THE CONCEPT OF REINFORCEMENT: EXPLANATORY OR DESCRIPTIVE?

François Tonneau
University of Guadalajara

ABSTRACT: The history of psychology has seen recurrent controversies on the circularity of reinforcement explanations, and behavior analysts disagree among themselves as to whether the concept of operant reinforcement is explanatory or descriptive. Some behavior theorists argue that the concept of reinforcement is merely descriptive, whereas others maintain that reinforcement explanations are acceptable provided extra precautions are taken. The issue of the circularity of reinforcement also has become embroiled in a more general problem, that of understanding what a scientific explanation is. Here I argue that the issue of demarcating scientific explanation from description takes two forms, and that once these two forms are distinguished most controversies vanish. Like the majority of scientific concepts, the concept of operant reinforcement is both descriptive and explanatory, and reinforcement explanations are never circular.

Key words: description, explanation, reinforcement, circularity

Skinner (1938) once portrayed his system as “positivistic” and as one that “confines itself to description rather than explanation” (p. 44). The notion that operant concepts do not provide any explanation of animal action has not fared well, however, and a number of behavior analysts (including a later Skinner, 1953, 1957) have argued that their discipline can and should explain behavior (e.g., Hineline, 1990). The issue concerns more than philosophical clarification for its own sake, and it has moved from the philosophical terrain to affect how researchers think of operant reinforcement. Is the concept of reinforcement descriptive, explanatory, or both? And if the concept of reinforcement is descriptive, are not explanations in terms of reinforcement necessarily circular? The issue of the circularity of reinforcement explanations has surfaced recurrently in the history of psychology (e.g., Meehl, 1950; Paniagua, 1985; Schnaitter, 1978) without reaching any clear resolution.

Among behavior analysts, Catania (1984, 1992) has attempted to defuse the objection of circularity by denying that the concept of reinforcement is explanatory. According to Catania, “reinforcement, stimulus control, elicitation, and so on are not explanatory terms but rather are names of phenomena” (1984, p. 714) and the concept of reinforcement is “descriptive rather than explanatory” (1992, p. 72). “Reinforcement” simply “names a relation between responses and the environment; it does not explain the relation” (Catania, 1992, p. 72). If correct,
Catania’s claim would undercut all criticisms of reinforcement explanations at the root, for if the concept of operant reinforcement was never meant to be explanatory, then the issue of circularity cannot even arise.

Other behavior analysts, however, have challenged Catania’s solution to the circularity issue and taken exception to his strictly descriptive stance on reinforcement. Hineline (2003), for example, has maintained that the concept of operant reinforcement can be, and often is, fully explanatory (p. 209), provided some caution is exercised. According to Hineline, the notion of reinforcement derives its explanatory force from its location in a wider network of behavioral concepts, from the demonstrated generality of reinforcement phenomena and from reinforcement being an instance of a selection process.

Thus, a concept widely used in behavior analysis finds itself surrounded with controversies about explanation and description. Among the efforts to clarify the issue in general terms, most have contrasted the explanatory stance of radical behaviorism with those of other scientific traditions (e.g., Moore, 2003a, 2003b) or have attempted to throw light on behavior analysis through a discussion of its historical and philosophical roots (e.g., Chiesa, 1994, Ch. 6). Lost among conceptual and historical evaluations, however, is a simple point that virtually all behavioral scientists can agree on, and that dissolves most controversies. Not all of them vanish, but those that remain are at least worth fighting for.

Two Issues

Any discussion of description and explanation should distinguish two issues. The first is to demarcate explanation from description in general. The problem is a difficult one because, among other things, virtually any description (e.g., “It is raining”) can serve as an explanation or be part of an explanation (“Why do you take your umbrella?”—“Because it is raining.”). How, when, and why a description qualifies as an explanation, and in particular a good explanation, remains a matter of philosophical debate (for an introduction to the issue see Woodward, 2002). With the deductive-nomological model of Hempel and Oppenheim (1948) serving as the starting point de rigueur, philosophers have explored the relations of scientific explanation to statistical relevance, causal processes, and theoretical unification without forgetting the pragmatic aspects of scientific discourse (e.g., Scriven, 1975).

Identifying the main characteristics of scientific explanation, however important for philosophical purposes, should be distinguished from another, more mundane, issue: that of demarcating the description of a phenomenon, B, from an explanation of this very phenomenon. The latter issue is straightforward. On virtually any view of scientific explanation, one explains a phenomenon (event, happening, state of affairs) by describing another. Of course, one does not explain a phenomenon (B) by describing it; rather, what one must do in order to explain B is to describe a phenomenon A distinct from B. Why is the temperature lower today in northern France (B)? Because of a mass of cold air coming from Russia (A). Why do objects accelerate in an inertial frame (B)? Because they are subject
to forces (A). Today’s lower temperature in northern France is a dated, singular fact, contrary perhaps to a general fact such as acceleration in an inertial frame. In either case, however, a phenomenon is properly explained by describing another (assuming the correctness of my meteorological explanation and of Newtonian mechanics, respectively).

The requirement that A and B be distinct does not prohibit B from overlapping with, or even being a proper part of, A. When I explain the polar bears’ fur color in terms of natural selection, I attribute fur color (B) to a complex state of affairs (A), natural selection for fur color, that has B as one of its parts (Endler, 1986). Yet clearly the two states of affairs, A and B, are distinct: the first (natural selection for fur color) includes reproductive success as a component, whereas the second (fur color) does not. In fact, one might argue that in a full scientific explanation A and B are never entirely distinct and that apparent counterexamples are simply shorthand for explanations in which A and B do overlap. If I say that Ivan’s dog salivates because of food, for instance, my explanation is just an elliptical way of stating that the dog’s salivation is an unconditional response to food (A), the unconditional response (B) being salivation. Again B is a proper part of A. The same argument applies to explanations in which a particular phenomenon is explained as a case of a physical law described in mathematical form. I can explain that the current through a circuit is \( i \) by taking note of the potential difference applied to it (\( V \)) and pointing out that the circuit follows Ohm’s law \( V = i R \) in having a total constant resistance value \( R = V/i \). The resultant current \( i \) is part of the \( (i, V) \) pair, which, in turn, is part of a whole linear function (corresponding to one point on its graph).

That A and B cannot be entirely distinct, while remaining nonidentical, is fully expected if scientific explanations are relational; in a scientific explanation A and B typically overlap because A includes B together with its relations to other phenomena (namely, the parts of A that are not B). The relations in question, however, must involve more than mere coincidence or succession (Owens, 1992), otherwise one could explain why it rained on Tuesday (B) merely by pointing out that it rained on Tuesday and Wednesday (A). In this example B (raining on Tuesday) is a proper part of A (raining on Tuesday and Wednesday), yet describing A in no way explains B, presumably because there is no relation of effect to cause between B and the remainder of A. But to say that scientific explanations require causation beyond happenstance (Owens, 1992) is not to say that scientific laws always express causal relations directly; laws of nature may instead express constancies that hold within sets of causes (Weissman, 1978).

Do psychological explanations follow my motto that one explains a phenomenon by describing another? Consider the explanation of behavior according to physiologism, cognitivism, and behaviorism (I use broad categories to fix ideas, and I borrow the term “physiologism” from Rachlin, 1987). Physiological explanations appeal to hormonal, biochemical, or neural events that are at least in part distinct from the behavioral phenomenon to be explained. A physiological explanation of behavior, then, is a description of the physiological variables that are supposed to explain it. Similarly, cognitivist explanations of
performance appeal to processes and structures (say, the “rules” and “representations” of Chomsky, 1980) that supposedly underlie behavior and are distinct from it. Cognitive psychologists typically intend descriptions of cognitive structure and processes to be *bona fide* descriptions of how the mind works (Baars, 1986). So again, from a cognitive perspective, psychologists explain a phenomenon (behavior) by describing another phenomenon (the workings of the mind).

The same schema holds of behavior analysis. It is uncontroversial that explanation in behavior analysis is historical (Branch, 1987). Behavior analysts explain behavior by describing the historical, in particular environmental, variables that lead to it. A history may be as short as the occurrence of an unconditional stimulus or as long as a protracted history of reinforcement and punishment (Hineline, 1990). In all cases, the explanation of a behavioral effect (B) involves a description of a particular history (A), jointly with the hypothesis that A is actually responsible for B. If the latter hypothesis is true, then A actually causes B, and describing the genealogical relation between A and B constitutes an adequate explanation of B.

Now cognitive psychologists and philosophers often question, if not fully deny, the viability of any behaviorist explanatory program. Their objections, however, have more to do with an alleged lack of generality of behaviorist explanations than with their *a priori* illegitimacy. Cognitivists argue, for example, that performance (especially complex linguistic performance) depends on a variety of interacting mental states and thus that the causal relations between environment and behavior, although genuine, are too idiosyncratic to sustain law-like generalizations (e.g., Dennett, 1981; Pylyshyn, 1984). But a causal explanation that may not generalize across settings is still an explanation, however crude or ultimately unsatisfactory.

Another way of contrasting behavior analysis and cognitive psychology is to distinguish transition theories from property theories (Cummins, 1983, 2000; Smith, 1988, 1994). Transition theories explain a system’s behavior by tracing them to previous causes, whereas property theories explain behavioral capacities in terms of the system’s internal organization. Capitalizing on this distinction, Cummins (1983) argues that “a good transition theory is not an explanatory theory” (p. 99), for though it does explain individual events, “this is not its primary scientific role. Its primary role is rather to *specify precisely* the explananda of explanatory theories, viz., the dispositional properties of systems. It is then the business of a property theory to explain these via analysis” (p. 99). Thus, like most cognitive scientists, Cummins believes that explanatory theories are of the property type, and that they embody the proper task of psychology. They are deeper, perhaps more scientific, and ultimately more satisfying. Molar behavioralists are bound to disagree and insist that their transition theories are explanatory as well as fully appropriate to psychology (e.g., Rachlin, 1994). But notice that Cummins (1983) nowhere denies that a transition theory explains particular events, such as the occurrence of a response or an increase in response rate. In fact, he explicitly states that a good transition theory does explain individual events (p. 99).
To sum up, psychologists and behaviorists of all persuasions can agree that one explains a particular phenomenon (B) by describing another (A), regardless of whether B involves particular events or law-like generalizations, and regardless of whether A is physiological, environmental, or cognitive. In fact, I can think of only one philosophy of science in which (at least some) explanations are not descriptions. This philosophy is instrumentalism (e.g., Gardner, 1979). Instrumentalism presupposes a clear demarcation between observational and theoretical terms, and it conceives of the latter as instruments or tools, justified only insofar as they are useful in predicting observables. From this perspective, a theoretical explanation is no more a description than a chisel or a hammer could be, for chisels and hammers, even good ones, do not describe anything. Thus, in the case of an explanation that involves theoretical concepts, instrumentalists will disagree with my claim that one explains a phenomenon by describing another.

In discussing whether the concept of reinforcement is descriptive or explanatory, however, I propose to set instrumentalist scruples aside because nobody in this dispute is an instrumentalist. Physiological psychologists are not instrumentalists about the brain, for example, and neither are cognitive psychologists instrumentalists about the mind, officially at least (but see Loftus, 1985). Chiesa (1994), a radical behaviorist, has promoted a view of explanation that comes close to instrumentalism in the case of cognitive constructs, which she admits only as temporary crutches to be discarded as soon as possible. But even though Chiesa may be an instrumentalist with respect to cognitive states, clearly she is not an instrumentalist about operant reinforcement. Neither are other behavior analysts, to my knowledge, although some of them may endorse an instrumentalist stance with respect to the mathematical concept of reinforcement “value” (e.g., Rachlin, 1989, p. 158). Finally, published discussions of the circularity of reinforcement explanations, either pro or con (e.g., Gallup, 1984; Paniagua, 1985), have typically proceeded without any adherence to instrumentalism.

Reinforcement Explanations

What, then, of the concept of reinforcement? Is it descriptive or explanatory? At this stage of my argument I hope that the answer is obvious. The concept of reinforcement, like most scientific concepts (and unless one adopts an instrumentalist stance about it), is both explanatory and descriptive. When behavior analysts attribute the prevalence of a response to reinforcement, they describe a particular type of relation between responses, subsequent events, and further responding (A), and in the process they explain why a response persists or increases in prevalence (B). Notice that A and B are distinct (even though they overlap), as they must be if the description of A is to serve as an explanation of B.

Now Catania (1992, p. 72) correctly points out that the term “reinforcement” names a relation between responses and the environment (A) and cannot be used to explain the relation in question (A), but no one has ever tried to do so. What behavior analysts have done instead is to use the term “reinforcement” as a
description of a complex relation between responses and the environment (A) to explain why a response is emitted at a certain rate (B). “Reinforcement” is both a description of A and an explanation of B. Catania’s further conclusion that “the term reinforcement is descriptive rather than explanatory” (p. 72) overlooks the fact that any scientific term can be both.

Among behavior analysts, and from a Machian perspective on causation, Moore (2003a) comes closest to realizing the dual nature of most scientific terms. On the assumption that explanations are descriptions of empirical regularities, however, he concludes that the distinction between description and explanation, while worth preserving, is at times tenuous (p. 181). Moore’s worry is easily set aside by recognizing two different issues, the issue of distinguishing explanation from description \textit{sui generis} and the issue of distinguishing the description and explanation of a particular phenomenon. The former distinction is both tenuous and difficult, but the latter distinction is not tenuous at all and is definitely worth preserving.

The notion that one explains a phenomenon (B) by describing another (A) also puts to rest the accusations of circularity that have been raised against reinforcement explanations. One of the early accusers, Postman (1947), was concerned that without hedonism “the law of effect can do no more than claim that the state of affairs resulting from a response in some way influences future responses. Such a statement is a truism and hardly lends itself to the rigorous deduction of hypotheses and experimental tests” (p. 501). Postman proposed to break the alleged circularity of reinforcement explanations (the “truism”) by identifying reinforcers with events that provoke pleasure. From Postman’s perspective, then, hedonism was one way to predict through independent criteria what events will be reinforcers. Absent such criteria, “the satisfying or annoying nature of a state of affairs can usually be determined fully only in the course of a learning experiment and cannot then be invoked as a causal condition of learning without circularity” (p. 497).

The statement that the state of affairs resulting from a response influences future responses is admittedly too vague to lend itself to hypothesis testing—yet this statement, however vague, \textit{is not} a truism. Accusations of circularity overlook the possibility of an activity increasing in frequency for reasons that have nothing to do with the states of affairs resulting from previous responses. If the activity is evoked invariably by an unconditional stimulus, for example, and if the latter increases in frequency, then the activity it evokes will also increase in frequency. Alternatively, the entire pattern of increasing activity may be a single macro-response evoked by a single unconditional stimulus (e.g., Killeen, 1975). The observed increase of response rate also may be a by-product of increasing deprivation in the absence of reinforcement (Sheffield & Campbell, 1954), or a target response can increase in prevalence merely because some competing activity is being prevented (e.g., Staddon, 1982).

Thus, there is no shortage of hypotheses to explain why a response increases in frequency. The hypothesis of operant reinforcement is only one of them, and to say that a response is reinforced is \textit{not} to say that this response increases in
frequency. It is to say that the temporal pairing or correlation (Baum, 1973) between the response and another event plays some causal role in increasing or maintaining the response rate. Even if we have no idea what this other event is or why it functions as a reinforcer, we are still making an empirical hypothesis, and one that may well be false. Plaintiffs of circularity confuse an explanation’s being made after the fact with an explanation’s being circular. A truly circular explanation would be something like “response rate increases because response rate increases,” but no one has ever said such a thing.

Of course, if we do not specify what event is a reinforcer for the target response, the hypothesis that the latter is maintained because it is reinforced—although a perfectly valid and noncircular explanation—does not lead us very far. A fruitful scientific hypothesis should at the very least specify what reinforcer is involved (food? water? eye contact?). The hypothesis can then be tested experimentally (and regardless of hedonism) by comparing the response rate in the current situation with what happens when the putative reinforcer is discontinued or presented independently of responding (Thompson & Iwata, 2005). Our explanatory efforts further improve if we can predict, on independent grounds, when a given event functions as a reinforcer (e.g., Hanson & Timberlake, 1983; Premack, 1959). But even if we have none of this, the claim that a response is prevalent because it is reinforced involves no circularity. Remember, to say that a response is reinforced is not to say that the response is prevalent; rather, it is to say that the response is prevalent because of a causal process of a certain type.

In another criticism of reinforcement explanations, Gallup (1984) states that “a change in behavior is not explained by relating it to reinforcement until reinforcement has been explained” (p. 627). This line of argument differs from Postman’s (1947), but it is equally misguided, for it is simply false that a change in behavior is not explained by reinforcement unless reinforcement is itself explained. An explanation of behavior change (B) in terms of reinforcement (A) does not answer any question about reinforcement itself (A), but it is still an explanation of behavior change (B), and it is never circular because, as we have seen, alternative explanations are always conceivable (Higgins & Morris, 1985). Of course, after asking what explains a change of behavior and answering in terms of reinforcement, we can ask about what explains reinforcement itself. Presumably, the answer to that question will involve yet another layer of explanation (and therefore description). But as we add a layer to another, we do not pass from nonexplanation to explanation; rather, we add an explanatory level to a hierarchy that already was explanatory. If one could not explain a phenomenon at level \( n \) unless one’s explanans\(^1\) was itself explained at level \( n + 1 \) (which basically is what Gallup implies), then one could never explain anything.

---

\(^1\) Here is how Hempel and Oppenheim (1948) define “explanandum” and “explanans”: “By the explanandum, we understand the sentence describing the phenomenon to be explained (not that phenomenon itself); by the explanans, the class of those sentences which are adduced to account for the phenomenon” (pp. 136-137). By this definition, explanandum and explanans are linguistic entities. In contrast, I use “explanans” to refer to a
Conclusion

Explaining B in terms of A is circular if it is logically impossible for A to fail to obtain when B obtains. Attributing John’s being a bachelor to his having no wife would be a case of circular explanation. Are there circular explanations in psychology? Perhaps a few. Although I have no hard data on this issue, I suspect that circular explanations in psychology and behavior analysis are infrequent at best, if not simply nonexistent, yet psychologists and behavior analysts regularly accuse one another of this largely imaginary sin. Just as mentalists have claimed that reinforcement explanations are circular, behaviorists have accused mentalistic hypotheses of circularity (e.g., Skinner, 1953).

Sometimes the accusations even bounce back. Gallup (1984), for example, quotes approvingly the Skinnerian argument that “a disturbance in behavior is not explained by relating it to felt anxiety until the anxiety has in turn been explained” (p. 627) before proceeding with a tu quoque about reinforcement. In fact, both accusations (Skinner’s as well as Gallup’s) are equally unfounded. It is no more circular to attribute a behavioral disturbance to felt anxiety than to attribute an increase in response rate to operant reinforcement. In either case alternative explanations are conceivable. Response rate may increase through mere elicitation, and the observed disturbance may be nothing more than a malingerer’s clever acting. The behaviorist and the mentalist will need to find better indictments against each other than that of circular reasoning.

What complicates the issue in the case of reinforcement is that some statements made about it were, arguably, circular. Consider Skinner’s (1983) “LAW OF CONDITIONING OF TYPE R. If the occurrence of an operant is followed by presentation of a reinforcing stimulus, the strength is increased” (p. 21). This so-called law is circular (that is, devoid of empirical content) unless supplemented by independent criteria for the reinforcing function of the stimulus in question (Meehl, 1950). Skinner’s (1938, p. 21) statement was offered as a law, though, not an explanation, and in any case it faded away from the literature. What remains today are various explanations of behavior in terms of operant reinforcement. These explanations may be wrong, and perhaps even bad in some cases, but circular they are not.

As for the term “reinforcement,” it is descriptive of a phenomenon (a relation between environment and behavior) while being explanatory with respect to another (response rate). In this respect “reinforcement” does not differ from other scientific terms such as “gravitation,” “magnetism,” or “osmosis.” Of course, in any particular case the maintenance of a response may or may not involve operant reinforcement; that it does is an empirical hypothesis to be evaluated on a case-by-case basis through further observation, experimentation, and inference.
My argument has a more general implication. When encountering a new psychological term, do not ask whether it is descriptive or explanatory. Rather ask, “Descriptive of what?” and “Explanatory with respect to what?”.

References