EXPLANATORY FICTIONS

Simple explanations usually consist of describing a cause that is proposed to account for some effect. Thus, when one’s car engine makes a loud clanking noise, it may be proposed that the noise is due to a worn or damaged main bearing. The loud noise is the effect and the damaged main bearing (or some aspect of its action) is the cause. We will use these terms, *cause* and *effect* without implying anything more than a functional relation between an independent variable (cause) and a dependent variable (effect).

A primary goal of science is to discover and describe the functional or causal relations that permit the prediction, control, and interpretation of dependent variables in terms of independent variables. However, in order to predict an effect from a cause, or to control the effect by manipulating the cause, or to provide a valid interpretation of the effect in terms of the cause, it must be possible to observe the cause independently of the effect. When the cause is just another name for the effect the explanation is said to be *circular*, and the supposed cause can be called an *explanatory fiction*. The essential feature of a circular explanation is that the cause is inferred from the same information that constitutes the effect. Or said another way, when a cause cannot be observed independently of the effect for which it is the explanation, such a cause is an explanatory fiction.

Circular explanations and their explanatory fictions are quite common in dealing with behavior. This probably comes about because it is quite useful to identify and label relatively stable characteristics of a person’s behavior in certain circumstances. The receiver of such information benefits by being able to anticipate and prepare for the relevant behavior. For example, for a teacher to be told that a child coming into her class has a hostile attitude toward authority, that another is not very intelligent, and that a third has a lot of musical talent will permit the teacher to behave more appropriately toward each child, assuming that the information is correct. As a name for a consistent pattern of behavior in reaction to certain environmental circumstances “*hostile attitude toward authority*” may be quite reasonable. In such usage, *attitude* simply refers to the fact that certain events typically generate certain kinds of behavior on the part of the child. The person giving the information to the teacher may well think of *attitude* as an internal quality that is causally responsible for the behavior on the basis of which the judgment about the attitude was made, but no contact was ever made with such an internal quality. The judgment was based on the statements of others or on observations of behavior under certain environmental conditions. Likewise, the teacher receiving the information may think of the attitude as an internal causal quality, but again, the usefulness of the information is solely in terms of its relevance to environment–behavior interactions.
A problem with such terms arises, however, when an explanation for specific instances of environment-behavior interactions is required. If it is asked why the child is so uncooperative when the teacher makes a simple and reasonable request, and the answer is "because of his hostile attitude," the attitude is no longer just a name for a collection of environment-behavior relations, but has become the explanation for the relations. Attitude has become an internal or mental event that supposedly explains the environment-behavior relation, but because the only evidence for the hostile attitude is the very same behavior that the attitude supposedly explains, it is essentially an explanatory fiction in this usage.

In the case of athletic and artistic performances, some individuals are systematically better than average. Ability or talent often appears as the name for such unusual effectiveness, and information couched in such terms may be quite helpful to the recipient. In the attempt to explain such unusual effectiveness, it may be possible to identify relevant environmental events. If as a child the person had ample opportunity to participate in such activities under highly effective training conditions—parents, siblings, and friends were all competent musicians and were highly supportive of the child's efforts—the current effectiveness may seem quite understandable. However, if obvious environmental causes cannot be found it is quite common for talent to be proposed as the explanation, and thus to function as an explanatory fiction. Why does she play the violin so well? Why does she learn new musical skills so readily? Because of her outstanding musical talent. As usual, the only evidence for the talent is the unusual musical performance, which is what makes talent an explanatory fiction in such usage.

The development of an effective way of talking about behavior has been made more difficult by the fact that terms like attitude and talent (and the thousands of other similar terms) have both a useful and a harmful application. As names for relatively consistent environment-behavior relations for particular persons, they are useful; as explanations for the environment-behavior relations, they are almost always explanatory fictions. At the present time, such fictions have additional plausibility if they seem to refer to cognitive or neural processes, or if they have something to do with computers.

Note that explanatory fiction is not a general term for an inadequate explanation, many of which are not explanatory fictions at all. Most of us would not consider a person's astrological sign to be an adequate explanation for her effective artistic behavior, but the date of one's birth can be observed quite independently of one's artistic behavior. Also, historical environmental causes, although often quite useful, may be faulty (a) because the events didn't actually occur or (b) because of insufficient evidence for the relation between those events and the relevant type of behavior. Some aspects of a person's behavior may be attributed to the fact that his mother really didn't want another child before he was born. Information about what the mother wanted may be inaccurate, however, and even if accurate, the explanation suffers because of the existence of many persons with this same behavior whose mothers wanted them and the existence of many unwanted children without the behavior.

In respondent conditioning, it is very common, but quite undesirable, to add explanatory fictions to the basic functional relations. After a dog has been conditioned to salivate to a tone by pairing the tone with food, it is often said that he salivates to the tone because of his expectation that the food will follow, because of his knowledge of the tone-food relation, or because he has formed an association between the tone and the food. The only evidence for the expectation, knowledge, or association, however, is the actual salivation when the tone sounds, which is the effect we are
trying to explain. (Sometimes the explanatory fiction is in verb form: the dog expects, knows, or associates; but it is still an explanatory fiction.) It is quite sufficient to state that he salivates to the tone because of the relevant history, namely the past correlation of tone with food, with further reference to the laws of respondent conditioning, of which the present case is a specific instance. Expectations, knowledge, and associations are simply inventions that add nothing to our understanding of the relations between environmental procedures and the resulting changes in behavior. Such explanatory fictions are even worse than useless, however, because they suggest to the listener (and also to the user) that some additional knowledge is available. Focusing on inferred inner causes also serves to distract us from a careful study of environment–behavior relations.

Of course, the conditioning procedure achieves its effect on current behavior by producing physical changes in the organism that persist over time, and it is these physical changes that are directly responsible for the changed behavior. Unfortunately, there is at present no physiological knowledge that supplements the knowledge that has been and is still being obtained by direct study of environment–behavior relations. And, even when more is known about the physiological changes underlying respondent conditioning, the concepts and principles of conditioning, which constitute an increasingly extensive and precise body of knowledge, will not be contradicted or in some sense rendered useless. It is, in fact, these concepts and principles that can guide the search for physiological understanding—they constitute what it is that one is trying to understand in physiological terms.

Explanatory fictions are also popular in explaining the facts of operant conditioning, although here they are more complicated because both motivation and learning must be dealt with. After reinforcing a food-deprived rat’s lever pressing with food, students often explain the animal’s behavior as a joint function of its desire for food and its knowledge that lever pressing will result in food delivery. The desire is inferred from the deprivation regimen, and the knowledge is inferred from the lever pressing. Without explanatory fictions, the facts can be adequately dealt with as follows: When some particular behavior on the part of a food-deprived organism is reinforced with food, that type of behavior becomes a part of the repertoire that occurs more frequently under food deprivation. This is, of course, a statement of the principle of operant conditioning. So why does the animal press the lever if not because of its desires and knowledge? Because (a) we previously reinforced lever pressing with food when the animal was food deprived, and (b) the animal is currently food deprived. It will also be necessary to make reference to the laws of operant conditioning, of which the present case is a specific instance.

The explanatory fiction knowledge is sometimes varied by substituting understanding, expectancy, association, cognition, belief, and more recently a cognitive structure is said to be relevant. Need is often mentioned instead of desire, and it might seem to have biological support; the rat needs food and presses the lever because if it doesn’t it will die of starvation. This use of need is an example of another type of inadequate explanation: explaining some current event in terms of a future event, which is called a teleological explanation. A future event that has not occurred cannot be considered to be an explanation of anything. The rat’s starvation, should it occur, is clearly in the future of our current lever presser, and cannot play a role in determining the current pressing. Of course, the fact that food deprivation causes an appropriate change in the animal’s repertoire is relevant to the evolution of such an animal. Those whose behavior is fine-tuned to biological needs must certainly be more likely to survive and pass on the relevant genes to their offspring. The behavior of the
present lever presser, however, must be explained in terms of the functional relations descriptive of its behavior, not the possible future results of such behavior.

Operant behavior, by definition affected by its consequences, is especially susceptible to teleological explanation. The syntax usually involves the term to or in order to, as when the rat presses the lever to get (or in order to get) the water or food that is being used as reinforcement. Of course, the proper role of consequences in explaining an instance of behavior is in terms of the past consequences of that behavior, not the consequence that may occur after that instance of behavior has taken place. The inadequacy of teleological explanations is widely recognized, however, and when a careless instance is questioned, the user will quickly come up with an explanatory fiction, which seems to solve the problem. A student who says that a rat is pressing the lever in order to get reinforcement, when asked why the pressing occurs before the reinforcement has been obtained, or during extinction, can easily reply that it is the rat's current belief that it will be reinforced that is the actual cause. This is an explanatory fiction, but at least it is not teleological. Current belief is no real improvement, but because its inadequacy is more complex, it may well go unchallenged.

As in the respondent case, the procedure of operant conditioning must surely alter the organism in some relatively lasting way, and it is this physical alteration that is directly responsible for the fact that current behavior differs from what would have been seen if the operant conditioning had not taken place. We look forward enthusiastically to advances in the neurosciences that will permit us to observe and even measure these relevant physical events within the organism. As with respondent conditioning, however, no useful physiological supplement to the knowledge obtained from the direct study of environment–behavior relations is currently available.

Explanatory fictions continue to crop up as we deal with more complex relations between the environment and behavior (such as generalization, discrimination, abstraction) and serve the same nonpurposes. Probably the safest practice in dealing with the behavior of any nonhuman species is to avoid all terms that refer to inner mental qualities or activities. Certainly one should avoid those identified above and their various synonyms. But what about human behavior? The behavioral position is that such terms are also explanatory fictions in relation to human behavior, but the argument becomes complex because of the introspective reality of such terms to the individual whose behavior they supposedly explain. This is a topic that requires an analysis of the type of verbal behavior that is under the control of private stimuli.