Bulletin of the Psychonomic Society
1984, 22 (6), 562-565

Philosophical criteria for psychological explanation

MICHAEL H. BIRNBAUM
University of Illinois, Champaign, Illinois

A psychological explanation consists of a set of premises postulated to account for the occurrence of some behavioral phenomenon. This paper contends that an explanation should satisfy at least five philosophical criteria: (1) The explanation must be *deductive*; that is, it should be possible to deduce the occurrence to be explained from the premises. (2) The explanation must contain empirical *meaning*. It should not be a contradiction to assert that the explanation can be falsified by empirical observation. (3) The explanation must be *predictive*; that is, in principle, if the explanation were known in advance, the occurrence would have been predicted. (4) The explanation should contain a *causal* argument; in principle, the explanation identifies a means for control of the behavior to be explained. (5) The premises should be *general*. Other testable predictions should follow from the premises. Thus, the premises used in an explanation should be usable in explanations of other phenomena. Every good explanation leads to a new experiment that threatens to disprove it.

Psychology is the science of behavior. As a science, it is concerned with the comparison of alternative explanations. Yet, many so-called "explanations" of behavior offered by psychology are not even wrong; rather, they are meaningless. The purpose of this paper is to argue that potential "theories" be subject to philosophical analysis before they are even considered as explanations. If a potential "theory" does not satisfy five philosophical criteria, it should not be considered a potential explanation and is not to be worthy of empirical evaluation. Quotation marks will be used to denote statements and arguments offered as "explanations," whereas the absence of quotes will denote that the criteria are satisfied. Of course, an explanation satisfying the philosophical criteria may be empirically wrong, which is the right of every explanation.

FIVE CRITERIA FOR EXPLANATIONS

C1. Deductive

An explanation of event E consists of a set of premises from which one could deduce the occurrence of E. This criterion may seem obvious in certain instances. For example, if someone asked,

"Why did Jones commit suicide?", (E1)

the event to be explained is Jones's suicide; it would not be explanation to reply,

"God works in mysterious ways," (P1)

for one cannot deduce E1 from P1.
More subtle problems occur; consider, for example, P1.1:

"Jones was unhappy." (P1.1)

The author's mailing address is: Department of Psychology, University of Illinois, 603 E. Daniel, Champaign, IL 61820.

If one were to add the second premise,

"All persons who are unhappy commit suicide," (P1.2)

the argument consisting of P1.1 and P1.2 would become deductive, assuming Jones is a person. Usually, however, the second premise would be of the form

"People who are unhappy are more likely to commit suicide than people who are happy." (P1.2a)

From these premises (P1.1 and P1.2a), we deduce that Jones had a higher probability of committing suicide, but we could not explain *his* suicide. Some philosophers would be satisfied with probability enhancement as a weaker form of explanation. A clearer view would be to restate the phenomenon to be explained as a probability, rather than as a single event.

Deduction, although necessary, does not satisfy, since it is possible to deduce meaningful or true conclusions from meaningless or false premises.

Why do teenagers masturbate? (E2)

Because

masturbation causes blindness, (P2.1)
all teenagers like to do anything that causes blindness, and (P2.2)
all teenagers do everything they like to do. (P2.3)

From these premises, which seem empirically false, we can deduce the conclusion that teenagers masturbate.

This example illustrates that just because one can deduce an empirically reasonable conclusion from some premises, one cannot conclude that the premises are true or reasonable. If the conclusion were false, however, then one could question the premises.

Another "explanation" for E2 is as follows:

A logically unverifiable devilish spirit exists in the id of all teenagers. (P3.1)

This spirit causes a teenager to masturbate whenever the level of libido in the id exceeds 200 ferg/cm³. (P3.2)

The level of libido in the id exceeds 200 ferg/cm³ for all teenagers. (P3.3)

This example shows that meaningless premises can be used to deduce meaningful conclusions. It would be a contradition to assert that one could refute premise 3.1; therefore, P3.1 is devoid of empirical consequences. Although the status of premises 3.2 and 3.3 is uncertain, measurement of libido is by definition impossible, because the id is part of the subconscious (which by definition, is unverifiable to even the mind in question). This example motivates the second requirement of an explanation, that an explanation be meaningful.

C2. Meaningful

An explanation of an empirical event must contain empirical meaning. The empirical meaning of a proposition is equivalent to the set of operationally specifiable testable implications. Empirical propositions often carry a great number of implications by making use of implicit, assumed laws of nature. For example,

"I am holding a coin in my hand"

implies that if I were to open my hand, an audience would report seeing the coin, that my hand would weigh more than a hand without a coin, that an x-ray would reveal the presence of a coin, etc. If an utterance has no testable implications, it is said to be devoid of empirical meaning, or *meaningless,* for short. An explanation does not consist of premises that are totally devoid of empirical meaning.

Negative thinking is important. One should ask how the premises can be falsified. If it is a logical contradiction to assert that at least one premise can be falsified, the theory is meaningless.

Some persons are confused by this requirement, and are occasionally fooled by circular arguments such as the following:

"If God did not exist, I could not continue to live."

"But how do you know God exists?"

"I'm still alive."

Or another:

"An animal eats when it is hungry."

"Hunger in animals is measured by the amount of food eaten."

Or another:

"Why did the child respond that the amount of water was the same when it was poured from the tall skinny glass to the low fat glass?"

"Because the child has acquired the concept of conservation of volume (in which the child also realizes that water is incompressible)."

"How do you know the child has acquired the concept?"

"Because he responded that the amount of water was the same."

These circular arguments are not considered explanations because they cannot be falsified except by denying the conclusion that they were designed to explain.

C3. Predictive

An explanation, in principle, would allow one to predict the occurrence and nonoccurrence of the event in advance. An explanation must not be ad hoc, in principle; it should be the case that if one knew the explanation and all of the states of nature in advance, one could have predicted the event to be explained.

Freudian accounts are famous for being devised after the fact to account for individual events.

"Why did Hinkley shoot Ronald Reagan?" (E4)

Because:

Hinkley had an unresolved Oedipal complex, (P4.1)

an unresolved Oedipal complex is resolved by either (a) killing your father, (b) loving him, (c) engaging in sporting rivalry, (d) ignoring him, or (e) something else that is symbolically equivalent. (P4.2)

Hinkley identified Reagan with his father. (P4.3)

If something is identified with something else, it will be treated identically. (P4.4)

Hinkley could not resolve his Oedipal complex by means of premise P4.2 (b), (c), (d), or (e). (P4.5)

Hinkley resolved his unresolved Oedipal complex. (P4.6)

In addition to being devoid of empirical content, the above explanation has many "free parameters" that are filled in appropriately and that are not specifiable in advance.

Consider the following predictive explanation:

Why were there 50,000 illegitimate pregnancies in Los Angeles last year? (E5)

The number of illegitimate pregnancies per year in a city is given by the equation $Y = aX$, where Y = number of illegitimate pregnancies, X = number of priests in the city, and $a = .5$. (P5.1)

The number of priests in Los Angeles is 100,000. (P5.2)

This "explanation" is deductive, meaningful (one could falsify both premises), and predictive. A predictive system, satisfying C1 to C3, can be extremely useful. Kepler's laws of astronomy lead to a predictive system of great great value. Indeed, Hospers (1953) defined an explanation as a general set of premises satisfying C1-C3. However, most people will not accept P5.1 and P5.2 as an explanation, despite the ability to predict.

C4. Causal

An explanation contains a causal argument. In principle, it identifies a means for the *control* of the event to be explained.

A causal statement is an empirical generalization that requires inductive evidence to establish its empirical status. The meaning of a causal statement is contained in its implications for the outcomes of experiments. In an

experiment, agents control the suspected cause and measure the hypothesized effect. If A causes B, it should be possible to control the occurrence of B by manipulating A.

An argument of the form

"A causes B; A occurred; therefore, B occurred"

is deductive, meaningful (if A is logically observable), predictive, and causal.

Inability to manipulate A in practice does not change the status of the explanation. Astronomers often theorize meaningfully about the behaviors of stars, whose masses are technically impossible to manipulate. However, their theories are meaningfully causal because, *in principle*, if the masses could be manipulated, the theories would dictate the consequences.

Often in the social sciences, correlational evidence is offered for a causal theory. For example, some might argue that priests cause illegitimate pregnancies on the basis of the correlation. The President's Commission on obscenity and pornography speculated that religious training causes sex criminality, since sex criminals were more likely to have received strict religious training than comparison groups who were not criminals. Correlational evidence should not be accepted as evidence for causation any more than ability to predict the future should be confused with the ability to control it.

It should also be noted that freely occurring correlations in nature often have the opposite sign from that of the true causal relationship. For example, patients who received high doses of penicillin last year are more likely to be dead this year than people who received no penicillin last year. However, when people with infections are assigned randomly to conditions, penicillin decreases the probability of death. Knowledge of the correlation between penicillin and death might be useful to an insurance agent who would avoid selling life insurance to people who received penicillin last year. However, the same agent would advise his policy holders to take penicillin for their infections, based on the experimental results.

Some authors would accept a deductive, meaningful, predictive, and causal argument as an explanation. However, with only these criteria, every event could potentially have its own explanation. The explanation for one occurrence of an event and another might be different. Similarly, a theorist might assume P1 in order to deduce E1 and then offer a contradiction of P1 in the explanation of E2. It seems desirable to take the premises seriously as attempts to formulate "laws" that are either true or not. Hopefully, explanations should produce some economy in our attempts to understand the world. To accomplish this economy, one additional condition is needed.

C5. General

Premises used in the explanation of one event should be used in the explanation of other events. In other words, an explanation should be part of a larger system that explains many events and has many testable consequences.

The greater the number and variety of implications of an explanation, the greater the power of the system. The word "general" has an unfortunate history in social psychology, in which it has assumed a very narrow meaning of "will this result work in the 'real' world?" Rather, generality refers to a much wider array of situations. When premises used in one explanation are used in others, the system containing both explanations is called *coherent.*

ALTERNATIVE EXPLANATIONS

If two different sets of premises satisfying Criteria 1-5 can both explain E, they are said to be alternative explanations of E. Alternative explanations of event E often lead to distinct predictions for the outcome of a new experiment. Experiments are devised and conducted to reduce the number of alternative explanations that are compatible with empirical evidence.

Example of Experimental Comparison

Gregory (1966) offered the following simple example:

Why does the perception of the world stand still even though the eyes are moved? (E6)

Sherrington and Helmholtz offered alternative theories, discussed by Gregory (1966). Both agreed that an internal brain mechanism determined perceptual location as follows:

$$L_{O:W} = L_{O:R} + L_{R.W}, \quad \text{(P6.1)}$$

where $L_{O:W}$ is the perceived location of an object in an external ("world") frame of reference, $L_{O:R}$ is the location of an object determined by the retinal location of its image and $L_{R:W}$ is the perceived location of the retina relative to the external, "world" frame of reference.

Sherrington proposed an *inflow* theory of $L_{R:W}$, which assumes the following:

Sensory mechanisms detect and compute the value of $L_{R.W}$, to determine perceived location of the eye. (P6.2a)

Helmholtz proposed an *outflow* theory as follows:

$L_{R:W}$ is determined by motor command, that is, by volition of desired location of the eye. (P6.2b)

Both theories satisfy the minimal criteria, C1-C5, for explanations. The theories are alternative explanations for the following events:

The world does not seem to move when we move our eyes. (E6)

A moving object is perceived as moving, whether the eyes follow or not. (E6.1)

A moving lighted object in a dark room is perceived as moving whether eyes follow or not. (E6.2)

However, the theories make distinct predictions for the following experiments:

(1) If the eye is passively moved (no volition), outflow theory predicts apparent movement, whereas inflow predicts no movement.

(2) If curare is injected into the eye muscles, volitions can be present without movement, since the drug brings peripheral muscle movements to a halt without causing unconsciousness. Outflow theory predicts that attempts to move the eyes to the right will produce apparent movement to the right, whereas inflow theory predicts no movement.

(3) Pushing on the eye with an afterimage should produce no apparent movement of the afterimage, according to outflow theory; it should produce apparent movement, according to inflow theory.

By performing these experiments and others, an empirical basis for preferring one theory over the other can be established. (Experiments have shown that there is apparent movement of an external object with passive movements of the eye, that there is apparent motion with volitions to move the eye under curare, and that there is no apparent movement of an afterimage with passive movement of the eye. These results are consistent with the outflow theory of Helmholtz.)

Nonexperimental Comparison of Alternatives

Predictive power. If the implications of one explanation are a proper subset of the implications of another, the theory with the greater number of implications is preferred.

Equivalent explanations. Two alternative explanations are said to be *equivalent* if the sets of empirically testable implications are identical.

Ockham's razor. If two explanations are equivalent, and one consists of a proper subset of the premises of the other, the alternative with fewer premises is preferred. In other words, premises that are not required to yield deductions should be deleted.

Simplicity. If two explanations are equivalent, the simpler one is preferred. Unfortunately, the concept of "simplicity" is not simple, because people often do not agree in their judgments of simplicity. Furthermore, theories often seem equivalent until a clever person thinks of an experiment to differentiate them.

CRITICISM OF EDUCATION IN PSYCHOLOGICAL THEORY

There are few good examples of psychological theories that satisfy the minimal philosophical criteria. Unfortunately, there are many examples of so-called "theories" of psychology that do not satisfy the minimal criteria. Furthermore, students of psychology have not been trained to distinguish between real theories in psychology and faddish movements that use the word "theory" in their titles.

In order to assess prior education and to impress upon students the limitations of their knowledge, I have asked several hundred psychology undergraduates and about 100 psychology graduate students to do the following homework assignment:

> Select any phenomenon from psychology or every-day life. Write down clearly the phenomenon you wish to explain, then write down a set of premises from which you can deduce the phenomenon. Show that your explanation satisfies the criteria, C1-C5. In other words, can you explain anything?

So far, no student has yet listed a set of premises that would satisfy the philosophical criteria for an explanation. A handful of graduate students, all of whom selected examples from physics, have come close, but no student has yet provided one explanation.

Psychology has for too long tolerated fuzzy thinking, vague "theory," and poorly conceived studies. More emphasis should be placed on the power of negative thinking. Investigators should ask themselves, "If my theory is wrong, how can I find out? What experiment seems most likely to disprove my theory?" If no experiment could conceivably disprove the "theory," then the theory is meaningless and deserves no further consideration. All too often, experiments are designed to confirm the investigators' intuitions of "If my theory is true, it seems as though I ought to be able to find X." Such studies are not diagnostic unless an opponent theory clearly fails to tolerate the prediction, in which case the study is really a test of the opponent theory. The introduction of all experiments should contain a list of clearly stated premises that are being tested by the experiment, and a statement of how the theory can be disproved.

In sum, an explanation consists of a set of premises from which one can deduce the phenomenon or phenomena to be explained. The premises are meaningful, predictive, causal, and general. The best explanations are those that can be easily disproved by experiment, but have withstood the assault of empirical evidence.

REFERENCES

GREGORY, R. L. (1966). *Eye and brain.* New York: McGraw-Hill.

HOSPERS, J. (1953). *An introduction to philosophical analysis.* Englewood Cliffs, NJ: Prentice-Hall.

(Manuscript received for publication June 15, 1984.)