Intentionalistic Explanations in the Social Sciences

JOHN R. SEARLE
University of California, Berkeley

The dispute between the empiricist and interpretivist conceptions of the social sciences is properly conceived not as a matter of reduction or covering laws. Features specific to the social sciences include the following. Explanations of human behavior make reference to intentional causation; social phenomena are permeated with mental components and are self-referential; social science explanations have not been as successful as those in natural science because of their concern with intentional causation, because their explanations must be identical with the propositional content of the mind of the actor, and because a social phenomenon exists only if people believe it exists. Elements of an apparatus necessary to analyze this problematic social ontology are given and include self-referentiality, constitutive rules, collective intentionality, linguistic permeation of the facts, systematic interrelationships among social facts, and primacy of acts over objects.

I.

For over a century now there has been a continuing debate about whether the forms of explanation appropriate to the social sciences are essentially the same as or radically different from those used in the natural sciences. On one side is the empiricist philosophical tradition, ranging at least from John Stuart Mill through the logical positivists. According to this view, the covering law model of explanation appropriate for the natural sciences is equally appropriate for subjects such as history, anthropology, linguistics, economics, and other social sciences. On the other side is the interpretivist or hermeneutic tradition which ranges at least from Dilthey in the nineteenth century through the twentieth-century followers of Wittgenstein. According to this tradition, there are special modes of explanation appropriate to human behavior. In the second tradition, for example, Dilthey claims that a special method which he calls Verstehen (literally, understanding) is essential to the social sciences. And more recently, Charles Taylor (1985) claimed that human beings are unique in that events are meaningful to them in a special way and that any mode of explanation adequate to accounting for human behavior must explain this meaning component.

An unstated but underlying feature of this debate is often the assumption that much larger issues are at stake. There is at least the suggestion that the issue is a version of the dispute between materialism, on one hand, and dualism and idealism, on the other. The positivist tradition insists that the physical world is the only world that there is and that consequently, the materialist models of explanation that are appropriate to physics and chemistry must be appropriate throughout the rest of science, otherwise there would be no explanation at all. In the interpretivist tradition, on the other hand, there is often the implied claim that not everything can be reduced to physics and chemistry, that some special mental (human, social) facts have a different sort of ontology from the ontology of physics and chemistry. This basic metaphysical dispute between materialism and dualism does not always rise to the surface in these discussions, but I think it is one of the underlying motivations for the persistence of the dispute. On one side are those who think with the positivists that if the material world is the only world there is, then the forms of explanation that have worked well in the material world for the hard sciences must work equally well throughout science. On the other side, in the interpretivist tradition, are those who feel that any such materialistic explanation must be too reductionist, that it must leave out the special features of human beings. It must leave out their mental or spiritual character. According to this view, mechanical forms of explanation are inadequate to capture the special features of human life and behavior; and on the most extreme versions of this view, explanations of human behavior are not even causal explanations. In their crudest and most polemical forms, the interpretivists see the empiricists as crass philistines, whereas the empiricists see the interpretivists as muddle-headed sentimentalists.

What are we to make of this dispute? I think we should be very suspicious of the terms in which it is traditionally posed. To begin with, explanations in the natural sciences do not universally employ the covering law form. For example, I recently had an occasion to look at half a dozen standard textbooks on the functioning of the brain. I do not recall a single case of a covering law explanation; indeed, there are
almost no references to "laws" of brain functioning. The operation of
the brain is explained in these books in the way that one might go
about explaining the operation of the internal combustion engine. The
brain is a physical system that performs certain functions, and the
textbooks describe how these functions are performed. The explana-
tions are indeed causal, but they do not appeal to covering laws.

Furthermore, though it does seem to me that there is a genuine
dispute between the empiricist and interpretivist conceptions of the
social sciences, I believe that it is a misconception of the nature of this
dispute to construe it as essentially about reductionism or the mind-
body problem. The question whether the social sciences require a
special mode of explanation distinct from that of the natural sciences
can be stated in a way which makes it independent of the debate
between materialists and dualists. It is possible, for example, to reject
dualism completely but still to think that there are certain special
logical properties of social science explanations of human behavior.
In my own case, for example, though I reject both dualism and
materialism as standardly conceived, I still think there is an interesting
dispute concerning the distinction between social science explana-
tion and natural science explanation. My own view—just to lay all
the cards on the table—is that the world consists entirely of material
particles and systems composed of material particles (the materialists
are right about that) but that the world also contains subjective mental
states that function causally in the production of human and animal
behavior (and this is usually—mistakenly—denied by materialists).
So, in the discussion that follows, I will be rejecting two commonly
accepted assumptions: first, that natural science explanations are
invariably covering law explanations and, second, that there is some
essential connection between the dispute about the nature of explana-
tion in the social sciences and the materialist-dualist dispute on the
mind-body problem.

It seems to me that the interpretivists are right in thinking that the
modes of explanation in the social sciences are in certain respects
logically distinct from the modes of explanation that we have grown
used to in physics and chemistry. My primary aim in this essay is not
to establish that the mode of explanation is different but simply to
describe some of the actual differences that we find and explore the
consequences of these differences. It is important to keep in mind in
this discussion that science progresses not only by the discovery of
new facts but by the discovery of new types of explanation. It is not

surprising that this should be the case because the discovery of new
facts often involves the discovery of new types of causes. Even if we
confine our discussion to causal explanations, the discovery of new
types of causes will result in the discovery of new types of causal
explanation.

II.

I do not have a well-worked-out theory of the nature of the explana-
tion of social phenomena, and consequently, it is not my aim here
to try to provide one. I have here the much more modest aim of
pointing to certain characteristic features of the structure of explana-
tions that we actually use in explaining social phenomena and of
calling attention to some features of the ontology of the social phe-
omena themselves. The basic theses that I wish to argue can be stated
as two separate but logically related points. First, characteristically,
explanations of human behavior, whether of individual behavior or
collective behavior, make reference to intentional causation. I will give
a more precise definition of intentional causation later, but the intu-
itive idea I can state now is that explanations of human behavior are
indeed causal (the empiricists are right about that), but that such
explanations, whether of individual behavior or collective behavior,
make reference to a special kind of causation—they require a certain
form of mental, or as I like to call it, "intentional" causation. Intention-
tal mental states such as desires and intentions represent certain
sorts of states of affairs, and intentional causation is that form of
causation by which mental states are causally related to the very states
of affairs that they represent. Thus, for example, a man whose thirst
causes him to drink is engaging in behavior that can only be explained
by intentional causation because his desire to drink causes the very
behavior that is represented in the content of the desire, namely, his
drinking. What is true of this very simple case is true on a much
grander scale of the explanation of wars, revolutions, social move-
ments, economic phenomena, and so on.

The second point I want to make is that there is a class of social facts
having certain logical features that make them quite unlike the phe-
nomena of physics and chemistry. In the case of phenomena such as
marriage, money, divorce, elections, buying and selling, hiring and
firing, wars and revolutions, the phenomena are—to speak vaguely
at this stage—permeated with mental components; furthermore, the facts in question are self-referential in an odd way because they can only be the facts they are if the people involved think that they are those facts.

I believe the best way to explore the differences between natural science explanation and social science explanation is to list the most striking differences between the basic features of the two types of science. I want to list obvious features, so that at this stage of the argument at least, what I say will seem uncontroversial.

The first and most obvious feature is simply that we have not had the sorts of success in the social sciences that we have had in the physical sciences. We do not have anything in sociology or history to compare with the rich theoretical apparatus that we have developed, for example, in physics. Indeed, there is a sense in which, for the most part, the social sciences have not advanced theoretically beyond a kind of systematized common sense. And the mode of explanation that is appropriate in the social sciences seems to employ categories that we are all familiar with from our ordinary pretheoretical experience. One of the great puzzling features of contemporary intellectual life, as the twentieth century comes to a close, is why the methods of the natural sciences have not produced in the social sciences results comparable to those that they have produced in such subjects as physics and chemistry.

It might seem that economics is an exception to this, since it is a formalized mathematical discipline, full of technical terms such as marginal propensity to consume, the multiplier effect, and unstable equilibrium. I believe that economics is not a counterexample at all. Typically what the economic theorist does is to take a lot of commonsense assumptions, such as the fact that consumers are trying to be better off and business people are trying to make money, and then idealize these assumptions (some people would say these assumptions are overidealized in economic theory) and work out the systematic implications. Thus the old chestnut in microeconomics that the rational entrepreneur sells where marginal cost equals marginal revenue is, in fact, a strict logical consequence of certain commonsense assumptions about "rational" economic behavior.

A second feature of the social sciences is one mentioned earlier, and it is, I believe, the most important: namely, that of intentional causation as one form of explanation in the social sciences. Intentional causation differs in an important respect from the sorts of causal phenomena that we are familiar with when we discuss such things as gravitation or nuclear forces, for intentional causation is that form of causation involving intentional mental states in virtue of the actual content of the mental states. Intentional causation is ascribed in statements such as the following:

General Robert E. Lee ordered an attack at Gettysburg because he feared that his men would become demoralized if they were told to retreat.

or

Many Democrats voted for the Republican candidate in the 1988 presidential election because they did not want a rise in taxes and they believed that Dukakis would raise the income tax.

Intentional causation is a reasonably precise notion and can be defined as follows: For any events x and y, x and y are related by intentional causation if and only if either x causes y or y causes x, and either x or y is an intentional state and the term which is not an intentional state is the conditions of satisfaction or part of the conditions of satisfaction of the intentional state (for a detailed discussion of this issue, see Searle 1983, chap. 5).

Intuitively, the idea is that mental states are a type of mental representation, and the special feature of intentional causation is that the intentional content causes the very state of affairs that it represents or is caused by the very state of affairs that it represents. In the case of General Lee, the desire to give an order to attack caused the giving of an order to attack, and ultimately, the desire to attack caused the attack. Furthermore, the desire to attack was itself caused by the desire not to demoralize the troops by ordering a retreat, together with the belief that anything other than an order to attack would cause such demoralization. In the case of the Democrats who voted for the Republicans, their desire to vote against increased taxes caused a desire to vote against the Democratic candidate, and that caused them to vote against the Democratic candidate.

A third distinction between the social and natural sciences is a direct consequence of the nature of intentional causation. The propositional content given by the theorists in the explanation of the behavior must be identical with the propositional content in the actual mind of the agent or agents whose behavior is being explained; otherwise, the behavior is not properly explained. For example, if we explain General Lee's behavior in terms of his fears and we explain the behavior of the Democrats in terms of their desires, in both cases the
explanation will be valid only if the actual content that we specify as the content of the fear or the desire is identical with the content in the mind of the agent who had the fear or the desire. Since in the case of intentional causation, the causation works only in virtue of the representational content in the mind of the agent, then any representation of intentional causation will simply be a representation of a representation, and to the extent that it is accurate, there must be an identity between the propositional content given by the theorists and the propositional content in the mind of the person or persons being theorized about. That feature is totally unlike explanations in the natural sciences. In the natural sciences, we simply cite those aspects of events under which they are causally related. So, we cite such phenomena as gravity, pressure, heat, and so on. But since the phenomena in question are not themselves mental, there is no question of there being an identity between the phenomena to be explained and the phenomena occurring in the explanation. I believe that it is this feature of social explanation which accounts for the fact noted by several authors (e.g., Winch 1958) that the terms used in the explanation of human behavior must be available to the agents whose behavior is being explained. We could not, for example, explain Lee’s behavior in terms of a fear of communism because as far as we know Lee never heard of communism.

It was a great breakthrough in the development of natural science when the medieval conception of all explanation as essentially intentionalistic was replaced in the seventeenth century by explanations appealing to general nonintentionalistic laws of nature. This was a tremendous advance, but the fact of that advance has blinded us to the fact that there are genuine empirical phenomena to which this covering law model of explanation is simply inadequate. It is inadequate because covering laws of explanation do not have this special feature of intentionalistic explanation. The special feature in question is that the content of the causal explanation and the mental content in the mind of the agent whose behavior is being explained must be identical because if the explanation is really valid, that very content must be functioning causally in the mind of the agent. To put this in brutally simple terms: When we explain human behavior, we are trying to explain people’s behavior in terms of what they want and what they do to try to get what they want, given what they believe. But terms like “want,” “try to get,” and “believe” are intentionalistic notions. We can only specify the particular wants, tryings, and believ-
in the sense that, for example, people are only playing a game if they believe that they are playing a game. Similarly, it is only a case of winning if people agree that it is the case of winning. Suppose, to take another example, I give a very large party, and at this party a large number of people are killed or injured. Suppose that the casualty rate is comparable to that of the battle of Gettysburg. All the same, as long as people think it is a party and not that it is a battle, then it is a party and not a battle, regardless of the casualty rate. Now, this sounds paradoxical for the following reason: In general, when someone thinks that x is f, if what one thinks is true, then there must be something completely independent of the fact that one thinks that x is f, that makes the case that x is f. But many social concepts are partly self-referential in this respect; part of what constitutes x’s being f is that people think that x is f. This is paradoxical because our notion of truth requires a distinction between the representation of the fact and the fact represented. But in these cases, the representation of the fact is partly constitutive of the fact represented, or at least this is the case for the participants in the fact. But how can our beliefs about social facts be true beliefs if the facts are partly constituted by the fact that we have those beliefs? The paradox dissolves when we see that thinking x is f in the case of social facts is only part of the constitution of the fact that x is f, and the concept expressed by f in all of these cases is simply a way of clustering a whole family of social practices. For social concepts, thinking that x is f involves thinking that whole patterns of behavior and social relationships are appropriate to the phenomenon in question. So, thinking that something is money or property or a party is not just a matter of thinking that certain labels apply but, rather, thinking that a set of attitudes and behavior are appropriate to the situation in its social context. But then, thinking that those attitudes are appropriate is itself partly constitutive of that social situation.

I believe that we will not begin to get a deep analysis of the foundations of the social sciences until we have a more thorough explanation than we now have of social facts. By an explanation of social facts, I mean an explanation of the ontology of social facts. The phenomena here that we are trying to analyze are so obvious and pervasive that it is almost impossible for us to see them. It is obvious to me that the piece of paper in my pocket is a dollar bill; it is also obvious to me that I saw a football game last Sunday. What is not obvious is what facts exactly about this piece of paper make it the case that it is a dollar bill and what facts exactly about the movements that I saw make it the case that some of the organisms involved were playing a game of football. Notice that it is not enough to say that what constitutes the movements being a game of football is that they are in accord with the rules of football because exactly the same physical movements might have been made as part of a dance, outdoor exercise, or religious ceremony. Furthermore, it is not enough to make the obvious point that the fact that this piece of paper is a dollar bill derives from the fact that it was printed as such by the United States Bureau of Engraving and Printing because that only forces the question back a step. What fact makes something the Bureau of Engraving and Printing, and why do we recognize this Bureau as in any way relevant to whether something is a dollar bill? What I intend to do now is briefly discuss the basic elements of the apparatus necessary to give us an analysis of the ontology of social facts.

The first element is self-referentiality. Social facts differ from natural facts in that they contain mental representations. But they differ from other sorts of mental facts in that the mental representations have the element of self-referentiality that I was just attempting to adumbrate. The thing is what it is only if people think that that is what it is.

Constitutive rules comprise the second element. In Speech Acts (Searle 1969), I made a distinction between constitutive rules and regulative rules. Regulative rules regulate antecedently existing forms of behavior. Constitutive rules not only regulate but create the very possibility of new forms of behavior. The rule that says “Drive on the right side of the road” regulates the activity of driving, but it does not create the very possibility of driving. It is possible to drive on the left-hand side of the road or in the middle of the road as much as it is to drive on the right-hand side. That rule is regulative. But the rules of chess do not in that way regulate the antecedently existing activity of pushing pieces around the chessboard. Rather, acting in accordance with at least a large subset of those rules is part of what constitutes playing chess. The rules do not just regulate but are constitutive of a new form of activity. I believe that similar remarks could be made about the constitutive rules of marriage, money, private property, football, and so on. I am not at all sure that every social fact requires systems of constitutive rules, but it seems to me a very large class of social facts only exists within such systems.

The third element is collective intentionality. Social facts require social behavior, and social behavior characteristically requires collec-
tive intentionality. Collective intentionality is manifested in cooperative forms of behavior, which we characteristically describe by saying it is not just the case that I am doing something and you are doing something but that we are doing something together. It is worth pointing out that most forms of conflict require collective intentionality. In order that two men should engage in a prizefight, for example, there has to be collective intentionality at a higher level. They have to be cooperating in having a fight in order for each of them to try to beat the other up. In this respect, prizefighting differs from simply mugging someone in an alley. The man who creeps up behind another man in an alley and hits him on the head is not engaging in collective behavior. But the man who sticks up a bank by pointing a gun at a teller and demanding money is trying to elicit collective intentionality. Collective intentionality, I believe, is a universal feature of social facts (for a theoretical account of the structure of collective intentionality, see Searle forthcoming).

Linguistic permutation of the facts is the fourth element. I do not believe that it is possible to have social facts without language. It is possible to have collective behavior without language, and indeed, many forms of animal behavior are precisely collective forms of behavior, expressing collective intentionality. But such forms of behavior do not yet constitute what I am calling social facts. No doubt, there is a continuum between collective behavior in the form of collective animal intentionality and full-blown human social facts, such as electing someone president of the United States. But although there is a continuum, still I believe there is a deep reason why full-blown social facts must be linguistic. This feature derives from the self-referential character of the concepts mentioned earlier here. Money is only money if people think that it is money; a game is only a game if people think that it is a game. But it is impossible for us to have these thoughts without a certain sort of vocabulary. It is not necessary to have the actual word money or some synonym of it, but there has to be a vocabulary appropriate to buying, selling, and exchange generally for us to be able to manifest the collective intentionality which invokes the constitutive rules of private property and the exchange of property in return for money. I do not fully understand this feature, but the hypothesis that I am suggesting might be expressed as follows: There are no social facts without language because it is characteristic of a social fact that in order to be the social fact that it is, it has to be regarded as that fact, but for it to be regarded as that fact there has to be some mechanism for expressing that regard. For human beings, that mechanism is essentially linguistic.

The fifth element comprises systematic interrelationships among social facts. Social facts cannot exist in isolation but only in a set of systematic relations to other social facts. Thus, for example, in order that anyone in a society could have money, that society must have a system of exchange, of exchanging goods and services for money, but to have a system of exchange, it must have a system of property and property ownership. Similarly, in order that societies should have marriages, they must have some form of contractual relationships, but to have contractual relationships, they have to understand promising and obligations.

It might seem that games are a counterexample to this general principle because, of course, games are designed precisely with the idea that they should be forms of activity that do not connect with the rest of our lives in a way that social facts characteristically do. Today's baseball game need have no consequences for tomorrow in the way that today's wars, revolutions, buying, and selling are intended precisely to have consequences for tomorrow and into the indefinite future. Nonetheless, even in the case of games, there are systematic dependencies on other forms of social facts. The position of the pitcher, the catcher, and the batter, for example, all involve rights and responsibilities, and their positions are unintelligible without an understanding of these rights and responsibilities, but these notions are, in turn, unintelligible without the general notion of rights and responsibilities.

The primacy of acts over objects is the sixth element. It is tempting to think of social objects as independently existing entities on analogy with the objects studied by the natural sciences. It is tempting to think that a government or a dollar bill or a contract is an entity in the sense that a DNA molecule or a techoptic plate or a planet is an entity, an object which admits of study as an object. In the case of social objects, however, the grammar of the noun phrases conceals from us the fact that, in such cases, process is prior to product. Social objects are always created by social acts, and, in a sense, the object is just the continuous possibility of the action. It is a consequence of the account that I have given thus far that social acts are prior to social objects. That is, it is a consequence of the facts that social concepts have the self-referential feature mentioned earlier here and that collective intentionality is essential to the constitution of social facts, that social objects are
themselves the product of collective intentionality manifested in the self-referential way that I have tried to characterize. Thus the object in my wallet is indeed a dollar bill and it is indeed an object; however, its status as a dollar bill—as opposed to its status as a piece of paper with ink on it—is constituted by its ability to function in a series of activities: buying, selling, paying bills, and so on. Furthermore, the self-referential concepts that I mentioned earlier serve to focus activities that people engage in in virtue of collective intentionality.

CONCLUSION

If social explanation has logical features different from explanations in the natural sciences, then it must be because social phenomena have factual features that are logically different from the facts of the natural sciences. I believe that such is indeed the case and I have tried to identify two sorts of facts involved: first, that the form of causation is essentially intentional causation and, second that social facts have a logical structure different from natural facts.

REFERENCES


The Problem of Other Cultures and Other Periods in Action Explanations

REx MARTIN
University of Kansas

This essay develops a general account of one type of explanation found in history in particular: that an individual action is conceived as an exemplification of a rather complex schema of practical inference, under the provision that the facts which instantiate the various terms of the schema have an intelligible connection to one another. The essay then raises the question whether historians, anthropologists, and their contemporaneous audience can have an internal understanding of the actions of others, where those others come from radically different cultures or times from the historians or anthropologists. An account is offered that, arguably, can resolve this problem and do justice to both the claim of internal understanding and the presumed cultural difference between the agents studied and the historians and anthropologists who do the study.

I. INTENTIONALIST EXPLANATIONS

Preliminary Remarks

One of the standard kinds of explanation is that in which an action of an agent is accounted for by reference to certain thoughts—motivations and beliefs—of the agent. We often call these the agent’s “reasons” for action.

I want in this section to give a brief account of such explanations, namely, intentionalist explanations. We can start with the claim that actions typically occur in a context of states of affairs and that the agent’s thoughts about any given state often help provide a motivation for action there. We also believe that the agent intends to bring about something with the action and that this something—this end to be achieved or relevant purpose—will resolve or help resolve the original situation that motivated the action in the first place. The action, in its turn, is a means to that end or part of accomplishing it.

I am much indebted to Allan Hanson and Jack Brice for helpful comments on an earlier draft of this essay.

Philosophy of the Social Sciences, Vol. 21 No. 3, September 1991 345-366
© 1991 York University, Toronto, and Contributors.