Causal Explanations of Behavior*

Merrilee H. Salmon†‡

Most discussions of causal explanations of behavior focus on the problem of whether it makes sense to regard reasons as causes of human behavior, whether there can be laws connecting reasons with behavior, and the like. This essay discusses explanations of human behavior that do not appeal to reasons. Such explanations can be found in several areas of the social sciences. Moreover, these explanations are both causal and non-reductionist. Historical linguists, for example, offer causal explanations of changes in how words are pronounced—and linguistic change in general—without appealing to human intentions. I use examples from linguistics, anthropology, and evolutionary psychology to discuss the importance of this sort of explanation and to examine its compatibility with recent philosophical accounts of causation.

1. Introduction. One of the most vexing problems in the philosophy of the social sciences is how to understand causal explanations of human behavior. Attributions of causes to behavior are common, and while accuracy in individual cases can be questioned, these causal explanations seem otherwise unproblematic—except to philosophers. For example:

1. Despair over financial losses caused a number of suicides after the stock market collapsed in 1929.
2. The soldiers fired as the van approached a checkpoint because they believed the driver was a suicide bomber.
3. Because Electra feared Orestes would not return, she asked her sister to help her kill their mother.

*Received February 2001; revised June 2003.

†To contact the author write to Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA 15260; e-mail: mhsalmon@pitt.edu.

‡An earlier version of this work was presented at a conference on varieties of scientific explanation, with a focus on the contributions of Wesley C. Salmon, held at Universidade da Coruña. That paper was published under the title "La explicacion causal en las Ciencias Sociales" in *Diversidad de la Explicacion Cientifica,* Barcelona: Editorial Ariel, S.A. I am grateful to Wenceslao J. Gonzalez, who organized the conference and edited the volume. I also want to thank this journal's anonymous referee for very useful comments.

Philosophy of Science, 70 (October 2003) pp. 720–738. 0031-8248/2003/7004-0004$10.00
Copyright 2003 by the Philosophy of Science Association. All rights reserved.

Few doubt that such causal accounts are roughly correct, even if simplistic or incomplete. Nevertheless, we can ask whether the causal relationships in the given examples can be understood in the same manner as in the following:

4. Lightening strikes cause many forest fires.
5. Frequent hand washing prevents many cases of infection.

In the first three examples but not the latter two, reasons (human beliefs or desires) are invoked as causes. Are reasons really causes or is causal talk merely metaphoric when reasons are involved? Whether or in what sense reasons are causes is an old philosophical problem that exposes deep disagreements about the nature of causation, especially psychophysical causation, and related issues of free will, holism, individualism, explanation and prediction. Opposing positions were set out clearly in the nineteenth century, along lines that continue to dominate contemporary discussions. These days, however, we are better equipped to deal more incisively with the issues because of gains in philosophical clarification as well as new scientific understanding. Before turning to recent work, a brief review of the old battle lines is in order.

On one side are Naturalists, who say that reasons can be causes in the straightforward sense. Some, like John Stuart Mill (1874) describe events as causally related whenever a regular connection exists between those types of events. Whether the events are ''mental'' or ''physical'' is irrelevant. Empirical observation and historical studies, Mill says, support the existence of regular connections between individual reasons (beliefs and desires) and behavior. For Mill the logic of human sciences begins with these low-level regularities. At the next stage of the science, Mill argues that additional regularities connect various types of human character with specific behavior patterns. Ultimately, underlying these mid-level laws of character, fundamental psychological regularities govern how the human mind organizes experience. Mill admits that despite the causal link between reasons and behavior, predictions of behavior are not always reliable. He attributes poor predictions to the multiplicity of interacting and counteracting causal forces, along with shifting or poorly grasped initial conditions—the same problems that affect some physical sciences, particularly meteorology. In the twentieth century, C.G. Hempel's (1965) work on explanations of human behavior shows many similarities to Mill's views inasmuch as regular connections between character and behavior, and reasons and behavior, are the basis for suitable covering laws for explanation.

Emile Durkheim (1897) also accepts the causal force of individual reasons, but insists that not all human behavior is amenable to this form of explanation. Durkheim, with a better grasp of statistical reasoning than Mill, maintained that *social facts,* such as the rate of suicide in a given

society, cannot be explained solely by individual reasons but require *social causes* for their explanation. By social causes, Durkheim meant mechanisms that operate at the social rather than individual level. In a famous study (Durkheim [1897] 1951), he used statistical data to try to discover causes for varying rates of suicide in different societies. He claimed to find the relevant mechanism in varying degrees of social integration. Although Durkheim worked to identify the mechanisms of social causation, he believed that the explanatory force of social causes was independent of such understanding. In the absence of plausible mechanisms, he referred to a ''collective impulse'' that was not merely a summary of various individual intentions. One example of a social fact that could not be accounted for with any known single mechanism was the surprisingly regular annual rate of "dead" letters in France. In a case such as this, Durkheim attributed causal force to the collective impulse. Surprisingly accurate predictions are possible, and—according to Durkheim—so is a weak sort of (social) causal explanation, even though the exact mechanism is unknown.

Opposed to this Naturalist tradition represented by Mill, Hempel, and Durkheim are Interpretivists, who at least from Wilhelm Dilthey (1833–1911) on, insist that the relationship between reasons and behavior is not causal, at least not in the same sense of "causal" that operates in the physical world. Some Interpretivists, such as R.G. Collingwood (1946) and Peter Winch (1958), claim that since reasons are *logically* connected with behavior, which is to say, reasons give behavior its meaning, and make it the sort of behavior it is, they cannot cause behavior. Interpretivists argue that the logical connection between intentions and behavior is a "matter of reason" quite different from any "matter of fact" (a contingent causal connection between cause and effect). Thus, following Hume, they maintain that ascribing reasons as causes of human behavior involves a category mistake. Other Interpretivists, such as the anthropologist Clifford Geertz (1973), also reject causal explanation of behavior, but not on the logical grounds cited by Winch. Geertz says the complexity and perversity of human beliefs and desires and the human capacity of free choice prevent us from ever finding nontrivial laws that could let us predict behavior accurately. Because Geertz implicitly accepts the symmetry of prediction and causal explanation, and recognizes the unreliability of predictions, he does not believe that causal explanation of behavior is possible.

Naturalists take comfort from D. Davidson's (1980) argument that Interpretivists such as Winch have failed to distinguish causal relationships as they exist in the world from descriptions of these relationships, some of which are framed in analytic sentences. There are limits to this comfort, however, for Davidson argues also that although reasons do cause behavior, there can be no psychophysical laws connecting the two. This,

he says, is because descriptions of reasons neither express nor can be translated into descriptions of the physical phenomena (presumably, brain states) that are the genuine causes of behavior. I will not discuss the merits of Davidson's argument here, or criticisms of it (but see McIntyre 1999), because I want instead to sidestep this conflict between Naturalists and Interpretivists by focusing on causal explanations of human behavior that do not appeal to reasons as causes.

To continue the discussion along these lines, we first need to specify the meaning of ''human behavior''. Interpretivists and Naturalists agree that the concern of the so-called social sciences is human *action,* as opposed to mere physical processes of the human body. The latter include digestive processes, normal breathing and other such processes that are more or less automatic and not normally under conscious control. In many cases, but not all, the difference between actions and bodily processes is clear. Deliberately holding one's breath to escape detection is an action. Normal breathing can also be an action—as when one consciously breathes normally as part of a yoga exercise. In most cases, however, normal breathing is merely a physical process. Mill was concerned to explain behavior that resulted from the beliefs and desires of agents, not physiological processes. Collingwood characterizes human *actions* as events whose complete description requires a reference to the thought of the agent.

Murders and making promises are examples of human actions in this sense. Collingwood says that the historian—his inclusive term for practitioners of any branch of the social sciences—is concerned ''with those events that are the outward expression of thought, and is only concerned with these so far as they express thoughts.'' As he says, ''the historian is not interested in the fact that men eat and sleep and make love and thus satisfy their natural appetites; but he is interested in the social customs which they create by their thought as a framework within which these appetites find satisfaction in ways sanctioned by convention and morality'' (Collingwood 1946, 216).

These characterizations of human behavior by the Naturalist Mill and the Interpretivist Collingwood are similar, but both lack precision. For example, is the historian to be concerned only with behavior in which the agent is *consciously* aware of intent when performing the action? Although both Naturalists and Interpretivists focus on consciously performed actions, these actually constitute a relatively small proportion of what we normally refer to as human actions. Conscious awareness—or paying attention to what we are doing—requires considerable effort. Conscious awareness does play a role when we attempt to solve a novel problem or make an important decision that requires us to process new information. Learning a new skill, such as driving a car, requires a great deal of

conscious attention, but once the skill is mastered, its exercise becomes automatic. Much of our social life is similarly automatic. Often, when we are learning new skills, including social skills, we recognize the need to pay attention to what we are doing. Little would be accomplished, however, if we had to give conscious attention to every instance of social behavior, and one goal of socialization is to make good social behavior automatic. This is not to deny that in many cases social life would be improved by increased consciousness of what we say and do.

Perhaps we could say that *actions* are behaviors that at some point required conscious attention, like driving a car, though later they may become automatic. It is difficult, however, to make a case that all our social behavior requires conscious attention at the learning stage. Many psychologists have argued that a lot of learning takes place in the absence of conscious attention (Hefferline et al. 1959, quoted in Gaulin and McBurney 2001, 114). Acquiring the rules of grammar in our native tongue, for example, is probably for the most part an unconscious learning process. Learning the grammar of a foreign language in school is an entirely different matter, and requires considerable attention. Like our native language, much of our other cultural knowledge is also acquired without any conscious attention at the learning stage. One example is knowledge of the appropriate distance to maintain when speaking face to face with another person. The social rule governing this action varies from culture to culture. Most people absorb the rule in the process of socialization, though neither the teaching nor the learning process is conscious. We become aware of this rule only when it is pointed out to us, when someone violates it, or when we are in a culture with a different rule from our own.

The possibility of unconsciously learned behavior is supported in part by the difficulty we have in explaining how we are able to do many of the things we have learned to do, or even realizing that we have learned a social skill. Other support for the claim that learning does not require conscious attention comes from the fact that forms of animal life that apparently lack consciousness nevertheless can learn. Despite its great importance, conscious behavior constitutes a relatively small part of human activity. Thus, it would seem that a theory of explanation of human behavior should not be limited to conscious behavior.

Could we say instead that voluntary, rather than conscious, behavior constitutes human action? Many of our actions that do not meet the criterion for consciousness are voluntary. Philosophers usually mean by *voluntary* actions those actions that the agent could have refrained from doing. For example, experienced drivers follow traffic rules with little or no conscious attention to the fact. Yet, they could refrain from doing so. The concept of voluntary action is closely associated with moral respon-

sibility. In the moral context, voluntary actions are those that are not coerced. We cannot be held liable for actions that were not within our power to avoid or change. While this notion of the voluntary is suitable for reasoning about moral matters, a different characterization of behavior may be more appropriate for social sciences. In the first place, it is odd to characterize some social behavior as voluntary when, as in the case of maintaining distance between speakers, most people are unaware and have never been aware that they are engaged in the behavior. In some cases, awareness of unconscious behavior comes easily when it is pointed out to us, and avoidance of the behavior may not be very difficult. In other cases, however, becoming aware of the behavior requires considerable effort, and perhaps even the aid of a teacher or therapist, and being able to change the behavior may pose a real challenge. Another reason for not restricting ourselves to the study of voluntary behavior is that coerced behavior is a legitimate concern of the social sciences. For example, coerced behavior is of particular interest to those who try to understand and explain cult behavior as well as gross acts of human cruelty.

Contemporary Interpretivists have proposed characterizing the relevant behavior as *meaningful* in the sense of *rule-governed*. Following Wittgenstein's analysis of linguistic behavior as rule-governed, Winch (1958) expands the notion of rules to include social institutions, customs, conventions, and the like, as well as laws and regulations. Just as linguistic activity is both constituted and regulated by the rules of grammar, so according to Winch, is all other social behavior constituted and regulated by social rules. Students of human behavior as diverse as Collingwood, Durkheim, and Max Weber have also emphasized the importance of social rules. Winch's notion of a social rule seems broad enough to cover various types of behavior that social scientists want to explain—conscious actions, voluntary actions, actions that were once consciously performed but have become automatic, and actions that are probably the result of unconscious learning. Even involuntary or coerced behavior can be rule-governed. Here then we seem to have a characterization of the relevant sense of human behavior that would be acceptable to both Naturalists and Interpretivists.

Important features of rule-governed behavior include the following:

1. There are correct and incorrect ways of doing things according to the existing rules;
2. Members of a society frequently negotiate and modify the rules of that society;
3. Rules—unlike natural laws—can be broken.
4. Rules do not link reasons to actions in the same way that natural laws link causally related events.

Even if we agree with the Interpretivists' claim that the link between a person's reasons and her behavior is not causal, we can still investigate the causal aspects of meaningful behavior. It is legitimate, for example to try to find causal explanations of why certain rules are adopted, or why rules change over time, or why some rules are widespread or extraordinarily stable. Such explanations need not appeal to reasons, yet are crucial to understanding human behavior. In what follows, I examine three different attempts to address this issue.

2. Historical Linguistics. Although it seems paradoxical to explain meaningful human behavior without citing reasons, such explanations are common and uncontroversial in the field of historical linguistics. No one could dispute that linguistic activity is meaningful. Equally indisputable is the physical basis of language, whether spoken or written. The physical basis is particularly prominent in the area of phonology, the study of how words are pronounced. Historical linguists have done a considerable amount of work investigating the causes of phonological change.

All living languages change constantly. The discovery of the regularity of sound change was an important empirical accomplishment that allowed linguists to exploit correspondences between sound and meaning. This regularity and the explanations it supports—like most explanations in historical linguistics—make no mention of the intent, motives, or reasons of speakers. This is true even though linguistic behavior is meaningful and rule governed (See M. Salmon 1996).

Although linguistic innovations may be unpredictable, some changes, once they have been introduced into a language, follow regular patterns. These patterns of linguistic change are the object of intensive study, and even form the basis of partial reconstructions of systems of undocumented prehistoric languages, such as Indo-European. Historical linguists offer causal explanations of how living languages arose from such unattested languages. They also offer causal explanations of observed patterns of change in living languages.

Consider linguists' study of changes in the way vowels are sounded. For example, some new ways of pronouncing vowels in particular contexts apparently enter the language because they require less muscular effort. Linguists call this process, in which a sound changes to accommodate neighboring sounds, *assimilation*. Linguists believe that such changes are regular, which means that the new sound, once introduced into the language, replaces the old sound in all similar contexts. In some instances, an expected assimilation may fail to occur, but linguists account for anomalies by showing that they are the result of other regular principles of linguistic change, such as linguistic borrowing or analogical change. Linguists regard the regularity of assimilation not as a proven theory for

reconstructing languages, but as an explanatory hypothesis that has a great deal of empirical support. What is important for our purposes is to note that the principle of assimilation explains linguistic behavior without referring to any intention or decision on the part of individual speakers to simplify pronunciation.

Because assimilation invokes structural features of the language, it is an example of an internal explanatory principle of phonological change. Most internal explanations of phonetic change that historical linguists offer can be called naturalistic causal explanations. Speech sounds themselves are physical things, produced by well-understood physical or physiological mechanisms—place the tongue and lips thus and so, emit a puff of air, and so forth. Linguists can measure the amount of physical effort required to produce vowel sounds in various contexts. A principle of conservation of effort does not require appeals to intention or desire, and can be framed in physical terms. Based on knowledge of some physics and physiology, an explanation of a vowel shift in terms of assimilation is apparently a naturalistic causal explanation. At the same time, these naturalistic explanations are not reductionist. There is no question of reducing language to sound patterns or to denying that linguistic meaning is important. After all, we must at least implicitly maintain that the meaning of a word remains stable when we note that a change in pronunciation has taken place. Let us now turn to another type of causal explanation of linguistic change.

In contrast to change that is closely linked to the structure of the language, external explanations of linguistic change are based on what happens when different linguistic groups come in contact with one another. The Spanish language now spoken in Spain contains many elements that entered the language during the Moorish occupation. Spanish as it is spoken in Latin America shows the influence of Spaniards' contact with indigenous populations. Historical documents and archaeological investigation provide evidence for contacts between different linguistic groups. Trade, military conquest, and political unions based on marriages among ruling families all contribute to linguistic change. Such contacts introduce new words and new ways of pronouncing words into a language, along with new people, goods, technologies, customs, and political systems.

External explanations invoke common causes and causal processes. Historical linguistics tells us that modern Spanish and modern Italian resemble one another because they both derive from a common ancestral language, Vulgar Latin. Regular sound changes as well as other regular changes distinguish Spanish and Italian. The causal processes that led to their differentiation include geographic barriers separating Rome from its provinces as well as social and political upheavals. These factors caused disruptions in communication, and allowed the Vulgar Latin that was

spoken in different relatively isolated parts of the old Roman Empire to develop into separate though related languages.

Individual decisions can and do affect linguistic change. Military and political decisions of leaders can determine the success or failure of conquests and trade agreements, and thus *a fortiori* they can promote or prevent certain kinds of linguistic change. Desires of particular individuals to improve their status by imitating speech patterns of the rich or powerful can cause widespread changes in vocabulary and in how words are pronounced. Decisions by individuals to engage in trade bring about changes in language, as traders seek a common linguistic ground for their negotiations. Nevertheless, beliefs and desires of individuals are hardly visible in historical linguists' explanations of linguistic change.

Several reasons can be given to account for this neglect or suppression of individual beliefs and desires. Information about individual decisions during the prehistoric periods studied by linguists is usually unobtainable. However, external explanations in historical linguistics rarely appeal to individual reasons even when records of human decisions exist. Linguistic investigations focus on the processes of internal and external change rather than the varying motives of agents of these processes. How an innovation enters the language seems less important than the processes that maintain it—if it is maintained. One might say that just as evolutionary biologists focus on selective processes rather than the mutations that are the ultimate source of change, so too linguists focus on what happens once novel items have entered the language.

Individual decisions are treated as historical accidents. One could investigate causes of the so-called ''accidents,'' but such causes are irrelevant to the work of historical linguists, which focuses on the usually unconscious processes of linguistic change rather than on sources of innovation. Linguists study problems such as why a novel feature is accepted in one area, but does not spread to others, and the nature of processes that encourage or impede linguistic change. Historical linguists recognize that while a new introduction into the language can result from a conscious decision of some individual, broad linguistic change operates for the most part in the absence of conscious awareness. The internal processes that linguists invoke to account for change, such as assimilation, are not processes that speakers normally recognize. Language is such a habitual activity that most speakers respond to the pressures of external change without conscious effort.

Some external explanations of linguistic change invoke physical causes in an uncontroversial way. For example, especially before the possibility of mass communication, geographical isolation of some members of a linguistic group posed a physical barrier to free communication. When subgroups of a linguistic community become physically separated from

one another, communication can diminish or cease. Consequently, the language that was once shared develops differently in the separated groups. Over generations, linguistic innovations in the separated subgroups tend to accumulate and undergo further change. In extreme cases, the subgroups become linguistically separated in much the same way that subgroups of a biological species diversify until they can no longer interbreed.

In summary, we can say that whereas historical linguists acknowledge that reasons play some part in linguistic change, their own causal explanations of linguistic change do not invoke reasons. Thus a large and important class of explanations of human behavior is incontrovertibly causal while being irrelevant to the debate concerning whether reasons can be causes. Moreover, both internal and external explanations of changes in how words are pronounced are nonreductionist, and appeal only to physical causes.

3. Anthropology. Most philosophers recognize the strong similarities between linguistic behavior and other forms of cultural behavior. Is it possible to study social behavior (other than linguistic change) using techniques similar to those of historical linguists? Dan Sperber, in *On Anthropological Knowledge* (1985) and *Explaining Culture* (1996) proposes a way to do this. He says that the part of anthropology that investigates the *persistence* of some cultural features and *patterns* of cultural change lends itself to scientific study. This study, called the "epidemiology of representations" asks questions similar to those asked by linguists who study phonological change.

Sperber distinguishes interpretive studies (ethnography) from the part of anthropology that conducts scientific studies of human culture. He believes that both are legitimate pursuits, and—in agreement with Interpretivist critics of social science—he denies that any useful scientific (causal) generalizations can be based on interpretations. In other words, he agrees that ethnography, the traditional interest of most anthropologists, is not a science. Nevertheless, he thinks that what he calls "anthropology proper" can be studied scientifically. Moreover, he believes that genuine causal explanations of patterns of cultural change and how cultural features are transmitted are possible. The similarity between Sperber's notion of a scientific anthropology and the linguists' study of sound changes is clear. Just as the historical linguists ignore the intentions of individuals when they study phonological change, Sperber believes a scientific anthropology—one that studies the distribution of beliefs and how some beliefs become stable enough to be part of a culture—can be developed without interpreting individual's beliefs and desires.

Sperber identifies beliefs with mental representations. He accepts the cognitive neuroscientists' account of mental representations as brain states

or processes. Support for this account comes from many research programs that use functional Magnetic Resonance Imaging (fMRI) to measure changes in blood flow in various areas of the brain when specific "mental" tasks are performed. Sperber distinguishes private beliefs that may not ever be transmitted to another person, or become widespread or well established enough to be considered a part of a culture, from those beliefs, expressed as spoken or written descriptions, that are widely shared, public sorts of things. These latter beliefs, or public representations, are cultural, and it is their establishment and distribution that are the subject matter of the epidemiology of representations. Thus it is the spread of beliefs rather than their content that is to be explained causally. Scientific anthropology, however, is not reductionist, and has limited applicability. The problem of interpreting the content of beliefs remains a legitimate pursuit within the interpretive (nonscientific) branch of anthropology, the discipline Sperber refers to as ethnography.

Just as linguists study internal factors, such as the change of a sound to accommodate neighboring sounds, and external factors, such as military conquests, that transform languages in more or less regular ways, so Sperber believes anthropologists should study psychological and ecological factors that contribute to spread of representations. An example of a psychological factor is the ease with which representations can be memorized. Given the current level of attention by neuroscientists and psychologists to fMRI studies of memory, it is not too implausible to suppose that we may one day have physical measurements of degree of ease of memory that compare to measurement of ease of production of speech sounds. Examples of ecological factors include the availability of an external memory storage facility, such as writing, and the institutional environment of a society (Sperber 1996, 85–96.) As with internal and external factors of linguistic change, psychological, and ecological factors can counteract or reinforce one another, and often both sorts of factors are involved in an explanation of cultural change.

Sperber describes two main sources of beliefs (mental representations): perceptions—along with the inferences based on perceptions—and communication. All beliefs depend on communication to the extent of acquiring the concepts necessary to characterize perceptions. But some beliefs, those Sperber calls "intuitive," depend for their distribution on a wide variety of perceptual experiences and many convergent communicative experiences. These beliefs are for the most part unconsciously acquired. They form the basis of our common-sense view of the world and our ability to communicate with one another. According to Sperber, these widely shared representations guide much of our behavior and also are the basis for "reflective beliefs." These latter beliefs are not grounded in perception, but depend almost entirely on communication for their

distribution (1996, 94). Reflective beliefs include both myths and advanced scientific theories. An interesting feature of reflective beliefs is that they can be believed in the absence of complete understanding. For example, a person who is incapable of understanding the proof of Gödel's Incompleteness Theorem, can nevertheless believe the conclusion of the proof because he trusts the authority of his friend who is a mathematician.

To illustrate how psychological and ecological factors affect the distribution of reflective beliefs, Sperber discusses three examples: the distribution of myths in a nonliterate society, the distribution of political beliefs, and the distribution of a complex mathematical belief. The transmission of a myth, such as "Little Red Riding Hood," depends mainly on psychological factors, such as its attractiveness and the ease with which it can be memorized. Myths are found in every culture, and survive independently of specialized institutions, such as schools. In contrast, the transmission of a political belief depends mainly on ecological factors, such as its relevance to existing political institutions. For example, the belief that all men are created equal would probably become established as a cultural belief only in connection with an institution that espouses some form of democratic government. The transmission of belief in Gödel's Theorem differs from transmission of myths and political beliefs because it depends heavily both on psychological (cognitive factors) and on the existence of institutions for communicating such abstract forms of knowledge. For those who fully understand the proof, cognitive factors are sufficient for assent, but for those who do not fully understand it, belief depends on institutional factors, such as the acceptance of the belief by experts. Causal explanations of how these beliefs become widespread appeal to ecological and psychological factors rather than to individual reasons, and Sperber contends that ecological and psychological factors are analyzable as material causes.

In the case of anthropology, Sperber makes a number of proposals about how naturalistic causal explanations can be constructed. First of all, since the transmission of representations is the subject at hand, he must show that the representations themselves are physical objects. Cultural representations are spoken and written sentences, inscriptions, performances, paintings, buildings, and so forth. All these are physical things, just as speech sounds are physical things. As to private mental representations, Sperber's view is that these are configurations of the brain, and thus physical objects as well. Nonmaterialists challenge this view; neuroscientists, however, are committed to it. Although the exact nature of the relationship between various brain states and mental representation is a subject of intensive and problematic research, some theory linking brain states and mental states seems required for a plausible naturalistic causal account of the social sciences.

If we can accept the physical nature of mental representations, the next step in providing naturalistic explanations of the transmission of ideas is to give a naturalistic account of the mechanisms that transform private mental representations into cultural representations. (Something like this a task for linguists as well. The person who hears a spoken word forms a mental representation of it, and uses that representation to guide his or her own production of the word, but the mechanism is not entirely clear.) A number of writers, most famously Richard Dawkins, who coined the term ''meme'' to describe a unit of culture (Dawkins 1982), have proposed natural selection as the mechanism that selects some private beliefs for cultural status. Sperber rejects this idea, primarily because of a crucial difference between mental representations and genes. Mental representations are not self-replicating. Genes are replicated, but ideas are transmitted, and transmission almost always involves some degree of modification. To see that this is so, consider what happens to a simple message when it is passed along from one person in an office to another person, from the second person to a third, and so forth. Even with the transmission of a single word, the sound pattern that is heard by the listener is not identical with the sound produced when the word is ''repeated.'' If the chain of transmission involves more than a few persons, the message is almost certain to be changed along the way, sometimes to the point where it hardly resembles the original version. Exact replication of representations is possible as a limiting case—for example, in the mechanically or electronically produced copies of a printed book—but is an exception to the rule. When an idea is transmitted, alteration is the norm. Genetic transmission is different. Because mutations occur relatively rarely, identical copies of genes are the norm, and this sort of stability is necessary for selection to work.

Sperber also rejects the ''influence model'' of the transmission of cultural items. This model explains the appearance of new cultural ideas as the results of input from different sources with varying degrees of influence on the transmitted item. The influence model is often invoked in social sciences as well as in ordinary explanations of the form of public representations, such as books, paintings, and musical compositions. Discussions of how African rhythms and Southern spirituals influence contemporary jazz are familiar examples. A great deal of history of ideas, of course, is concerned with tracing influences as a way of understanding and explaining cultural forms. It would be pointless to deny the truth of many claims about influence. Sperber's chief argument against this approach to transmission is similar to his criticism of the meme approach: it does not allow for genuine novelty in transmitted items. Once weights are assigned to the various influences, they are treated as inputs that result in a completely determined output. Mental representations of transmitted

items, Sperber insists, can contain more information than that represented by the synthesis of inputs (Sperber 1996, 106).

Sperber's own view is that representations *transform* themselves as a result of a constructive cognitive process. Of course, some degree of resemblance is required among individual mental representations if they are to be considered representations of the same cultural idea—just as there must be some degree of resemblance among individual pronunciations if they are to be considered pronunciations of the same word. Nevertheless, the process by which representations are transmitted is neither one of producing a copy of itself or producing a composite drawn from various influential representations.

Sperber's choice for a mechanism is called "attraction." Attraction involves the convergence of one's "affective and cognitive processes with those of many people towards some psychologically attractive type of view in the vast range of possible views" (Sperber 1996, 106). Attraction is a statistical concept, and does not involve some physical object that attracts similar representations. For example, we judge various versions of myths, such as "Little Red Riding Hood," as better or worse, but this does not mean that they are more or less faithful copies of some ideal version. Instead, the tellings of this tale tend to converge on a version that is psychologically attractive in the sense of having such features as being more coherent and easier to remember. Again, it is helpful to compare the simpler case of convergence on a version of how a word is pronounced. The factors involved in attraction are both psychological and ecological. That is to say, environment supplies the inputs to individuals and the mental organization of the individuals determines which inputs are processed, how they are processed and how they affect behavior. In Sperber's account, large-scale changes in culture are ultimately the result of such small-scale changes as might occur with shifts in convergence on the best version of "Little Red Riding Hood" or the correct way to pronounce a vowel in a given context.

Clearly, Sperber's account of a physical causal basis for the study of cultural change and the transmission of cultural features is programmatic and dependent on the success of current theories of neuroscience. At the same time, the theory is not hampered by the constraints of traditional psychophysical causation, that is, telling a causal story of showing how a particular private mental representations (beliefs) gives rise to a bit of behavior. Sperber's ecological factors are similar in many respects to linguists' external causes for linguistic change, and while complex, can be analyzed in material terms.

4. Evolutionary Psychology. Although Sperber rejects selection as the mechanism for cultural change, his views have much in common with the

work of evolutionary psychologists such as Tooby and Cosmides (1992). Evolutionary psychologists are committed to the view that just as evolutionary pressures have shaped the human body, so have they shaped the human mind. This does not mean that every human trait has been shaped by natural selection, but rather that various components of the mind (called "modules") have been shaped for special but broadly defined tasks, such as using spoken language. Studies of the brain that use fMRI techniques offer support for mental modules because they display regular connections between certain kinds of mental stimulation and activity (blood flow) in specific areas of the brain, such as the frontal lobes. Evolutionary psychology has the task of demonstrating the selective advantage of various human features, and also of accounting for dysfunctional or obviously disadvantageous features. The selective advantage of having a spoken language is fairly obvious, whereas the selective advantage of some other admirable but costly human activities, such as playing a musical instrument very well or engaging in higher mathematics, is not so clear. Just as evolutionary biologists have the burden of explaining how harmful alleles, such as the one for malaria, persist in a population, so do evolutionary psychologists have the burden of explaining the persistence of apparently dysfunctional behavior, such as infanticide, suicide, and the like.

Evolutionary psychologists also want to explain how humans can be culturally diverse despite being very similar in their psychological make-up. They do this by pointing out the importance of environmental factors. Explaining cultural change in evolutionary terms involves special difficulties that are not involved in explaining biological change. The failure of cultural units to replicate has already been discussed. In addition, the time scale for evolutionary change is very large compared to the scale for cultural change. This problem is partly answered by arguing that many of the modules that are shaped by evolution are designed to offer "facultative" responses to environmental pressures. Just as skin tans in various degrees when exposed to sunshine, cultural responses to environmental stimuli can be very rapid. There is probably a language processing module in the brain of every human, but the first language the human learns to speak is a facultative response to the language available in his or her environment.

While the difficulties that face a full-blown science of evolutionary psychology are formidable, I believe that some progress has been made in meeting the challenges just mentioned. The program is committed to formulating and testing empirical hypotheses within the broad framework of evolutionary theory and to offering explanations that invoke physical mechanisms to account for human behavior. Let us look briefly at two examples that test hypotheses used to explain human behavior in terms of

promoting survival of their genes through favoring persons genetically related to them. Neither example appeals to intentions of the agents involved. The first example is the explanation of why in some societies, a man takes little responsibility for the welfare of the children of his own wife, but invests heavily in the children of his sister. This pattern of behavior is called "matrilineal inheritance." The evolutionary hypothesis invoked to account for it is that men expend resources on children in proportion to their (likely) degree of genetic relationship to them. One would suppose then that the societies in which men do not assume primary responsibility for their wives' children are societies in which women are likely to engage in extramarital sexual relations. In such societies, the husband's paternity is in doubt. If he invests heavily in his wife's children, he may be supporting a child who is not genetically related to him. Since maternity is almost never in doubt, however, he can be certain that he and his sister share some genes. His sister's children carry some of those genes as well. By caring for his sister's children, he can be certain that he is investing in children that carry some of his genes. Several empirical studies, including an important one by Hartung (1985), have been designed to test the hypothesis that matrilineal inheritance and paternal probability are correlated in the way described. The results of all of them support the hypothesis (Gaulin and McBurney 2001, 346–349).

The second example involves an investigation of homicide from the standpoint of evolutionary psychology (Daly and Wilson 1988). Drawing on 1970 census data for the city of Detroit, they investigated domestic homicides. In domestic situations involving both kin (blood relatives) and non-kin (such as spouses and stepchildren), they found that each type of non-kin is killed more often than expected by chance, and each type of kin less often than expected by chance. On average, according to the summary of this work in Gaulin and McBurney "coresident non-kin are at eleven times greater risk of domestic homicide than coresident kin" (2001, 324).

Both these examples lend support to the view that we are designed by evolution to be more inclined to aid and less likely to harm those who are genetically related to us than those who are not. Such results are offensive to some people because they believe that the results suggest that such behavior is inevitable, or even "nobody's fault." They worry that showing that a behavior is the result of evolutionary pressure casts some doubt on human moral agency. These interpretations, however, are incorrect. Behavior is shaped both by the psychology of individuals and the environment. Neither of these factors is immune to change. Evolutionary traits are not all rigid, particularly those that are facultative responses to the environment. Changing environments can modify behavior. The homicide study, for example, shows that non-kin are more vulnerable than kin in a domestic situation. In the light of such knowledge, moral agents can

modify domestic environments to make non-kin less vulnerable. Psychological modification is also an option, and has shown to be successful in some counseling programs for those prone to domestic abuse.

5. Physical Causation in the Social Sciences. Recent philosophical theories of physical causation, such as Wesley Salmon's (1998) account of physical causation in terms of causal processes and interactions and Phil Dowe's conserved quantity theory of physical causation, have been well-received among philosophers of the physical sciences. Briefly, the views maintain that causal connections exist in the physical world and can be discovered by empirical investigation. According to Dowe (2000), "a causal process is the world line of an object that possesses a conserved quantity, and a causal interaction involves the exchange of conserved quantities." Yet, even those who are sympathetic to these accounts of causation find it difficult to understand how they could be applied to the social sciences. Paul Humphreys, for example, remarks that "[Salmon's theory of causation] commits us to physicalism in a way that might give some pause" (2000). According to Humphreys, "This means that all anthropological and sociological causation, for example, must be accounted for by the transfer of mass-energy, linear and angular momentum, and so on. And so all explanations in those sciences must ultimately be given in terms of physical causation" (2000). Despite the extravagance of this explanatory goal, we have seen that some social scientists are willing to embrace it.

In this paper I have tried to illuminate three different proposals about scientific explanation of human behavior that are consistent with Salmon's account of causality in the physical world, and I regard this is a mark in their favor. One—linguists' explanation of phonological change—is widely accepted, and the other two are programmatic and controversial. Part of the reason for the success of the historical-linguistics example, is that its scope is clearly limited to an area which while social is close in content to areas in physiology or physics. Sperber's program is ambitious, and controversial because of its dependence on theories of neuroscience, but he also admits that the scope of his causal explanations is limited. I think that strong analogies between his proposals and historical linguistics bode well for its success. Evolutionary psychology is probably the most controversial of the three examples. But whatever its scientific flaws, it is a serious research program that is committed to physical causation and to devising and conducting experiments that test the theory. The attempts of historical linguists, anthropologists like Sperber, and evolutionary psychologists to bring the social sciences into line with the other sciences is a serious enterprise that motivates scientific research in other fields such as neuroscience. This work can be appreciated without denigrating the importance of interpretive work, such as is carried out in traditional

ethnography. We all recognize that significant insights into human behavior can be gained from interpretive studies in anthropology, history, and sociology—as well as from literary work. Nevertheless, it has become increasingly clear that such interpretive studies cannot be generalized in the manner required by a true science. Alternatively, the possibility of scientific causal explanations of human behavior has enormous practical as well as philosophical significance, because knowledge of causes is crucial to promoting social welfare and modifying behavior that dangerous and destructive.

The need for such causal knowledge and control is now clear even to humanists. Consider for example Jonathan Glover's recent work: *Humanity: A Moral History of the Twentieth Century* (1999). Glover recounts and analyzes in chilling detail moral disasters of the last century, including genocide in Nazi Germany and Rwanda, mass killings of civilians in Hiroshima and My Lai, along with terrorism and mass killing in Stalinist Russia and Cambodia. Glover, whose credentials as a humanist are impeccable, and whose opening motto is taken from Collingwood's autobiography, begins and ends his book with a plea for understanding the causes that will allow us to control such evil behavior. At the close of his introduction, he apologizes for the dark subject matter of the book, but goes on to say "We need to look hard and clearly at some monsters inside us. But this is part of the project of caging and taming them" (1999, 7). His concluding remarks are a fitting conclusion for this paper:

> To avoid further disasters, we need political restraints on a world scale. But politics is not the whole story. We have experienced the results of technology in the service of the destructive side of human psychology. Something needs to be done about this fatal combination. The means for expressing cruelty and carrying out mass killing have been fully developed. It is too late to stop the technology. It is to the psychology that we should now turn. (1999, 414)

REFERENCES

Collingwood, Robin G. ([1946] 1956), *The Idea of History*. Oxford: Oxford University Press.

Daly, Martin, and Margo Wilson (1988), *Homicide*. New York: Aldine de Gruyter.

Davidson, Donald (1980), *Essays on Actions and Events*. Oxford: Oxford University Press.

Dawkins, Richard (1982), *The Extended Phenotype*. Oxford: Oxford University Press.

Dowe, Phil (2000), *Physical Causation*. Cambridge: Cambridge University Press.

Durkheim, Emile ([1897] 1951), *Suicide*. New York: The Free Press.

Gaulin, Steve, and Donald McBurney (2001), *Psychology: An Evolutionary Approach*. Upper Saddle River, NJ: Prentice Hall.

Geertz, Clifford (1973), *The Interpretation of Cultures*. New York: Basic Books.

Glover, Jonathan (1999), *Humanity: A Moral History of the Twentieth Century*. New Haven: Yale University Press.

Hartung, John (1985), "Matrilineal Inheritance: New Theory and Analysis", *Behavioral and Brain Sciences* 8: 661–670.

Hempel, Carl G. (1965), *Aspects of Scientific Explanation*. New York: The Free Press.
Humphreys, Paul (2000), Review of *Causality and Explanation,* by Wesley C. Salmon, *Journal of Philosophy* 97: 523–527.
McIntyre, Lee (1999), "Davidson and Social Scientific Laws", *Synthese,* 120 (3): 375–394.
Mill, John Stuart (1874), *A System of Logic,* 8th ed. New York: Harper and Bros.
Salmon, Merrilee (1996), "Causal Explanations of Linguistic Behavior", *Theoretical Linguistics* 22:195–205.
Salmon, Wesley (1998), *Causality and Explanation.* Oxford: Oxford University Press,
Sperber, Dan (1985), *On Anthropological Knowledge.* Cambridge: Cambridge University Press.
——— (1996), *Explaining Culture*. Oxford: Blackwell Publishers.
Tooby, John, and Leda Cosmides (1992), "The Psychological Foundations of Culture" in Jerome Barkow, Leda Cosmides, and John Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press, 19–136.
Winch, Peter (1958), *The Idea of a Social Science*. London: Routledge and Kegan Paul.