Simon Fraser University Undergraduate Journal of Philosophy, Volume 1, Issue 1. 2017.
*Counterfactual Accounts and Equilibrium Explanations*


# Counterfactual Accounts and Equilibrium Explanations

Fernando Villasenor, Robert Munro

*What part should counterfactuals play in our scientific explanatory accounts? In this paper we examine James Woodward's role of counterfactual explanation, Christopher Hitchcock and James Woodward's view of counterfactual explanatory depth, and a possible extensions to the equilibrium explanation proposed by R.A. Fisher and Elliot Sober. In the argumentative portion of the paper, equilibrium explanation is examined and an attempt is made to bridge the two accounts.*

**Keywords: Counterfactual, Equilibrium, DN Model, Scientific Explanation**

Counterfactual questions ask "what-if?" What if the circumstances had been different? How would what followed be different, if at all? A first-run account of Scientific Explanation, the Deductive-Nomological (DN) Model, made no explicit reference to counterfactual ideas in its formulation. Instead the DN model claimed that to provide an explanation was to provide a deductively valid argument with the phenomena we wished to explain as its conclusion.[1] An essential feature of the DN model was the utilization of a law of nature in its premises. This account of explanation did not make use of a notion of causation; instead positing that there was something inherent to the argument structure it provided which would entail an explanatory account.

> Consider the following two examples of DN explanations given by James Woodward[2]:
> An explanation for a Raven's being black:
>
> i) All ravens are black.
> ii) x is a raven.
> iii) ∴ x is black.

An explanation for the magnitude of the electric field created by a long straight wire with a positive charge uniformly distributed along its length.:

> Given the differential form of coulomb's law: $dE = (1/4\pi\varepsilon)(dq/s^2)$ and via integration over each individual contribution (over each infinitely small dq) yields the result that the field is at right angles to the wire and has intensity given by $E = (\lambda)/(2\pi\varepsilon^* r)$ where

[1] Carl Hempel and Paul Oppenheim, *Studies in the Logic of Explanation* (Philosophy of Science).
[2] James Woodward, Making Things Happen, (Oxford University Press, 2003), 187.

r is the perpendicular distance to the wire and the charge density along the wire.

Argument 1 uses the generalization that all ravens are black as its law-like premise, while argument 2 uses Coulomb's law for its law-like premise. Both arguments fall under the umbrella of the DN model, but, as Woodward notes, argument 2 would strike most readers as a deeper or more satisfying explanation than argument 1. Unlike argument 1, argument 2 has a built in capability to answer what-if questions. It has the capacity to answer how the conclusion would change under different initial values. It is a relationship in which, had there been some manipulation of the initial starting conditions, it would have elicited a different effect. This is precisely Woodward's proposal: explanation is a matter of exhibiting counterfactual dependence.

Woodward's view overlaps with David Lewis' account. Both treatments assign importance to the notion of counterfactual dependence in elucidating causal notions. But first of all, what is counterfactual dependence? And what does it mean for a counterfactual to be true? In Lewis's treatment of counterfactuals, he stipulated that the counterfactual 'If A had been the case, then B would have been the case' would be non-vacuously true just in case it were to take less of a departure from actuality to make A true along with B true than it would to make A true without B.[3] To handle these similarity orderings Lewis devised an ordering of possible worlds based on their similarity to ours.[4] [5]The similarity conditions consisted (roughly) in avoiding big violations of the natural laws, and maximizing the regions throughout which a perfect match of the facts prevail in our world and the world being compared.

For Lewis, given events c and e, and letting O(c) and O(e) be the propositions that events c and e occur, then e is counterfactually dependent on c (e depends causally on c) iff the counterfactuals (1) and (2):

1. $O(c) \rightarrow O(e)$
2. $\sim O(c) \rightarrow \sim O(e)$

both hold true. One key location where Lewis' view and Woodward's view differ is that Lewis's goal in devising his similarity ordering of worlds was to make the counterfactuals true that we expected to be true. On Woodward's account of explanation, no such predetermined orderings exist. Woodward's account of causation instead provides a natural answer of how we may go about determining the truth levels of these claims. Often we see attempts to describe relationships between variables in the (simplified) form $Y = bX + U$, where b stands for the coefficient effect on X that produces an effect in Y, with U representing the unaccounted for causes of Y beyond X. Of special interest to us, remarks Woodward, are when these relationships contain an autonomous structure, where

[3] David Lewis, Counterfactuals, (Cambridge, Harvard University Press, 1973).
[4] Ibid.
[5] David Lewis, Philosophical Papers, Volume II (New York, Oxford University Press, 1987).

'autonomy' is the degree within which the relationship remains invariant under various possible manipulations or 'interventions'.[6]

Another similarity between Lewis' and Woodward's account of causal explanation are that explanations are not binary affairs as the DN model suggests, but rather something which can be possessed in differing degrees. One of Lewis' objections to the DN account is that the DN model only provides a 'small cross section of the causal history [of an explanandum]'.[7] Lewis believed that a deeper explanation consisted in providing additional information on antecedent causes in the causal chain leading up to the explanation event. Woodward and Hitchcock propounded a different notion of depth. For them, depth corresponded to the invariance present in the explanation, or in other words, the amount of different 'what-if' questions the account was capable of answering.[8] In this sense explanation isn't tied up in the occurrence of a universal law, as it is possible to proffer a good explanation in the absence of a law but in the presence of a good generalization holding invariant over a wide class of changes. The strength of explanations concerning efficacy of treatment interventions on ill patients, or effects of positive news coverage on number of votes for political candidates can be evaluated using these criteria.

Thus the interventionist-counterfactual account claims to expand on the DN model, with explanations evaluated not only on their ability to provide a link from explanans to explanandum, but a link that is possible to exploit due to stability over a range of counterfactual conditions. In this manner, criteria for counterfactual explanations as detailed above gives explanations without needing law-like premises to hold universally, which seems to correspond with our intuitions of the existence of explanations without law-like generalizations in science.

**Equilibrium and Counterfactual Explanatory Depth: Basics**

Woodward and Hitchcock propose that explanatory depth lies in the depth of a given generalization's strength, otherwise characterized as its invariance. Generalizations don't have to be 'universal' but they must be stronger than mere 'accidental' generalizations, e.g. 'All bald schoolteachers in district 7 have syphilis'. A decent (if somewhat shallow) example of a generalization that does provide explanatory power of this type could be the case of a flower's growth. Given Y: the height of the flower, $X_1$: amount of fertilizer, and $X_2$: amount of soil, we could set up a reasonably general explanation accounting for the height of the flower based on the inputs of fertilizer and soil along with some random error input U (accounting for effects on the height outside of soil or fertilizer):

---

[6] James Woodward, Causation and Explanation in Econometrics: *On the Reliability of Economic Models* (1995).
[7] David Lewis, *Philosophical Papers Volume II* (New York, Oxford University Press, 1987).
[8] Christopher Hitchcock and James Woodward, *Explanatory Generalizations, Part II: Plumbing Explanatory Depth* (Nous 37, no. 2, 2003)

(1) $Y = aX_1 + bX_2 + U$

Under Woodward's and Hitchcock's account, (1) provides a reasonable account for explanation since it holds over a range of alternative conditions. Presumably Y would be explainable within some space ($c < X_1 < d$, $e < X_2 < f$ where c, d, e, f fall in the set of the positive Real Numbers). Soil and Fertilizer aid in a flower's growth but not indefinitely. At some point adding soil or fertilizer would no longer aid in the flower's growth (for some value of d where $X_1 > d$ – for some presumably large value d – the relationship would fail to hold). Soil is necessary for flower growth, but too much and perhaps the flower would no longer have access to sunlight.

Clearly (1) is generalizable but it has its limits. But (1) would be a considered a somewhat shallow explanation for flower height when compared to some alternate explanation that took into account more relevant factors holding over a wider range of instances. Perhaps a more detailed account concerning the flower's biology would hold true over a deeper range of conditions and as such would be classified as a 'deeper' explanation.

Now consider the case of equilibrium explanation proposed by Elliot Sober. (Sober, 1982) R.A. Fisher's equilibrium explanation stated that if a population were ever to depart from equal numbers of males and females, that there would be a reproductive advantage favouring parental pairs overproducing the minority sex.[9] The equilibrium explanation exhibits very high amounts of generalization strength.

> "Equilibrium explanation shows how the event would have occurred regardless of which of a variety of causal scenarios actually transpired."[10]

Sober argues that by situating these causal trajectories in a deeper more encompassing structure, we are able to provide explanations that are more explanatory even though they provide less information about the cause. Regardless of where the population may have lied when equilibrium conditions were disturbed, the system would have the potential to return to equilibrium through a wide variety of possible causal paths. It is this knowledge, and not knowledge of any singular causal event providing the explanatory power here.

What separates equilibrium, and counterfactual explanation? Woodward's account seems to rely on an exclusively causal account and the degree to which this causal account can be generalized, emphasizing differing levels of independent variable treatments, which in turn reflect or predict differing levels of the dependent variable to be explained. The equilibrium account on the other hand seems insistent to avoid referencing any exclusive (or singular) causal events. Sober's equilibrium explanation purports to exhibit explanatory

[9] Elliott Sober, *Equilibrium explanation*, (Philosophical Studies 43, no.2, 1983).
[10] Ibid., 202.

depth through the sheer irrelevance of a range of antecedent states. This, Sober argues, enables equilibrium explanation to supersede an explanation of any individual causal trajectory.

Perhaps a feature of equilibrium explanation that makes it enticing to differentiate from the interventionist account is the 'structure' or the 'system' which encompass this far ranging dependence. In an equilibrium explanation it doesn't seem to be necessary or even desirable to speak of 'interventions,' because most of these interventions would be inconsequential to the return to equilibrium.

However, what of the fact that the explananda of equilibrium explanations can be represented as a binary affair? Consider the following mapping f of some arbitrary state of a system A containing a population with two sexes that prevails after an initial disturbance from its 1:1 equilibrium. Let:

> f(A) = 1 if the population settles back into equilibrium following some arbitrary disturbance, and
> f(A) = 0 otherwise.

Save for very few (probably very large) interventions, sex equilibrium will return to 1:1, usually resulting in the mapping of f to 1. Throughout a large range of causal chains we end up with a sex equilibrium returned to 1:1. The explanation seems to be so broad that it seems silly to think of it as explainable under an 'interventionist' account, where almost any change in antecedent conditions ends in the return to equilibrium. But, just as we re-conceptualized the explanation of the flowers height to the presumably more final reproductive cause of the flower, we can also step back from the 'final' 1:1 sex equilibrium and more closely examine what path brought us there.

But perhaps this is a restrictive mapping of the explananda. Consider the explanation accounting for the height of the flower (1) $Y = aX_1 + bX_2 + U$. It is clear in this case that the explanandum variable Y in (1) could be equal to a much more diverse set of numbers than 0 and 1. Y could be any number on the real interval [0, x] where x is some reasonable upper limit on the height of the flower's growth.

**Can the Two Accounts be Bridged?**

We suggest that the equilibrium explanation be thought of as a special case counterfactual explanation – a case in which the outcome has reached a sufficient level of invariance under changes to initial conditions. The explanation holds throughout a large amount of varying antecedent states, given the satisfaction of certain fairly broad conditions. Sober argues that equilibrium explanations account for the return to a 1:1 sex balance are more explanatory than any single tracing of a particular causal path back to 1:1 sex balance.

Even though the equilibrium explanation 'fails' to account for specific causal trajectory of the event, it isn't necessary when providing explanations of this sort. Answers to what-if questions don't seem to be particularly conducive to understanding as they all lead us to the same result.

Maybe different types of disturbances produce different paths to 'realignment' with the equilibrium? It seems easy to imagine cases where we would like to know how fast or slow the population would realign. What about disturbances that would result in a return to equilibrium but in an oscillating fashion as opposed to an asymptotic or linear return? Each different return would carry the potential for interest outside of the equilibrium discussion. Maybe different types of returns to equilibrium would exhibit vastly different interactions with the environment proving to be of considerable interest to people in a relevant field of study. There seem to be many different instances of 'return to equilibrium' that aren't elucidated by simply asserting the return to equilibrium.

Just as we reformulated the explananda in our explanation of flower height (1) to map exclusively onto the set 0,1, we can break apart the equilibrium explananda and divide different causal paths to the equilibrium and differentiate them by their respective mappings.

Perhaps some explanatory model in the form of:

(2) $Y = a_1Z_1 + a_2Z_2 + \ldots + a_iZ_i + U$

could account for different possible returns to the 1:1 equilibrium, where the Z's are all factors contributing to initial disturbances with coefficients reflecting the degree to which these factors influence the type of return to equilibrium. Unlike the 1:1 equilibrium explananda, our newly minted Y in explanation (2) includes additional information about the causal pathway thus now expressed as dependent on the antecedent conditions, meshing much more nicely with Woodward's interventionist framework.

Do equilibrium explanations provide more depth with binary explananda or with many different realizable explananda? The answer seems to depend on the questions we ask. If all that is important to the person requesting an explanation be the equilibrium event itself, then we would grant that it is neither necessary nor desirable to provide an account tracing the token causal trajectory. However, as argued above, perhaps it is possible to re-conceptualize flower growth as having a binary explanation, but still this doesn't make other explanations with more specificity irrelevant. Likewise, considering the equilibrium explanation, sometimes we may wish to know more about the equilibrium event than its simple return to sex balance. Given what we believe to be the compatibility of the two accounts, there is a sense in which equilibrium explanations are best thought of as one case

of a more general counterfactual analysis.

## Bibliography

Fisher, Ronald A. "The genetical theory of natural selection." 1930. doi:10.5962/bhl.title.27468.

Hempel, Carl G., and Paul Oppenheim. "Studies in the Logic of Explanation." *Philosophy of Science* 15, no. 2 (1948), 135-175. doi:10.1086/286983.

Hitchcock, Christopher, and James Woodward. "Explanatory Generalizations, Part II: Plumbing Explanatory Depth." *Nous* 37, no. 2 (2003), 181-199. doi:10.1111/1468-0068.00435.

Lewis, David K. *Counterfactuals*. Cambridge, Mass: Harvard University Press, 1973.

Lewis, David. *Philosophical Papers Volume II*. New York: Oxford University Press, 1987.

Lewis, David. "Causation." *The Journal of Philosophy* 70, no. 17 (1973), 556. doi:10.2307/2025310.

Sober, Elliott. "Equilibrium explanation." *Philosophical Studies* 43, no. 2 (1983), 201-210. doi:10.1007/bf00372383.

Woodward, James. *Making Things Happen*. Oxford University Press, 2003.

Woodward, James. "Causal Interpretation in Systems of Equations." *Synthese* 121 (1999), 199-247.

Woodward, James. "Causation and Explanation in Econometrics." *On the Reliability of Economic Models*, 1995, 9-61. doi:10.1007/978-94-011-0643-6_2.