Scientific Explanation
C Mantzavinos, University of Athens, Athens, Greece

Abstract
There are three main approaches to scientific explanation in the philosophical literature. The unification approach claims that science explains by fitting the particular facts and events within a general theoretical framework. The mechanistic approach claims that science explains by identifying mechanisms. According to the manipulationist approach an explanation ought to be such that it can be used to answer a “what-if-things-had-been-different question.” The article examines whether these three approaches are compatible or not in the case of the social sciences, and it concludes by defending explanatory pluralism.

Introduction
The three philosophical approaches to scientific explanation that are currently dominating the discussion are the unification approach, the causal/mechanistic approach, and the manipulationist approach. The unification approach claims that science explains by fitting the particular facts and events within a general theoretical framework. The mechanistic approach claims that science explains by identifying mechanisms understood as entities and activities organized such that they are productive of regular changes from start to termination conditions. Finally, the manipulationist approach claims that explanatory activity consists mainly in answering the “what-if-things-had-been-different question.” I will address the issue of compatibility of these three approaches with a specific focus on the social sciences. Are those philosophical accounts of explanation compatible with each other and if yes, in what sense and how exactly?

The Three Main Current Philosophical Theories of Explanation
The philosophical discussion on the theory of scientific explanation centered for many decades around the epoch-making essay of Carl C Hempel and Paul Oppenheim, ‘Studies in the Logic of Explanation.’ Even when philosophers developed alternative views, the Hempel and Oppenheim (1948) paper served as the default position. At the end of many decades of debate three powerful schools of thought seem to dominate the scheme today: the unification approach, the mechanistic approach, and the manipulationist approach.

The unification thesis, whose chief proponents are Friedman (1974) and Kitcher (1981, 1989), holds that scientific understanding increases as one decreases the number of independent assumptions that are required to explain what goes on in the world. 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 Friedman (1974, 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, it still remains somehow close to it (mainly in virtue of its insistence on deductivism).

Some new and important work on the unification approach that stresses different dimensions of unification has been produced (for a review Psillos, 2002), and it is characteristic for its relevance that any new theoretical endeavor on scientific explanation feels obliged to take a position vis-à-vis this approach. For the purposes of this paper it will be Kitcher’s formulation that will be preferred, which focuses on the scarcity of patterns of derivation. Unification 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.

The causal/mechanical approach to scientific explanation was born mainly as an attempt to repair the two most serious problems of the received view, i.e., the problem of causal asymmetries (associated with the famous flagpole counterexample) and the problem of relevance (associated with the famous example of the hexed table salt). Besides Railton (1978, 1981) and Humphreys (1981, 1989), it is Wesley Salmon who has most prominently argued in favor 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 aims not only 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 pointing to causal processes: According to his
theory those processes (and only those) are causal that are capable of transmitting a mark.

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. 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). Salmon’s causal/mechanical approach was mainly inspired by physics, and his mark-transmission (and later in Salmon (1998) the conserved quantity-transmission) theory of causal processes was tailored to physics, whereas 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 (1) they are not limited to linguistic representations and logical inference, but employ frequently diagrams to characterize mechanisms and simulations to reason about them; (2) the fact that mechanisms involve organized systems of component parts and operations provides direction to both the discovery and testing of mechanistic explanations; and (3) models of mechanisms are developed for specific exemplars and are not represented in terms of universally quantified statements (Bechtle and Abrahamsen, 2005; Bechtle, 2006; Darden, 2006; Bogen and Machamer, 2011).

In the cognitive sciences, the mechanistic approach points to the fact that explanations in neuroscience describing mechanisms are multilevel and integrate multiple fields (Craver, 2007; Bechtle, 2008). Finally, in the social sciences, though the problem of interpretation of meaningful actions remains important (Mantzavinos, 2005, 2012), a great number of both philosophers and practicing 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; Demeulenaere, 2011). The search for causal mechanisms is often combined with the position of methodological individualism (Mantzavinos, 2009), but more recently also with the position of methodological localism (Knight, 2009; Little, 2009). All in all, judging from the recent wealth of publications on the mechanistic approach and how it can be applied on many different domains of science, one would apparently not be very wrong to state that this approach is in the process of becoming the new mainstream in the theory of explanation.

The manipulationist approach of Jim Woodward is the third philosophical account of explanation that has gained prominence in recent years. This approach was designed as an alternative to the common view that explanation involves subsumption under laws. According to Woodward (2000, 2003) whether or not a generalization can be used to explain 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 spatiotemporal 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,” i.e., 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.

Summarizing, it seems that there are three viable accounts in the philosophical theory of explanation: the unification approach, the mechanistic approach, and the manipulationist approach. The question is whether some or all of them are compatible, and if yes, in which sense. The fundamental issue this article is tackling is whether one can build bridges between the different theories of explanation.

**The Question of the Compatibility of the Three Philosophical Theories of Explanation**

The main question, thus, is whether the unification approach, the mechanistic approach, and the manipulationist approach to explanation are compatible, and, if yes, how exactly. Kitcher (1985) has christened the unification approach as a ‘top-down’ approach since scientific explanations are supposed to generally unify the understanding of the world by fitting the particular facts and events within a general theoretical framework – explanations serve to systematize knowledge in the most efficient way. The mechanistic approach is, on the contrary, according to Kitcher’s (1985) terminology, a ‘bottom-up’ approach, since it aims at identifying mechanisms and provides explanatory knowledge by virtue of showing how they work and what outcomes they tend to bring about. Are these two ways of looking at scientific explanations compatible with one another and with the manipulationist approach?

The question of the compatibility of the three approaches must be formulated more sharply. Are the three approaches compatible and in what sense? I want to distinguish between two senses:

**Compatibility in the weak sense:** The three approaches are compatible in that each approach applies in different domains of science.

**Compatibility in the strong sense:** Uses of different approaches in the same domain would lead to the same conclusions regarding explanatory relations.

I only defend the compatibility in the weak sense. Prima facie, at least, the three approaches are not necessarily mutually exclusive – each one offers a reasonable way of construing explanations. They might be taken as representing indeed three different, but compatible, aspects of scientific
explanation. Since scientific understanding is after all a very complex matter, there is every reason to suppose that it has different dimensions. In other words, there is no a priori reason to assume that a single theory of philosophical explanation must necessarily cover all dimensions of explanatory activity. However, this prima facie contention needs to be elaborated and defended.

The core issue that seems to divide the unification approach on the one hand and the mechanistic and manipulationist approaches on the other seems to be the question whether explanations must necessarily include causes, or in other words whether any genuine explanation has to take into consideration the causes involved. The causal/mechanical approach takes this for granted and goes on to specify how exactly causality is tied to explanation. The unification approach, on the other hand, contends that the concept of causal dependence is derivative from that of explanatory dependence. It is from the ‘because’ that one can infer the ‘cause’ and not vice versa. What is distinctive about the unification view is, thus, that it proposes to ground causal claims in claims about explanatory dependency rather than vice versa. This approach accounts for the intuition that appeals to shadows not explaining the heights of towers because shadow heights are causally dependent on tower heights. It suggests that the view of causality in this and similar cases rather stems from an appreciation of the explanatory ordering of beliefs. Put more simply, one can only identify something as a cause because this something provides the explanation of the phenomenon at hand – the concept of causal dependence is derivative from that of explanatory dependence. Adopting a position and making a decision in favor of the primacy of causal or explanatory dependence goes hand in hand with adopting the respective metaphysical commitments and coming to grips with Hume’s legacy that causal judgments are epistemologically problematic.

This difference in the approaches seems to be really fundamental since it concerns the thorny metaphysical question about the nature of causality. Although more and more philosophers – surprisingly – hasten to jump on the train of causality in order to reach explanation, cautiousness is necessary. Not all explanation is causal explanation; the most prominent examples of noncausal explanation include mathematical, geometrical, and equilibrium explanation (Sober, 1983). If all these are accepted as genuine cases of explanation, then it is definitely wrong to insist on including causal arguments in every explanatory account. However, this core issue remains, and, in my judgment, constitutes the main dividing line between the unificationist approach on the one hand and the mechanistic and manipulationist approaches on the other, and in the end this is the main reason that a real bridge between them is impossible.

However, the question of their compatibility remains open and should be addressed (Strevens, 2008). It is necessary to take a close look at the scientific practices prevailing in different domains of science in order to inquire whether the three approaches are compatible in the sense that one of them best accommodates the scientific practices in some domains of science and the others accommodate best the scientific practices in other domains of science. Should this be the case, then the compatibility of the three approaches would not refer to the theoretical bridges in the philosophical theory of explanation intended to bring the three approaches closer together, but rather to the domains of application of the three approaches – the compatibility in the weak sense could be defended. I will devote more space to this task and mainly focus on the social sciences.

Compatibility of the Three Approaches of Explanation in the Social Sciences

The picture in the social sciences seems to be very disparate. On the one hand, the most theoretically developed field is probably neoclassical microeconomics, for which the unification approach with its insistence on deductivism seems to offer the best account. On the other hand, in sociology and political science one is confronted with a multitude of theoretical approaches, which do not aim at establishing a global theoretical framework but insist instead on highlighting specific, middle-range causal mechanisms as the means of explaining certain aspects of the social world. The mechanistic approach to explanation can clearly better accommodate these cases. Finally, it seems possible to capture a lot of empirical work carried out in the social sciences with the interventionist account of Jim Woodward. This is especially true for econometrics, which provides the example par excellence of the kind of toolbox used in trying to answer a “what-if-things-had-been-different question,” the question that occupies center stage in Woodward’s manipulationist account of explanation. I now will elaborate on my claim that each philosophical account of scientific explanation best captures a different domain of social science.

Neoclassical Microeconomic Theory

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 Leon Walras, Carl Menger, and William Stanley Jevons, a theory of price has been devised based on marginal utilities. Marshall’s Principles of Economics 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 behavior of households and firms in markets. The general theoretical framework that underlies this neoclassical theory of markets is the rationality hypothesis. The hypothesis of utility maximization plays the fundamental role in driving economic research, and it claims to offer a theoretical account covering all those cases where two or more individuals exchange goods under conditions of scarcity. Neoclassical microeconomic theory offers a global theoretical framework for both a partial analysis of a single market and a total analysis of all markets in an economy, using the utility maximization hypothesis and focusing on the properties of economic equilibrium.

As Mäki (2001) and Mäki and Marchionni (2009) correctly point out, the ideal of explanatory unification is prevalent in economics, and more specifically in neoclassical economic theory. One of the great merits of the theory is supposed to consist precisely in its unifying power, i.e., its ability to subsume a great range of economic phenomena under a unique descriptive and explanatory scheme. Kitche

R's (1989, 432) verdict seems to fit exactly to the case of neoclassical
Now, neoclassical microeconomic theory is, and has always been, under serious attack. Alternative research programs have always prevailed. Institutional Economics is the most prominent case at hand (Mantzavinos, 2001; Mantzavinios et al., 2004) (with a few Nobel Prizes awarded to a series of leading figures working in this research program), and more recently Behavioral Economics. This is not the place to proceed to the evaluation of the alternative research programs in economics or to present the battery of critical arguments against neoclassical economics that has been shaped and elaborated on by a series of economists and philosophers. What is important here is the following: There is an opinio communis among critics and defenders of neoclassical economics alike that one, if not the main, virtue of the approach is its power to unify diverse phenomena under a single descriptive and explanatory scheme. It is due to this specific merit of the theory that it is probably the only piece of social science that is offered in a standardized way in every single introductory textbook taught in every single economics department around the globe.

The ideal of unification has been pushed even further by applying the neoclassical toolbox to the study of phenomena other than the traditional economic ones. Buchanan (Buchanan (1975); Buchanan and Tullock (1962); Breman and Buchanan (1985)), Downes (1957), and Olson (1965) – just to name a few protagonists of introducing neoclassical economic theory into the study of politics – have given birth to a whole new discipline, Public Choice (Mueller, 2003), which is a consistent step toward a unification of economic and political phenomena. Besides, the same patterns of derivation of neoclassical economics are used over and over again to explain such diverse social phenomena as crime and family (Becker, 1976; Becker and Posner, 2009), and law and sex (Posner, 2010). It is clear that the unificationist theory of explanation can provide a better account and accommodate better the practices of the scientists working within the neoclassical economic theoretical framework than the alternative philosophical theories of explanation. Taking into consideration that neoclassical economic theory avails of a unique reputation amidst the social sciences – rightly or wrongly so is not under dispute here – the unification approach to explanation seems to be tailored to capture the reasons for its popularity.

Social Mechanisms in Political Science

A great deal of political science, when not of a descriptive or purely historical nature, deals with laying bare the working properties of social mechanisms or opening up the black boxes between independent and dependent variables (e.g., Elster, 2007, ch. 2) by identifying the specific causal patterns at work. In comparative politics in political science, for example, the focus of political scientists is to highlight the mechanisms that exhibit a relative constancy across countries and epochs. A prominent piece of theoretical work that aims at discovering a specific social mechanism that I want to briefly present here is Kalyvas’s (2006) The Logic of Violence in Civil Wars.

Kalyvas differentiates between the broad concept of civil war and the phenomenon of civil war violence. His study is innovative inter alia for not treating violence and war as the same phenomena as has previously been done in the literature, but for focusing on violence within civil wars as a dependent variable. In his model, civil war itself is an exogenous shock onto a society, which can activate invisible networks of grievances among its individuals. His theoretical approach breaks civil war violence down into two basic categories: indiscriminate violence is executed massively, whereas, in contrast, selective violence describes targeted aggression toward individuals identified as enemies with the help of specific local information about their actions. The aim is to provide an explanation of the spatial variance of the dependent variable, i.e., of violence in civil wars – characterized by its barbarism and the fact that victims and executioners used to live together peacefully before the outbreak of the civil war. The independent variable that Kalyvas highlights as the most important is territorial control exercised by opposed fractions. He explains the level of violence in a territory as a function of the degree of control that the warring factions exert over a particular territory.

The mechanism that is at work starts with individual actors, who are assumed to be rational in their activity in the sense of trying to further their self-interest within a given geographical region. Civil wars are very often warred by means of irregular warfare, and despite the frequency and planning that are involved in indiscriminate violence, the warring factions soon discover it to be counterproductive. Armed groups soon realize that selective violence better furthers their interest, but then their principal problem becomes the removal of the uncertainty about who the enemy is and who is neutral in an environment where competitors are able to hide behind civilian population. The solution to this problem requires armed groups to collect specific information about individuals, which in turn gives individual noncombatants incentives to cooperate. Since enemies can be hidden among apparent supporters of a community, the exertion of selective violence requires information about specific people, which is most easily collected from individual noncombatants acting as denouncers; and thus selective violence becomes “a joint process, created by the actions of both political actors and civilians” (Kalyvas, 2006, p. 209, emphasis in original). The probability of denunciations grows in a nonmonotonical way with the level of control that one side has in a given area since differing levels of control provide different incentives to potential civilian informers and thus leads them to different calculations with respect to the benefits of denunciation. The more control one side has in a given geographical territory, the less the denouncers will face risks of retaliation by the other side and thus the higher will be their propensity to collaborate and provide information – very often because of grievances against specific individuals dating back to the peaceful period. Selective violence is the outcome of the interplay of many individual decisions that take place in
certain social contexts defined precisely within the model. The evidence provided stems from regional data of the Greek Civil War, and it accounts for two-thirds of violence variance, and the validity of the model outside the Greek sample is tested through a great range of both historical and anecdotal accounts about civil wars around the world.

The mechanistic model of explanation can clearly best accommodate such pieces of social scientific research. The aim of this kind of scientific endeavor is not to establish a global theoretical framework and provide the unification of a great range of phenomena, but to insist instead on highlighting specific, middle-range causal mechanisms as the means of explaining certain aspects of the social world. The mechanistic approach to explanation is clearly the appropriate philosophical model in those cases.

**Econometrics**

Econometrics originally emerged as a field at the intersection of economics, mathematics, and statistics, aiming at the provision of numerical values for the parameters of economic relationships, for example, elasticities, marginal values, etc. and at the verification or falsification of economic theories. The main insight offered by econometrics is that economic relationships contain a random element, which is typically ignored by economic theory that standardly postulates exact relationships between the various economic magnitudes. Econometrics offers a range of methods for dealing with the random component of economic relationships.

Since the heart of econometrics lies in regression equation models that specify a functional relationship between dependent and independent variables, including also a ‘measurement error,’ econometric techniques have been included over the years in the toolboxes of other disciplines in the social sciences. Regression equations take the following form:

$$Y = B_1X_1 + B_2X_2 + \cdots + B_nX_n + U$$

where $X_1, X_2, \ldots, X_n$ are the independent variables, $B_1, B_2, \ldots, B_n$ are the coefficients, $Y$ is the dependent variable, and $U$ the error term (which makes the model a stochastic one).

Now, such equations cry for a causal interpretation, which can shed light upon the relationship between different variables in a specific setting. This is exactly what the philosophical theory of explanation proposed by Woodward offers: a natural causal interpretation of such regression equations. Each of $X_1, X_2, \ldots, X_n$ are (direct) causes of $Y$, and these causal relationships are understood as holding for each individual in the population of interest "in the sense that for each such individual (the regression equation) characterizes the response of the value of $Y$ possessed by that individual to some range of interventions that change the values of $X_1, \ldots, X_n$ for that individual. On this interpretation, the error term $U$ also has a causal interpretation: it represents the combined influence of all the other causes of $Y$ besides $X_1, \ldots, X_n$ that are not explicitly represented in (the regression equation)" (Woodward, 2003, p. 321).

A great bulk of applied work in economics and the other social sciences consists in establishing the truth of those kinds of quantitative relationships. The core issue is that social scientists very often do not aspire to establish the truth of general laws, but rather to establish the truth of quantitative relationships that have been tested successfully and hold in specific populations and under specific circumstances. Judged from a God’s point of view, those generalizations might be judged insufficient on many grounds; however, scientific knowledge in the social sciences de facto consists of such quantitative relationships that are invariant and hold only in specific settings. Neither the unificationist nor the mechanistic approach to explanation can adequately capture these scientific practices. Woodward’s manipulationist account on the contrary seems to be tailored to them.

Summarizing, the discussion has shown that there definitely is compatibility between the three philosophical approaches to explanation, but this does not allow them to be synthesized into one approach; rather they are compatible in that they can be applied in different and distinct domains of social scientific knowledge. In other words, they are compatible because they refer to different explanatory enterprises that constitute parts of different domains of social sciences and that are all perfectly legitimate. Insofar the weak compatibility thesis is tenable.

**Conclusion: Towards Explanatory Pluralism**

Since it is commonly accepted that philosophy of science as a discipline can only be useful if it takes the scientific practices seriously, and it is clearly the case that in different domains different kinds of scientific practices prevail, trying to establish a monolithic philosophical theory of explanation that is supposedly good for everything cannot be an acceptable strategy. In a nutshell, the goal of a philosophical account of explanation should not be to capture the explanatory relation, but rather to capture the many ways in which explanations are provided in the different domains of science.

The position that emerges then and that I wish to adopt is that of an explanatory pluralism. This is a pluralistic account that allows for different ideal types of explanation, i.e., different exemplary accounts of what an explanation consists of, which are good as means of classification of different types of explanatory activities that are offered in different domains of the (social) sciences. This explanatory pluralism is different from the explanatory ecumenism that Jackson and Pettit (1992) endorse insofar as it is independent of any type of commitment to causality. In other words, I disagree with what Jackson and Pettit call ‘causal fundamentalism,’ i.e., the view that every explanation must provide information on the causal history of what is to be explained (a view also endorsed by Lewis (1986, p. 217) in his famous verdict that "to explain an event is to provide some information about its causal history"). The explanatory pluralism that I propose is a genuinely pluralistic position that makes as few commitments as possible to causality (and other metaphysical issues). It does not preclude that explanations in some domains of science are offered, debated, and criticized in terms of figuring out causal patterns, processes, or mechanisms; it also permits scientific explanations that provide scientific understanding by, for example, unifying phenomena without invoking causes.

The main danger for a defender of such a position is the possibility of sliding into a pure sociological theory of scientific explanation that just provides descriptions of different explanatory practices or, even worse, to make such a position serve a purely apologetic function by providing ex post legitimation.
of any and/or all explanatory scientific practices. However, philosophy of science has a normative function to serve, so that a philosophical theory of explanation that intends to be pluralistic must clearly accommodate normative considerations. Working out the normative dimension of explanatory pluralism is clearly a desideratum. However, any kind of normative considerations cannot be formulated in abstracto, but must be founded on the existing scientific practices. The aim of this article has been the moderate one of establishing the claim that the philosophical theories of explanation currently on offer do not exclude each other but can rather capture some of those practices and thus of reorienting the discussion toward the feasibility of the position of explanatory pluralism.

See also: Causation: Physical, Mental, and Social; Causes and Laws: Philosophical Aspects; Functional Explanation: Philosophical Aspects; Historical Explanation, Theories of: Philosophical Aspects; Methodological Individualism: Philosophical Aspects; Rational Choice Explanation: Philosophical Aspects; Social Science, The Idea of.

Bibliography


