model, and accepting the view that there should be space for the explanation of unlikely events, Railton [1981, 160] suggested that a legitimate explanation of a chancy *explanandum* should consist in

> a) "law-based demonstration that the explanandum had a particular probability of obtaining";
>
> and b) a claim that, "by chance, it did obtain".

Take the case of a very unlikely event such as a Uranium-238 nucleus *u*decaying to produce an alpha-particle. The mean-life of a U-238 nucleus is 6.5 X 10^9 years, which means that the probability p that such a particle will produce an alpha-particle is vanishingly small. Yet, events like this *do* happen, and need to be explained. Railton [1981, 162-3] suggests that we construct the following two-step explanation of its occurrence.

The *first* step is a straightforward *DN*-explanation of the fact that particle u has a probability p to alpha-decay during a certain time-interval Δt.

> (1a) All U-238 nuclei not subjected to external radiation have probability p to emit an alpha-particle during any time-interval Δt.
>
> (1b) u was a U-238 nucleus at time t and was not subjected to any external radiation during time-interval $[t,\ t + \Delta t]$.

Therefore, (1c) u has a probability p to alpha-decay during time-interval $[t,\ t + \Delta t]$.

This step does not yet explain why the particular particle u alpha-decayed. It only states the probability of its decay. So, Railton says, the *second* step is to add a "parenthetic addendum" [1981, 163] to the above argument. This addendum, which is put *after* the conclusion (1c), says:

(1d) u did indeed alpha-decay during the time-interval $[t,\ t + \Delta t]$.

If, in addition, the law expressed in premise (1a) is explained (derived) from the underlying theory (quantum mechanics, in this example), then, Railton [1981, 163] says, we have "a full probabilistic explanation of $u's$ alpha-decay". This is an instance of a *DNP*-explanation.

The addendum (1d) is not an extra premise of the argument. If it were, then the explanation of why did particle u alpha-decay would be trivial. So, the addendum has to be placed *after* the conclusion (1c). Still, isn't there a feeling of dissatisfaction? Have we really explained why u did alpha-decay? If we feel dissatisfied, Railton says, it will be because we are committed to determinism. If, on the other hand, we take indeterminism seriously, there is no further fact of the matter as to why particle u did alpha-decay. This is a genuine chancy event. Hence, nothing else could be added to steps (1a)-(1d) above to make them more explanatory than they already are. Note that I have refrained from calling steps (1a)-(1d) an argument because they are not. Better, (1a)-(1c) is a deductively valid argument, but its conclusion (1c) is *not* the *explanandum.* The *explanandum* is the "addendum" (1d). But this does not logically follow from (1a)-(1c). Indeed, Railton defends the view that explanations are not necessarily arguments. Although arguments, (and in particular *DN*-arguments), "play a central role" in explanation, they "do not tell the whole story" [1981, 164]. The *general schema* to which a *DNP*-explanation of a chancy event ($G_{e,t0}$) conforms is this (cf. [1981, 163]):

(2a) For all x and for all t ($F_{x,t} \rightarrow$ Probability ($G_{x,t}$)= r)
(2b) A theoretical derivation of the above probabilistic law
(2c) $F_{e,t0}$

(2d) Probability($G_{e,t0}$)=r
(2e) $G_{e,t0}$.

(2e) is the "parenthetic addendum", which is *not* a logical consequence of (2a)-(2d). As for (2a), Railton stresses that the probabilistic generalisation must be a genuine probabilistic law of nature. The explanation is true if both the premises (2a)-(2c) *and* the addendum (2e) are true.

There are a number of important features of the *Deductive-Nomological-Probabilistic* model that need to be stressed.

- *First*, it shows how the *DN*-model is a limiting case of the *DNP* model. In the case of a *DN*-explanation, (2e) just is the conclusion of the *DN*-argument — so it is no longer a "parenthetic addendum".

- *Second*, it shows that all events, no matter how likely or unlikely they may be, can be explained in essentially the same way. In schema (2) above, the value of probability r is irrelevant. It can be anywhere within the interval (0,1]. That is, it can be anything other than zero.

- *Third*, it shows that single-case probabilities, such as the ones involved in (2a), can be explanatory. No matter what else we might think of probabilities, there are cases, such as the one discussed in Railton's example, in which probabilities can be best understood as fully objective chances.

- *Fourth*, it shows how probabilistic explanation can be fully objective. Since (2a)-(2d) is a valid deductive argument, and since the probability involved in (2a) is "a law-full, physical single-case" probability (cf. 1981, 166), the *DNP*-account does not fall prey to the objections that plagued the *Inductive-Statistical* model. There is no problem of ambiguity, or epistemic relativisation. Single-case probabilities need no reference-classes, and by stating a law, premise (2a) is maximally specific.

- *Fifth*, it shows how probabilistic explanation can be freed of the requirement of nomic expectability as well as of the requirement that the *explanandum had to* occur. So, it accommodates genuinely chancy *explananda*.

- *Sixth*, by inserting premise (2b), it shows in an improved way how explanation can be linked with understanding.

Since this last point is of some special importance, let us cast some more light on it. The Hempelian tradition took explanation to be the prime vehicle for understanding in science. But, as we have already seen, it restricted understanding of why an *explanandum* happened to showing how it should have been expected to happen, given the relevant laws. In particular, it demanded that understanding should proceed via the construction of arguments, be they *Deductive-Nomological*, or *Deductive-Statistical* or *Inductive-Statistical*. Railton's *Deductive-Nomological-Probabilistic* model suggests that understanding of why an *explanandum* happened cannot just consist in producing arguments that show how this event had to be expected. The occurrence of a certain event, be it likely or not, is explained by placing this event within a *web* of

> inter-connected series of law-based accounts of all the nodes and links in the causal network culminating in the explanandum, complete with a fully detailed description of the causal mechanisms involved and theoretical derivations of all covering laws involved. [1981, 174]

In particular, explanation proceeds also with elucidating the *mechanisms* that bring about the *explanandum*, where this elucidation can only be effected if we take into account the relevant theories and models. Railton [1981, 169] rightly protested against the Hempelian view that all this extra stuff, which cannot be captured within a rigorous *DN*-argument, is simply "*marginalia*, incidental to the 'real explanation', the law-based inference to the explanandum". He does not doubt that an appeal to laws and the construction of arguments are important, even indispensable, features of explanation. But he does doubt that they exhaust the nature of explanation.

19 ON HISTORICAL AND TELEOLOGICAL EXPLANATION

Hempel's central idea, we have seen, has been that all explanation is nomological and that all explanations are arguments. This idea was meant to capture all cases of scientific explanation — not only in the natural sciences but also in historical and human sciences as well. Indeed, the very first systematic presentation of the Deductive-Nomological Pattern appeared in a paper titled "The Function of General Laws in History", in the *Journal of Philosophy* in 1942. There, Hempel advanced the DN-model in an attempt to capture *historical* explanation, in particular. This move was a radical break with a whole philosophical tradition that flourished especially on the Continental Europe, a tradition that took historical explanation to be essentially *sui generis*.

The root of this tradition is neo-Kantianism. According to Wilhelm Windelband, there is a fundamental methodological distinction between the natural and the historical sciences. He called 'nomothetic' the method suitable for the natural sciences. They are based on universal and demonstrative judgements. They aim to specify the laws of nature and strive to reveal nomological connections among events. The historical sciences, Windelband thought, are characterised by the ' idiographic' approach. They aim at individual and concrete events. The method of historical sciences, Windelband thought, is based on value-judgements, i.e., on judgements about what events are important and why. Before Windelband, the German historian and philosopher Johann Gustav Droysen introduced a distinction between *explanation* and *understanding* (in German *Erklären* and *Verstehen*) and claimed that while the natural sciences aim to explain the phenomena, the historical sciences aim to understand the phenomena that fall within their purview. Many continental thinkers, most notably Wilhelm Dilthey, took up these ideas and developed them into a whole theory of historical explanation the basis of which is the idea that explanation in history relies on a *sui generis* method of *empathic* understanding: the re-creation in the mind of the historian of the mental milieu, the motivations, the feelings, the reasons for action etc., of the historical subject (that is, the object of the historian's study). In contradistinction to explanation, historical *understanding* was thought to have psychological and intentional elements. It was not supposed to require knowledge of causes, but knowledge of *reasons*. In his influential *The Idea of History* (1946), R. G. Collingwood put forward three basic theses of historical understanding. First, in order for the historian to understand the actions of some historical subject, he must understand the thoughts that these actions express; second, once these thoughts have been grasped, the historian fully understands these actions and hence there is no further requirement for finding the causes that produced them, or the laws, if any, that govern them; and third, understanding these actions in terms of the thoughts they express requires re-thinking of the thoughts of the historical subject by the historian. All history, Collingwood, said, is "the re-enactment of past thought in the historian's own mind".

It was against this tradition that Hempel reacted by claiming that

> Historical explanation, too, aims at showing that the event in question was not 'a matter of chance', but was to be expected in view of certain antecedent or simultaneous conditions. The expectation referred to is not prophecy or divination, but rational scientific anticipation which rests on the assumption of general laws. [1942, 39]

To the obvious charge that many historical explanations fail to state any laws, Hempel replied that, as they stand, they are explanation *sketches*: when the sketches are filled out, the reference to laws (be they historical or psychological) will be made explicit. Occasionally, Hempel thought, the laws will turn out to be probabilistic. Still, the historical explanation, when fully spelt out, will be an inductive-statistical argument. Hempel, then, proposed a unified theory of understanding: there is only one type of understanding and is tantamount to offering a proper nomological explanation of the *explanandum* (be it a natural or a historical event). Understanding, he thought, has nothing to do with empathy and everything to do with nomic expectability.

Hempel, then, was adamant that all explanation consists in the subsumption of the *explanandum* under general laws. In line with the logical empiricist ideal, he thought that he could thereby secure the objectivity of scientific explanation. But is it plausible to think that objectivity requires thinking of explanation as subsumption under laws? Perhaps, Hempel was too quick to classify all explanations that fail this requirement as pseudo-explanations. The problem runs deep. For the issue at stake is precisely whether we can talk of laws of history (or of other special sciences). Hempel was aware of the problem, but he did not face it squarely. He offered only a few examples of historical laws, e.g., that populations tend to migrate to regions that offer better living conditions. But even when these laws are available, the explanation of an individual historical event, e.g., the migration of population X to region Y, will not be a straightforward deduction of the *explanandum* from the *explanans*, the reason being that the law-like statement is too vague or holds only *ceteris paribus*, that is, when other things are equal. His thought was that these laws should be filled out with detailed information that covers the *explanandum*. But it is obvious that the more detailed the law becomes (and hence the more apt to cover the individual case at hand), the less universally applicable it is; hence the less apt it becomes to be deemed a *law*.

This last thought was a key element of William Dray's critique of the Hempelian model of explanation. For Dray the problem is not that historical 'laws' are too complex or vague, but rather that they are not proper laws. On the positive side, he claimed that historical explanation is *rational* (not causal) explanation: to explain (that is, to understand) a historical action is to state its *reasons* and not its causes. In his defence of Collingwood's approach, Dray was sensitive to the charge that since causal explanation requires laws (because the *explanandum* must be necessitated, in some sense, by its cause), if historical explanation is causal explanation, historical explanation must be nomological too. His reply to this was to *deny* that historical explanation is causal explanation. Since, however, explanation must, somehow, necessitate the *explanandum*, he suggested (as the best way to explicate Collingwood's idea) that the necessity involved in historical explanation is *rational* necessity. Hence, we can explain a certain historical action by showing that it was rationally necessary: the action was the rational thing for the agent to do on the occasion under consideration. Dray went further by claiming that rational explanations of the sort he and Collingwood envisaged are complete. Further causal considerations (in terms of natural or historical laws) are irrelevant to the explanation of a historical action.

There are several objections to Dray's idea, but the most telling one comes from Davidson's famous assertion that reasons can be causes. In asserting this, Davidson unravelled the root of the problem faced by the generic pattern of explanation that underwrites Dray's suggestion, viz., the thought that *intentional* explanation is *sui generis* (non causal) explanation. Intentional explanation, a species of which is Dray's rational explanation, refers to the explanation of actions. It has been suggested that it proceeds by citing the intentions and the beliefs of the actors as the *explanans* of certain actions. According to Davidson, intentional explanation is causal explanation, since, as he put it "the primary reason for an action is its cause" [1963, 4]. This might be taken to imply that intentional explanation is *singular* causal explanation. For, it seems, for a certain action A and its cause (or reason) C, there may not be a general law, expressed in the vocabulary of A and C, such that it is the case that whenever C then A. Davidson was ready to grant this last claim, but he denied that it follows from this that causal explanation has to be singular. As we have already seen him arguing, he notes that singular claims of the form c caused e entail that *there is* some causal law that is instantiated in the particular causal sequence of events. This law might not be expressible in the vocabulary of the particular cause (reason) and the particular effect (action). Indeed, as he stressed, the relevant law might be expressed in neurological or physical vocabulary. Davidson's reconciliation of intentional and causal explanation keeps the view (central to the intentional approach) that explanation need not cite laws to be acceptable and good but it also retains the view (central to the Hempelian nomological approach) that causal explanations involve laws. We have already seen that Hempel's reaction to a Davidson-style compromise was that it amounts to claiming that there is a treasure somewhere without giving us any guidance as to how we should find it. The irony is that in the particular case of historical explanation Hempel came quite close to a Davidson-style attitude.

Intentional explanation has been described as a species of teleological explanation. We have already seen Leibniz going for this view. More recently, this view has been defended by von Wright [1971]. Teleological explanations, advanced by Aristotle and defended by Leibniz and Kant among others, are 'future oriented".

Whereas in a typical causal explanation the earlier-in-time cause explains the later-in-time effect, in teleological explanations, as traditionally understood, the later-in-time effect (that is, the aim or purpose for which something happened) explains the earlier-in-time cause (that is, why something happened). The typical locution of a teleological explanation is: *this* happened in order that *that* should occur.

A big challenge to our thinking about explanation throughout the centuries has been the issue of whether teleological explanation is *sui generis* or whether it can be subsumed under causal or mechanistic explanation ordinarily understood. To most empiricists the very idea of teleology (that is, the existence of purposes and aims in nature for the sake of which things are done) was an anathema. Vitalism, the view that the explanation of life and living organisms cannot be mechanical but should proceed in terms of vital forces or principles, was taken to be the paradigm-case of a non-scientific theory. Many working biologists and philosophers of science devoted time and energy in trying to show how biological phenomena can be explained mechanically, with no reference to vital forces and the like. But it is fair to say that even if vitalism was neutralised, the idea that biological explanations are, in some sense, teleological survived. The idea was that there is a special type of teleological explanation, viz., *functional* explanation, that is indispensable in biology.

Teleological statements can be classified as goal-ascriptions or as function-ascriptions. Goal-ascriptions state the goal or aim towards which a certain action is directed. Function-ascriptions state the function performed by something (e.g, a biological organ, or an organism, or an artefact). Goal-directed explanation explains actions in terms of their goals, while functional explanation explains the presence of some item in a system in terms of the effects that this items has in the system of which it is a part. I will not discuss these issues in great detail. But it is fair to say that goal-directed explanations can be causal in a very straightforward sense. For instance, as Ernest Nagel has noted, intentional explanation is causal in the sense that the motives, desires and beliefs of an agent explain her actions. So it is not that the goal causes the action. Rather, the agent's desire for a certain goal, together with her belief that certain actions will achieve this goal, bring about the action (as a means to achieve an end). More sophisticated forms of goal-directed behaviour are also explainable causally (see [Nagel 1977]).

What about, then, functional explanation? In biology, it is typical to explain a feature (a phenotypic characteristic) of a species in terms of its contribution to the enhancement of the chances of survival and reproduction. It is equally commonplace to explain the properties or the behaviour of the parts of an organism in terms of their functions in the whole: they contribute to the adequate functioning, the survival and reproduction of the whole. The explanation of the beating of the heart by appeal to its function to circulate the blood has become a standard example of such a functional explanation. Functional explanations are often characterized by the occurrence of teleological expressions such as 'the function of', 'the role of', 'serves as', 'in order to', 'for the sake of', 'for the purpose of'.

It seems, then, that functional explanations explain the presence of an entity by reference to its effects. Hence, they seem to defy a strict causal analysis.

The pervasiveness of functional explanations posed a double problem to all those who denied teleology. Given that they are not present in physics, given, that is, that explanation in physics is nonteleological, the presence of functional explanations in biology suggested that biology was an underdeveloped (or immature) science. But this was absurd given the scientific successes of evolutionary biology. If, on the other hand, it was accepted that there is an indispensable special type of explanation in biology, then the methodological monism that was taken to characterise all science was in danger.

Hempel and Nagel took it upon themselves to show how functional explanation can be understood in a way that has no serious teleological implications. Hempel [1959], to be sure, was sceptical of the possibility of functional explanation. It is no accident that he titled his paper "The Logic of Functional Analysis". One of his main problems was the presence of functional equivalents, i.e., the existence of different ways to perform a certain function (for instance, artificial hearts might circulate the blood). Take then the statement: The heartbeat in vertebrates has the function of circulating blood through the organism. Would it be proper to explain the presence of heartbeat by claiming that it is a necessary condition for the proper working of the organism? If it were, we could construct a proper deductive explanation of the presence of the heartbeat. We could argue thus: The presence of the heartbeat is a necessary condition for the proper working of the organism; the organism works properly; hence, the organism has a heart. But the existence of functional equivalents shows that the intended conclusion does not follow. At best, all we could infer is the presence of one of the several items of a class of items capable of performing a certain function. Hence, Hempel thought that explanation in terms of functions works only in a limited sense and have only heuristic value.

Faced with the problem of functional equivalents, Nagel [1977] suggested that if a sufficiently precise characterisation of the type of organism we deal with is offered, only one kind of mechanism will be apt to fulfil the required function. For instance, given the evolutionary history of *homo sapiens*, the heartbeat is the only mechanism available for the circulation of the blood. If Nagel is right, there can be a proper deductive account of functional explanation. According to him, here is the form that functional explanations have (illustrated by his favourite example):

> *Functional ascription*: During a period when green plants are provided with water, carbon dioxide, and sunlight, the function of chlorophyll is to enable the plants to perform photosynthesis.
>
> *Functional explanation*:
>
> 1. During a stated period, a green plant is provided with water, carbon dioxide, and sunlight.
> 2. During a stated period, and when provided with water, carbon dioxide, and sunlight, the green plant performs photosynthesis.

3. If during a given period a green plant is provided with water, carbon dioxide, and sunlight, then if the plant performs photosynthesis the plant contains chlorophyll.

Conclusion: Chlorophyll is present in the green plant.

More schematically:

(A) This plant performs photosynthesis.

(B) Chlorophyll is a *necessary condition* for plants to perform photosynthesis.

(C) Hence, this plant contains chlorophyll.

This is a deductive-nomological explanation of the presence of chlorophyll in green plants. Premises (1) and (2) state specific conditions, and premise (3) is a law-like statement. Any appearance of teleology in the functional explanation is gone. But, as Nagel [1977, 300] notes, this is *not* a causal explanation of the presence of chlorophyll. The reason is that

> the performance of [photosynthesis] is not an *antecedent* condition for the occurrence of [the chlorophyll], and so the premise [3] is not a causal law.

He concludes:

> Accordingly, if the example is representative of function ascriptions, such explanations are *not* causal — they do not account causally for the presence of the item to which a function is ascribed.

If functional explanations are not causal, what do they achieve? They make evident the role of an item within a system. Nagel was adamant that this is a legitimate role of explanation. To explain is not necessary to cite causes. Explanation is also accomplished by finding the effects or consequences of various items. As he [ibid.] put it:

> inquiries into effects or consequences are as legitimate as inquiries into causes or antecedent conditions; (...) biologists as well as other students of nature have long been concerned with ascertaining effects produced by various systems and subsystems; and (...) a reasonable adequate account of the scientific enterprise must include the examination of both kinds of inquiries [i.e., inquiries into causal antecedents and inquiries into effects or consequences].

Functional explanation is then made to fit within the deductive nomological model, but at the price of ceasing to be causal. Obviously, there are two ways to react to Nagel's suggestion. One is to try to restore the causal character of functional explanation. The other is to deny that explanations have to be arguments. Both ways were put together in Larry Wright's [1973] *etiological* model of functional explanation.

According to Wright [1973, 154], functional ascriptions are explanatory in their own right.

> Merely saying of something, X, that it has a certain function, is to offer an important kind of explanation of X.

For instance, when it is said that the fact that plants have chlorophyll is functionally explained by noting the role that chlorophyll plays in enabling the plants to perform photosynthesis, this is a genuine explanation. It does not have to be an argument of any sort. An explanation is an answer to a why-question and the question 'why do plants have chlorophyll?' is adequately and fully answered by "providing a perfectly respectable etiology; [by] provid[ing] the reason chlorophyll is there" [1976, 100]. For Wright the problem of functional equivalents does not arise, since, as he said, it is based on a false assumption, viz., that an explanation of why a certain item performs a certain function must exclude the possibility that anything else could have performed it. All that is necessary for functional explanation, according to Wright, is that the item (e.g., chlorophyll) was *sufficient* in the circumstances to perform the given function.

'Etiology' means finding the causes. Etiological explanation is *causal* explanation: it concerns the causal background of the phenomenon under investigation. To be sure, Wright's etiological explanation is causal in an extended sense of the term: it explains how "the thing with the function got there" [1973, 156]. The basic pattern of functional explanation is:

(F)

The function of X is Z iff:

(i) X is there because it does (results in) Z,

(ii) $Z$is a consequence (result) of X's being there.

For instance, the function of chlorophyll in plants is to perform photosynthesis iff chlorophyll is there because it performs photosynthesis and photosynthesis is a consequence of the presence of chlorophyll. Clause (ii) is particularly important because it makes explicit the asymmetry of functional explanation: that the function Z is there because of X and not the other way around. An important feature of Wright's account is that it is especially suitable for explanation in biology, where the notion of natural selection looms large. The etiological explanation of natural (biological) functions is in terms of natural selection: they are the results of natural selection because they have endowed their bearers with an evolutionary advantage. Consequently, etiological explanation does not reverse the causal order: a function is performed because it has been causally efficacious *in the past* in achieving a certain goal. Wright, however, insists that etiological explanation is teleological in an important sense: it is future-oriented. As he [1976, 105] put it:

> To deny the propriety of teleological explanation (. . .) is to deny the obviously right answer to [some] question[s]: namely, that [something] is there because of its (. . .) consequences.

According to Robert Cummins [1975] approaches like the ones mentioned above fail to capture what is distinctive in functional explanation, viz., that it explains a capacity that a system has in terms of the capacities of its parts. To ascribe a function to an item i which is part of a system S (that is, to say that item i functions as . . .) is to ascribe to it some capacity in virtue of which it contributes to the capacities of the whole system S. So, functional explanations explain how a system can perform (that is, has the capacity to perform) a certain complex task by reference to the capacities of the parts of the system to perform a series of subtasks that add up to the system's capacity. For instance, the capacity of an organism to circulate blood is functionally explained by reference to the capacities of certain parts of this organism, viz., the capacity of the blood to carry oxygen, the capacity of the heart to pump the blood, the capacity of the valves to direct the blood from the lungs to the organs etc. We may then say that the heart has the *function*, within the organism, to circulate the blood, (by virtue of its capacity to pump the blood), but we do not thereby explain the presence of the heart in the organism, as the standard conception of functional explanation would have it (cf. [1975, 762]). Cummins sums up his position thus:

> To ascribe a function to something is to ascribe a capacity to it which is singled out by its role in an analysis of some capacity of a containing system. When a capacity of a containing system is appropriately explained by analysing it into a number of other capacities whose programmed exercise yields a manifestation of the analysed capacity, the analysing capacities emerge as functions. [1975, 765]

Cummins's view can be called the *causal role* theory of functional explanation. On this theory functional explanation does not answer why-is-it-there questions at all, but how-does-it-work questions.

This is not, of course, the end of the story for conceptions of functional explanation. The debate as to how exactly functional explanation should be understood, especially in connection with biological explanation, is pretty much alive today.

20 A CONCLUDING THOUGHT

In light of the preceding discussion, which has barely scratched the surface of our thinking about explanation, it should be obvious that there is no consensus on what explanation is. Perhaps, the very task of explaining explanation, if by that we mean the advancement of a single and unified account of what explanation is, is futile and ill-conceived. Perhaps, *explanation* is a loose concept that applies to many things; it is such that it can be partially captured by different models and accounts. Perhaps, the only way to understand explanation is to embed it within a framework of kindred concepts and try to unravel their interconnections. Indeed, the concepts of *causation*, *laws of nature* and *explanation* form a very tight web. As it should be evident by now, hardly any progress can be made in any of those, without relying on, and offering accounts of, some of the others. All we may then hope for is some enlightening accounts of the threads of the web formed by these concepts.

ACKNOWLEDGEMENTS

Many thanks to Theo Kuipers for useful comments. Research for this paper was funded by the framework EPEAEK II in the programme Pythagoras II.

BIBLIOGRAPHY

[Aristotle, 1984] Aristotle. *The Complete Works of Aristotle*. Jonathan Barnes (ed.), 2 vols., Princeton N.J.: Princeton University Press, 1984.

[Armstrong, 1983] D. M. Armstrong. *What Is a Law of Nature?* Cambridge: Cambridge University Press, 1983.

[Bromberger, 1966] S. Bromberger. Why-questions. In R. G. Colodny (ed.), *Mind and Cosmos: Essays in Contemporary Philosophy of Science*. Pittsburgh: Pittsburgh University Press, 1966.

[Carnap, 1928] R. Carnap. *The Logical Structure of the World*. Berkeley: University of California Press, 1928.

[Carnap, 1974] R. Carnap. *An Introduction to the Philosophy of Science*. New York: Basic Books, 1974.

[Cartwright, 1983] N. Cartwright. *How the Laws of Physics Lie*. Oxford: Clarendon Press, 1983.

[Collingwood, 1946] R. G. Collingwood. *The Idea of History*. Oxford: Clarendon Press, 1946.

[Cummins, 1975] R. Cummins. Functional analysis. *The Journal of Philosophy*, 72: 741–765, 1975.

[Davidson, 1967] D. Davidson. Causal Relations", *The Journal of Philosophy* **64**: 691-703, 1967.

[Davidson, 1963] D. Davidson. Actions, reasons and causes. *The Journal of Philosophy*, 60, 1963 . Reprinted in *Essays on Actions and Events*, Oxford: Oxfrod University Press, 1980.

[Descartes, 1644] R. Descartes. *Principles of Philosophy* (1644). In *The Philosophical Writings of Descartes*, Vol. 1. Translated by J. Cottingham, R. Stoothoff, and D. Murdoch, Cambridge and New York: Cambridge University Press, 1985.

[Dowe, 2000] P. Dowe. *Physical Causation*. Cambridge: Cambridge University Press, 2000.

[Dretske, 1977] F. I. Dretske. Laws of nature. *Philosophy of Science*, 44: 248–68, 1977.

[Friedman, 1974] M. Friedman. Explanation and scientific understanding. *Journal of Philosophy*, 71: 5–19, 1974.

[Harré and Madden, 1975] R. Harré and E. H. Madden. *Causal Powers: A Theory of Natural Necessity*. Oxford: Basil Blackwell, 1975.

[Hempel, 1942] C. G. Hempel. The function of general laws in history. *The Journal of Philosophy*, 39: 35–48, 1942.

[Hempel, 1959] C. G. Hempel. The logic of functional analysis. In L. Gross (ed.), *Symposium on Sociological Theory*. New York: Harper and Row Publishers, 1959.

[Hempel, 1965] C. G. Hempel. *Aspects of Scientific Explanation*. New York: The Free Press, 1965.

[Hume, 1739] D. Hume. *A Treatise of Human Nature* (1739). L. A. Selby-Bigge and P. H. Nidditch (eds.), Oxford: Clarendon Press, 1978.

[Hume, 1740] D. Hume. *An Abstract of A Treatise of Human Nature* (1740). L. A. Selby-Bigge and P. H. Nidditch (eds.), Oxford: Clarendon Press, 1978.

[Kant, 1781] I. Kant. *Critique of Pure Reason* (1781). N. Kemp Smith (trans.), New York: St Martin's Press, 1965.

[Kant, 1786] I. Kant. *Metaphysical Foundations of Natural Science* (1786). J. Ellington (trans.), Indianapolis and New York: The Bobbs-Merrill Company, INC, 1970.

[Kitcher, 1976] P. Kitcher. Explanation, conjunction and unification. *The Journal of Philosophy*, 73: 207–12, 1976.

[Kitcher, 1981] P. Kitcher. Explanatory unification. *Philosophy of Science*, 48: 251–81, 1981.

[Kitcher, 1985] P. Kitcher. Two approaches to explanation. *The Journal of Philosophy*, 82: 632–9, 1985.

[Kitcher, 1986] P. Kitcher. Projecting the order of nature. In R. E. Butts (ed.), *Kant's Philosophy of Science*. Dordrecht: D Reidel Publishing Company, pages 201–35, 1986.

[Kitcher, 1989] P. Kitcher. Explanatory unification and causal structure. *Minnesota Studies in the Philosophy of Science*, 13, Minneapolis: University of Minnesota Press, pages 410–505, 1989.

[Kripke, 1972] S. Kripke. *Naming and Necessity*. Oxford: Blackwell, 1972.

[Leibniz, 1686] G. Leibniz. *Discourse on Metaphysics* (1686). In *Discourse on Metaphysics, Correspondence with Arnauld, Monadology*, G. Montgomery (trans.). The Open Court Publishing Company, 1973.

[Leibniz, 1698] G. Leibniz. *Monadology* (1698). In *Discourse on Metaphysics, Correspondence with Arnauld, Monadology*, G. Montgomery (trans.). The Open Court Publishing Company, 1973.

[Leibniz, 1973] G. Leibniz. *Philosophical Writings*. M. Morris and G. H. R. Parkinson (trans.). London: Everyman's Library, 1973.

[Lewis, 1973] D. Lewis. *Counterfactuals*. Cambridge MA: Harvard University Press, 1973.

[Lewis, 1986] D. Lewis. Causal explanation. In his *Philosophical Papers, Vol. II*. Oxford: Oxford University Press, pages 214–40, 1986.

[Mackie, 1977] J. L. Mackie. Dispositions, grounds and causes. *Synthese*, 34: 361–70, 1977.

[Malebranche, 1674–5] N. Malebranche. *The Search After Truth* (1674–5). T. M. Lennon and P. J. Olscamp (trans.). Cambridge and New York: Cambridge University Press, 1997.

[McMullin, 2001] E. McMullin. The impact of Newton's Principia on the philosophy of science. *Philosophy of Science*, 68: 279–310, 2001.

[Mill, 1843] J. S. Mill. *A System of Logic: Ratiocinative and Inductive* (1843). London: Longmans, Green and Co., (8th ed.) 1911.

[Nagel, 1977] E. Nagel. Teleology revisited. *The Journal of Philosophy*, 75: 261–301, 1977.

[Psillos, 2002] S. Psillos. *Causation and Explanation*, Chesham and Montreal: Acumen and McGill-Queens University Press, 2002.

[Railton, 1981] P. Railton. Probability, explanation and information. *Synthese*, 48: 233–56, 1981.

[Ramsey, 1928] F. P. Ramsey. Universals of law and of fact (1928). In D. H. Mellor (ed.), *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*. London: Routledge and Kegan Paul, 1978.

[Reichenbach, 1956] H. Reichenbach. *The Direction of Time*. Berkeley and Los Angeles: University of California Press, 1956.

[Salmon *et al.*, 1971] W. Salmon *et al.*. *Statistical Explanation and Statistical Relevance*. Pittsburgh: University of Pittsburgh Press, 1971.

[Salmon, 1984] W. Salmon. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press, 1984.

[Salmon, 1985] W. Salmon. Conflicting conceptions of scientific explanation. *Journal of Philosophy*, 82: 651–4, 1985.

[Salmon, 1989] W. Salmon. *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press, 1989.

[Schlick, 1932] M. Schlick. Causation in everyday life and in recent science. In H. L. Mudler and B. F. B. De Velde-Schlick (eds.), *Philosophical Papers*, Vol. 2 (1925–1936). Dordrecht, Netherlands: D. Reidel, 1979.

[Scriven, 1962] M. Scriven. Explanations, predictions and laws. *Minnesota Studies in the Philosophy of Science*. 3, Minneapolis: University of Minnesota Press, 1962.

[Shoemaker, 1980] S. Shoemaker. Causality and properties. In P. van Inwagen (ed.), *Time and Change*. D. Reidel, pages 109–35, 1980.

[Suppes, 1984] P. Suppes. *Probabilistic Metaphysics*. Oxford: Blackwell, 1984.

[Thayer, 1953] H. S. Thayer (ed.). *Newton's Philosophy of Nature: Selections from his Writings.* New York and London: Hafner Publishing Company, 1953.

[Tooley, 1977] M. Tooley. The nature of laws. *Canadian Journal of Philosophy*, 7: 667–98, 1977.

[von Wright, 1971] G. H. von Wright. *Explanation and Understanding.* London: Routledge & Kegan Paul, 1971.

[Woodward, 2000] J. Woodward. Explanation and invariance in the special sciences. *The British Journal for the Philosophy of Science*, 51: 197–254, 2000.

[Woodward, 2003] J. Woodward. *Making Things Happen: A Theory of Causal Explanation.* New York: Oxford University Press, 2003.

[Wright, 1973] L. Wright. Functions. *Philosophical Review*, 82: 139–168, 1973.

[Wright, 1976] L. Wright. *Teleological Explanations: An Etiological Analysis of Goals and Functions.* Berkeley & London: University of California Press, 1976.