1 Introduction

A philosophical theory of explanation must solve many problems. It must provide a descriptive account of the explanatory activity of scientists in the different domains of science. It must show how this activity differs from the provision of commonsensical explanations permanently offered in everyday life by ordinary people and how the explanatory practices evolve over time in different social arenas. It must also account for the fact that explanatory activities very often, if not always, take place under conditions of a cognitive division of labour in which different participants to the explanatory enterprise undertake different tasks and assume different roles. In a nutshell, a philosophical theory of explanation must be able to provide an adequate descriptive account of the different facets of the explanatory activity, a task which is much harder than it prima facie appears.

Besides, it must also provide standards for judging the quality of the outcomes of the explanatory activity. Some explanations are better in certain respects than others, and a set of norms is necessary for providing judgements of their quality. Some explanations provided in everyday life and in science can be more accurate, simpler or closer to the truth than alternatives that are on offer. They might be able to provide better understanding of the subject matter that they are supposed to deal with, and they might be much easier to use when one wants to intervene in the world on the basis of them. In short, normative rules of adjudicating between the explanations offered in different social contexts are needed and must be worked out and debated critically.

These diverse aims that a philosophical theory of explanation must accomplish are of both a *descriptive* and a *normative* nature. Stated differently, a philosophical theory of explanation is to provide

solutions to a series of problems, both descriptive and normative, using not only philosophical resources but whatever resources are also available from other disciplines. The aim of this book is to establish the claim that this can be best done if one theorizes in terms of explanatory games rather than focusing on the explication of the concept of explanation. This is one way to defend the position of explanatory pluralism, and possibly the most successful one.

2 The Wrong Question: What Is an Explanation?

Though most of the thoughts and arguments in currency in the modern theory of explanation can be found one way or another in the works of some past philosopher from Aristotle to John Stuart Mill, the great advancements in logic and the explosion of technical sophistication of the philosophers of the twentieth century have led to a more systematic treatment of scientific explanation, giving birth to a whole sub-discipline in the philosophy of science dealing with this issue.[1] Since the classic contribution of Hempel and Oppenheim (1948), the modern theory of explanation has largely come to reflect the virtues and vices of the analytic tradition: precise formulations and carefully exposed arguments on the one hand, but also a passionate insistence on logical aspects at the cost of more substantial aspects on the other. The primary question that the philosophical theory of explanation has tried to answer has been "what an explanation consists in". As an answer to this question, Hempel has famously maintained that to explain a singular event is to show how this event can be expected to happen if one takes into account the laws that govern its occurrence together with its initial conditions. For Hempel, an explanation is a valid deductive argument whose premises include law-like statements and initial conditions and whose conclusion states that the event to be explained did occur. An explanation amounts to a statement of the nomic expectability of the explanandum event, and the concept of explanation is according to Hempel primarily epistemic.

[1] For a competent and precise shorter review starting with Aristotle and including the main models of explanation prevailing in the present discussion, see Psillos (2007). See also the monographs by Salmon (1990), Psillos (2002), Ruben (2012) and Weber *et al.* (2013).

One set of reactions to the so-called received view of scientific explanation[2] has centred around the provision of counterexamples to Hempel's model, ranging from shadows explaining the height of poles[3] to magicians hexing salt[4] and much more. These counterexamples were originally meant to highlight different difficulties of this account having to do with its specific features, that is, that it did not invoke any notion of causality, that it did not problematize sufficiently the pragmatic aspect of explanations and so on. However, all the drawbacks of this account are really symptoms of a more general problem: Hempel's covering-law model of explanation was ultimately designed as a unitary model that was supposed to account for *all and every* kind of explanation provided in the different domains of science.

The more constructive critics of Hempel offered alternative models of explanation showing how they could better account for the cases in which Hempel's model failed. Just to name the most influential ones: the *causal mechanistic model*, which claims that an explanation consists in the identification of mechanisms understood as entities and activities organized such that they are productive of regular changes from start to termination conditions;[5] the *unification model*, which

2 See Salmon (1990, p. 8), who has coined this term.

3 In the literature the vertical flagpole example is usually referred to, but in his seminal paper Bromberger refers to the height of a telephone post to which a taut wire is connected – exemplifying, of course, the same point. See Bromberger (1966, p. 105).

4 See Kybourg (1965) and Salmon (1970).

5 The causal/mechanistic approach to scientific explanation was born mainly as an attempt to repair the two most serious problems of the received view, that is, the problem of causal asymmetries (associated with the famous flagpole counterexample) and the problem of relevance (associated with the famous example of hexing the table salt). Besides Railton (1978, 1981) and Humphreys (1981, 1989), it is Wesley Salmon who has most prominently argued in favour of bringing "cause" back into "because". The straightforward way to remedy the main problems of the Hempel–Oppenheim model is supposed to consist in integrating a theory of causality into the theory of explanation or, in other words, in providing scientific explanations by identifying causes of events and/or processes. Since this approach tries to take account of the explanatory practices in science (mainly physics), it does not only aim at derivations of low-level laws and generalizations from higher-level theories but also at elucidating the mechanisms at work. To explain is, thus, to expose the internal workings, to lay bare the hidden mechanisms, to open the black boxes that nature presents to us. This view makes explanatory knowledge into knowledge of the hidden mechanisms by which nature works. Salmon (1984) has tried to specify the notion of mechanism by

claims that explanations are deductive arguments that provide understanding by fitting the particular facts and events within a general theoretical framework;[6] the *pragmatic account* of explanation, which

pointing to causal processes: according to his theory, those processes (and only those) are causal that are capable of transmitting a mark. Besides, he has endorsed an ontic conception of explanation rejecting the epistemic and modal conceptions. (For a criticism of the ontic conception see Wright (2015).)

The first decade of the new millennium has seen an explosion of work in this direction. On the one hand, the seminal paper of Machamer *et al*. (2000) has provoked further the "Thinking About Mechanisms". On the other hand, further work on causality has been produced, and the accounts of causality have reached a higher level of technical sophistication than any of the accounts in the past (Cartwright 2007). Mechanisms as "entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions" (Machamer *et al*. 2000, p. 3) should be sought in order to be able to explain how a phenomenon comes about or how some significant process works – this is the main message of the mechanistic approach to explanation which comes, of course, in different variations (Glennan 2002, Colombo *et al*. 2015). The search for mechanisms goes hand in hand with three claims: (1) explanations should provide causes (or reasons); (2) explanations should make phenomena intelligible; (3) explanations should exhibit the continuity among the explaining parts (Machamer 2009). Whereas Salmon's causal/mechanical approach was mainly inspired by physics, and his mark-transmission theory of causal processes was tailored to physics (and later in Salmon (1998) and also in Dowe (2000) the conserved quantity-transmission), the modified mechanistic approach has extended its reach to the life sciences, the cognitive sciences and the social sciences. Defenders of this approach in the life sciences claim that mechanistic explanations differ from more traditional, nomological explanations because (a) they are not limited to linguistic representations and logical inference but employ frequently diagrams to characterize mechanisms and simulations to reason about them; (b) the fact that mechanisms involve organized systems of component parts and operations provides direction to both the discovery and testing of mechanistic explanations; and (c) models of mechanisms are developed for specific exemplars and are not represented in terms of universally quantified statements (Wimsatt 1976, p. 671, Bechtel and Abrahamsen 2005, Bechtel 2006, 2011, Darden 2006, Bogen and Machamer 2011). In the cognitive neurosciences, the mechanistic approach points to the fact that explanations in neuroscience describing mechanisms are multilevel and integrate multiple fields (Craver 2007, Bechtel 2008, Harbecke 2010). Finally, in the social sciences, a great number of both philosophers and practising scientists hold the view that social scientific explanations require the discovery of the underlying causal mechanisms that give rise to the outcomes of interest (Hedström and Swedberg 1996, Schmid 2006, Demeulenaere 2011). The search for causal mechanisms is often combined not only with the position of methodological individualism (Elster 2007) but more recently also with the position of methodological localism (Little 2009, Knight 2009). It has also been argued that narrative explanations in historical science are descriptions of epiphenomenal mechanisms (Glennan (2010) and a critical discussion by Currie (2014)).

6 The unification thesis, whose chief proponents are Friedman (1974) and Kitcher (1981, 1985, 1989), holds that scientific understanding increases as we decrease the number of independent assumptions that are required to explain what goes on in the world.

claims that explanation is not a relationship like that of description, that is, a relationship between theory and fact, but rather a three-term relationship, that is, between theory, fact and context;[7] the *manipulationist account* of explanation, which claims, relying on invariant generalizations rather than covering laws, that an explanation primarily answers a "what-if-things-had-been-different question", that is, that an explanation primarily enables us to see what sort of difference it would have made for the explanandum if the factors cited in the explanans had been different in various possible ways;[8] and the *kairetic model*, which claims

It seeks laws and principles of high generality with the aim of constructing a coherent world picture and fitting particular facts within this framework. Besides, it is not committed to the world picture being deterministic since it is perfectly compatible with the position that basic laws can be irreducibly statistic. The thrust of the central argument of this approach is nicely summarized in the following quote from the classic paper of Michael Friedman (1974, p. 15): "I claim that this is the crucial property of scientific theories we are looking for; this is the essence of scientific explanation – science 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." The unification approach, being, of course, different from the received view of Hempel–Oppenheim, still remains somehow close to it. This is mainly in virtue of its insistence on deductivism, as is the case in Kitcher's approach which focuses on the scarcity of patterns of derivation. Unification, according to Kitcher, is reached by deriving descriptions of many types of phenomena using one or a few argument patterns over and over again respecting certain constraints, stringency being the most important one. Some new and important work on the unification approach has been produced which stresses different dimensions of unification (e.g., Schurz and Lambert 1994, Schurz 1999, Bartelborth 2002, Bartelborth 2007, ch. 6, Colombo and Hartmann 2015, Nathan 2015, Petkov 2015, and the criticism of Gijsbers (2007) and for a review Psillos (2002)) and it is characteristic for its relevance that any new theoretical endeavour on scientific explanation feels obliged to take a position vis-à-vis this approach.

7 According to van Fraassen (1980), one of the most prominent defenders of the pragmatic account of explanation, the discussion of explanation went wrong at the very beginning because explanation was conceived of as a relationship like that of a description, that is, a relationship between theory and fact. However, it is really a three-term relationship, that is, between theory, fact and context. Both van Fraassen and Achinstein (1983) claim that explanations are answers to why-questions. Why-questions are essentially contrastive: the question "Why P?" is elliptical for "Why P rather than P′, P′′ . . .?" The same words can thus pose different contrastive why-questions. This account of explanation aims at completing the syntactic and semantic aspects of any exposition of scientific explanation by highlighting the pragmatic side of it. See also Faye (2012, ch. 3) and Faye (2014, pp. 183ff.).

8 The manipulationist approach of Jim Woodward is designed as an alternative to the common view that explanation involves subsumption under laws. According to Woodward (2000, 2003, 2014), whether or not a generalization can be used to explain

that explanation is a matter of finding, by way of using a kairetic criterion, which of the causal influences on a phenomenon are relevant to its occurrence, demanding more specifically that the explanation is not missing parts and that every aspect of the causal story represented by the explanatory model makes a difference to the causal production of the explanandum.[9]

All these models of explanation – even though their designers intended them to be alternatives to Hempel's model – still remain in the same research tradition of producing unitary models in order to capture what is supposedly the main aim of (theoretical) science, explanation.[10] If not explicitly, then at least implicitly, the main philosophical project of the philosophers working in this tradition

has to do with whether it is invariant rather than with whether it is lawful. A generalization is invariant if it is stable or robust, in the sense that it would continue to hold under a relevant class of changes. For example, a generalization can be invariant even if it has exceptions or holds only over a limited spatio-temporal interval. A relationship among some variables (or magnitudes) X and Y is said to be causal if, were one to intervene to change the value of X appropriately, the relationship between X and Y would not change *and* the value of Y would change. In a nutshell, an explanation for Woodward ought to be such that it can be used to answer what he calls a "what-if-things-had-been-different question", that is, the explanation must enable us to see what sort of difference it would have made for the explanandum if the factors cited in the explanans had been different in various possible ways. For a very useful recent discussion of Woodward's account see Franklin-Hall (2014).

9 Strevens (2008) proposes the merging of the causal and the unification approach in what he calls the kairetic account of explanation – inspired by the ancient Greek *kairos*, meaning a crucial moment. His approach is clearly a causal account of explanation which claims to have some of the advantages of the unification approach: explanation consists in difference-making of one or more causal factors, and those explanations that are in principle reductionist and expressed in the vocabulary of physicalism are deemed better explanations. According to this approach, the relations of causal influence are invariably physical, so that all sciences, including the social sciences, must state causal relationships using a physical vocabulary. For a discussion of the kairetic account along with the manipulationist account of Woodward see Jansson (2014).

10 This is also true for those philosophical accounts of explanation which have been developed in parallel or immediately after Hempel's covering-law model, as, for example, the account of Karl R. Popper as formulated in his *Logik der Forschung* (1934) and in "Naturgesetze und theoretische Systeme" (1949) and later in his "The Aim of Science" (1957), and also the account of Richard B. Braithwaite in his *Scientific Explanation. A Study of the Function of Theory, Probability and Law in Science* (1953) and of Ernest Nagel in his *The Structure of Science: Problems in the Logic of Scientific Explanation* (1961).

consists in offering an explication of the concept of explanation or, stated more neutrally, in answering the "What is an explanation?" question. The development of the precise meaning of the concept of scientific explanation occupies centre stage in all those approaches.[11] Nobody can oppose – and I do not either – the famous dictum of John Stuart Mill (1843/1974, p. 464) that "[t]he 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". However, this cannot be but a mission of peripheral importance to a philosophical theory of explanation for the simple reason that the outcome of this endeavour can only be a more precise concept of explanation – to be used in the discourse about the solution of the descriptive and normative problems of scientific explanatory activity making up the core of the philosophical enterprise.[12]

But even if one disagrees about the nature of the philosophical enterprise that a philosophical theory of explanation should launch and follow,[13] a cursory glance at the prevailing scientific practices shows that the unitary models of explanation on offer have only a limited range of application. It is simply a matter of fact, thus, that their resources cannot capture anything but the explanatory activities of *some* areas of (theoretical) science. The claim that each of them

[11] And in nearly all other approaches. See, for example, Halonen and Hintikka (2005, pp. 55ff.) and the discussion of two senses of explanation – the subjectivist and the objectivist – by Bird (2005).

[12] Surprisingly, the only text that endorses this view is Noretta Koertge's "Explanation and Its Problems" (1992, p. 86): "What strikes me as unsatisfactory about the current philosophical discussion of explanation is not its failure to match our intuitions about flagpole shadows or mayors with paresis. Rather it is the paucity of explicit *theories* of explanation – the absence of systematic philosophical generalizations in which the competing explications or models of explanation play a central role. I suggest that we reverse the order of investigation. We should begin by asking what *problems* a good theory about scientific explanation might reasonably be expected to solve. Only then can we begin to sketch such a theory. [. . .] I believe it is only by focusing on the philosophical problem-situation that we can transcend the forty years of explication of *explanation* [. . .]."

[13] For a discussion of the level of generality of theories of explanation, see Nickel (2010).

raises, that is, that it is supposed to accommodate *all and every* scientific activity, is not tenable. I will briefly focus on the case of the social sciences, providing three examples in a very sketchy form that are intended to show that the unitary models of explanation have at best limited application.

3 A Brief Outlook on the Social Sciences

The social sciences constitute a very disparate domain of science replete with debates and all kinds of controversies, and this is not the place to even start reviewing them. I will only select three fields in order to exemplify my argument. Since the three more influential models of explanation are the causal mechanistic model, the unification model and the manipulationist model, I will show that only one of those three models is suited to each case that I will discuss. My aim is not only to show that one philosophical model of explanation fits the respective case but also that the other two do not.

3.1 NEOCLASSICAL MICROECONOMIC THEORY

The most theoretically developed field of the social sciences is probably neoclassical microeconomics: It avails of a well-developed mathematical formulation, has a great range of applications and is indeed the only piece of social scientific knowledge that is offered in a standardized way in every single economics textbook. The standard neoclassical microeconomic theory is based on the theoretical construction of utility maximization. Since the marginalist revolution in the 1870s and the pioneering works of Carl Menger (1871), William Stanley Jevons (1871) and Leon Walras (1874), a theory of price has been devised based on marginal utilities. Alfred Marshall's Principles of Economics (1890/1920) then provided a systematic account of the interplay between demand and supply on product and factor markets. Today, neoclassical microeconomic theory provides a standard axiomatization of the behaviour of households and firms in markets. The general theoretical framework that underlies this neoclassical theory of markets is the rationality