



12-1997

An Intervening Cause Counterexample to Railton's DNP Model of Explanation

Stuart Gluck

Steven Gimbel
Gettysburg College

Follow this and additional works at: <https://cupola.gettysburg.edu/philfac>

 Part of the [Philosophy of Science Commons](#)

Share feedback about the accessibility of this item.

Gluck, S. and Gimbel, S. (1997). An Intervening Cause Counterexample to Railton's DNP Model of Explanation. *Philosophy of Science* 64(4), 692-697.

This is the publisher's version of the work. This publication appears in Gettysburg College's institutional repository by permission of the copyright owner for personal use, not for redistribution. Cupola permanent link: <https://cupola.gettysburg.edu/philfac/11>

This open access article is brought to you by The Cupola: Scholarship at Gettysburg College. It has been accepted for inclusion by an authorized administrator of The Cupola. For more information, please contact cupola@gettysburg.edu.

An Intervening Cause Counterexample to Railton's DNP Model of Explanation

Abstract

Peter Railton (1978) has introduced the influential deductive-nomological-probabilistic (DNP) model of explanation which is the culmination of a tradition of formal, nonpragmatic accounts of scientific explanation. The other models in this tradition have been shown to be susceptible to a class of counterexamples involving intervening causes which speak against their sufficiency. This treatment has never been extended to the DNP model; we contend that the usual form of these counterexamples is ineffective in this case. However, we develop below a new version which overcomes these difficulties. Thus we claim that all of the models in this tradition, DNP included, have an equal status with respect to sufficiency.

Keywords

Railton, explanation, philosophy of science

Disciplines

Philosophy | Philosophy of Science

Discussion: An Intervening Cause Counterexample to Railton's DNP Model of Explanation*

Stuart Gluck and Steven Gimbel†‡

Department of Philosophy, Johns Hopkins University

Peter Railton (1978) has introduced the influential deductive-nomological-probabilistic (DNP) model of explanation which is the culmination of a tradition of formal, non-pragmatic accounts of scientific explanation. The other models in this tradition have been shown to be susceptible to a class of counterexamples involving intervening causes which speak against their sufficiency. This treatment has never been extended to the DNP model; we contend that the usual form of these counterexamples is ineffective in this case. However, we develop below a new version which overcomes these difficulties. Thus we claim that all of the models in this tradition, DNP included, have an equal status with respect to sufficiency.

Traditionally, there have been two major approaches to scientific explanation in the philosophy of science: formal models (e.g., Hempel, Salmon) and ordinary language (e.g., Scriven, Achinstein). Peter Railton (1981) has attempted to wed these traditions through his distinction between an ideal explanatory text and explanatory information. The ideal explanatory text is the sufficient, rigorous, true explanation of a given explanandum governed by formal conditions akin to those employed in the formal model tradition; explanatory information is any of the set of explanation fragments that might be offered in particular

*Received August 1996.

†Send reprint requests to the authors, Department of Philosophy, 347 Gilman Hall, 3400 N. Charles Street, Johns Hopkins University, Baltimore, MD 21218-2690.

‡We would like to give special thanks to Robert Rynasiewicz and Peter Railton for extremely useful comments and discussion. Peter Achinstein, Karen Neander, Christopher Grau, Charles Choo, Stephen Dillingham, and Ray Rennard also provided helpful suggestions. Our appreciation goes to David Lunkin, who graciously created Figure 1.

Philosophy of Science, 64 (December 1997) pp. 692-697. 0031-8248/97/6404-0010\$2.00
Copyright 1997 by the Philosophy of Science Association. All rights reserved.

circumstances when the sort of pragmatic and contextual factors with which the ordinary language approach is concerned deem the full ideal text unnecessary or unwanted. Thus each tradition amounts to an aspect of the full account of explanation. The potential of this project to forge a new framework for understanding scientific explanation has been commented upon by such notable philosophers as Wesley Salmon:

Given Railton's conception of explanatory information and the ideal explanatory text, it seems to me that much of the long-standing battle—going back to the early skirmishes between the logical empiricists and the ordinary language philosophers—can be resolved. I see this achievement as one foundation upon which a new consensus might be erected. (1990, 162)

To play the role of the ideal explanatory text for probabilistic explanation, Railton developed the deductive-nomological-probabilistic (DNP) model. It possesses properties which are claimed to make it uniquely suited to assume this place. The DNP model departs sharply from traditional probabilistic explanations in that it is deductive, rather than inductive, in nature. As a result, DNP explanations are either true or false independent of our epistemic situation. The criteria of truth are the "validity of the argument" in conjunction with the truth of the premises. True DNP explanations are limited to truly indeterministic processes since any attempt to explain deterministic situations by appeal to the DNP model must include a covering law of probabilistic form which would necessarily be false, as it assigns a probability to a nonprobabilistic process.¹ Additionally, requirements of maximal specificity are disowned as unnecessary, as the criteria of truth ensures that "unspecific" premises are false.

The form of a DNP model is the following:

- (a) *Derivations of the laws (b) from the relevant theoretical accounts of the mechanisms at work.*
- (b) *Probabilistic covering laws.*
- (c) *Descriptions of the relevant circumstances.*

- (d) *Statement of the probabilities of occurrence of the events to be explained.*
- (e) *A parenthetic addendum to the effect that the events to be explained did or did not actually occur.*

1. This fact comes about because Railton adopts a propensity account of probability, so that deterministic cases cannot simply be assigned probabilities of 0 or 1. Further, the propensity account allows the ascription of probabilities to single cases.

(b), (c), and (d) form a DN inference to the probability of occurrence of the event to be explained, with the usual requirements for such an inference (Hempel 1965). Or, more formally:

- (a') *A derivation of (b') from the theoretical account of the mechanism at work.*
 (b') $\forall x \forall t [Fx, t \rightarrow \text{Prob}(Gx, t) = p]$
 (c') Fe, t_0

 (d') $\text{Prob}(Ge, t_0) = p$
 (e') $(Ge, t_0) \sim (Ge, t_0)$

for a single event, *e*, having probability *p* of occurring at time *t*₀.

Before attempting to formulate an intervening cause counterexample to the DNP model, it will be helpful to see how this class of counterexamples have been applied to the other models in the formal tradition (Achinstein 1983). Consider, for instance, the following DN explanation of why Jones died within twenty-four hours after time *t*:

- (f) *Jones ate a pound of arsenic at time t.*
 (g) *Anyone who eats a pound of arsenic will die within twenty-four hours of doing so.*

 (h) *Jones died within twenty-four hours after time t.*

Now suppose that while Jones did eat a pound of arsenic at time *t*, he also was hit by a truck at time *t'*, such that $t < t' < t + 24$ hours, and this traffic accident was the actual cause of Jones' death. In this case, the above explanation still satisfies all of the criteria for a good DN explanation of Jones' death despite clearly being unacceptable as it appeals to the wrong cause. Therefore, one can conclude that the DN model is not sufficient for scientific explanation; and similarly for other models in the tradition (Achinstein 1983).

That the above example is inapplicable to the DNP model is obvious—it is a deterministic case. If we alter it to an indeterministic case (supposing somehow that it is truly indeterministic) the DNP explanation, unlike the DN example, is invalidated by the intervening cause. Suppose that the situation is the same as before except now the probability of death within twenty-four hours for one who has eaten a pound of arsenic is .95. The DNP explanation of Jones' death is:

- (i) *A derivation of (j) from the relevant toxicological theory.*
 (j) *Anyone who eats a pound of arsenic has a probability of .95 of dying within twenty-four hours.*
 (k) *Jones ate a pound of arsenic at time t.*

(l) *The probability of Jones dying within twenty-four hours of t is .95.*

(m) *(Jones did die within twenty-four hours of t .)*

Now we introduce the truck striking Jones. If the probability of Jones dying due to truck impact is 0, then the traffic accident is explanatorily irrelevant and the above DNP explanation remains valid. On the other hand, if the probability is non-zero, then (l) is false because the probability of Jones' dying is no longer .95. Thus the above DNP explanation is invalidated. A valid deductive argument with all true premises must yield a true conclusion. So the falsity of (l) implies the violation of one or both of the criteria of truth. Since the form of the argument is valid, at least one premise must be false; we agree with Railton (personal correspondence) that the law (j) is the culprit. There is a multitude of statistically relevant factors which may intervene during the pertinent interval such that the probability asserted in (j) is not reliable. The truth of (j) may be salvaged by the inclusion of a *ceteris paribus* clause to account for these factors. Even including such a clause, our point stands since the clause itself would rule out the original DNP explanation in light of the new information. In that case, the *ceteris paribus* clause introduces antecedent conditions in the law (j) which are not met, invalidating the inference.²

What is needed for an effective intervening cause counterexample is an intervening cause which is statistically irrelevant but explanatorily relevant. Consider the following twist on Schrödinger's cat (see Figure 1). We have a cat in a plastic box with a metal floor. In the box is a

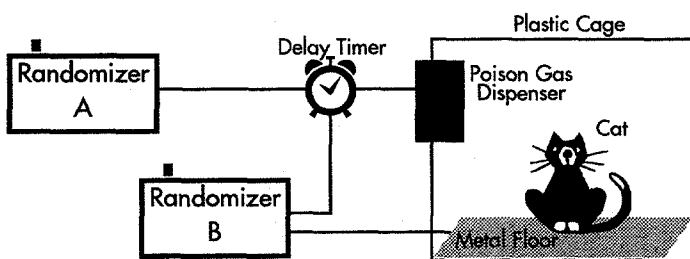


Figure 1. A twist on Schrödinger's Cat

2. Nancy Cartwright (1983) argues that truth and explanatory power are at odds as regards scientific laws because this sort of invalidation will occur in every case. Her solution is to jettison the requirement of truth and with it one of the key underpinnings of the formalist tradition. While this is an interesting view, we choose to argue here against the sufficiency of the DNP model from within its framework.

poison gas dispenser connected to a randomizer device, call it A. When a button on A is pressed, a detector within the device is activated for some specific time period. Within the randomizer is a quantum system with a probability of .5 of decaying during a time period of the same duration. If the detector senses a particle from the decay process, it sends an electrical signal to a delay timer device placed along the connection between the randomizer and the poison gas dispenser. After 11 hours and 50 minutes, the delay timer sends the signal along to the dispenser, which releases the gas, killing the cat. If the detector in the randomizer does not sense a particle, nothing happens.

Also part of this elaborate set-up is a second randomizer of a slightly different sort, call it B, which requires independent initiation. Like A, B contains a quantum system with a .5 probability of decaying during an appropriate time interval and a detector, but this detector also performs a spin measurement on any particle detected. Spin-up and spin-down results are equiprobable. B does nothing if no particle is found. If a spin-up particle is detected, B sends a signal to the delay timer which either deactivates it if it had been previously activated or does nothing otherwise. Provided that a spin-down particle is detected, the same signal is sent to the delay timer, but, additionally, an electrical pulse is sent to the metal floor of the box, electrocuting the cat.

Supposing the button on A has been pushed and the cat dies, we may have the following DNP explanation:

- (n) *A derivation of (o) by appeal to the relevant quantum mechanical and biological theories and the mechanisms of the entire system [including both A and B].*
 - (o) *Whenever the button on an A randomizer in this type of system is pushed at time t, there is a .5 probability of the cat dying within 12 hours of time t.*
 - (p) *The button on A was pushed at time t and there was a cat in the box.*
-
- (q) *The cat had a probability of .5 of dying within 12 hours of time t.*
 - (r) *(The cat did die within 12 hours of time t.)*

Now we introduce the intervening cause with the following additional information. Within 11 hours and 50 minutes of time *t*, the button on B was pressed. A spin-down particle was observed by the detector in B. The delay timer, which had been activated by a signal from A, was deactivated by another signal from B.³ B also sent an electrical pulse which electrocuted the cat.

3. A similar version of the counterexample can be formulated in which no particle is detected in A so that the delay timer is never activated. This version is at least as potent a counterexample as well, although, strictly speaking, the nomenclature “intervening

Notice that, in contrast to the Jones example, *ceteris paribus* concerns are irrelevant. Obviously, (o) does not explicitly contain a *ceteris paribus* clause, and to be a true law, it may require the presence of such a rider, though perhaps implicitly. But these concerns do not effect the validity of this explanation because the process associated with B is accounted for in the derivation (n) so that it is part of the core content of (o) and is not an external factor of the sort for which a *ceteris paribus* clause is meant to account.

With these concerns dispelled, we contend that the above case constitutes an intervening cause counterexample. The proposed explanation continues to satisfy all of the requirements for an acceptable DNP explanation; the probability cited in the conclusion (q) remains correct, the law (o) is true, and the inference is valid, as the derivation (n) takes into account the possibility of B's activation. However, it is clearly not the right explanation since it turns upon the activation of A, which played no role in the cat's death, whereas the pushing of the button on B, the real cause, is nowhere mentioned. The cat, after all, was electrocuted, not poisoned. So as with the other models in the tradition, the DNP model of explanation is susceptible to intervening cause counterexamples. Thus it fails to provide sufficient conditions for scientific explanation and is unable to serve as an ideal explanatory text.

REFERENCES

- Achinstein, P. (1983), *The Nature of Explanation*. New York: Oxford University Press.
 Cartwright, N. (1983), *How the Laws of Physics Lie*. New York: Oxford University Press.
 Hempel, C. (1965), *Aspects of Scientific Explanation*. New York: Free Press.
 Railton, P. (1978), "A Deductive-Nomological Model of Probabilistic Explanation", *Philosophy of Science* 45: 206–226.
 ———. (1981), "Probability, Explanation, and Information", *Synthese* 48: 233–256.
 Salmon, W. (1990), *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press.

cause counterexample" is not appropriate, as there would be no antecedent "cause" to be preempted.