Mechanisms and Functional Hypotheses in Social Science

Daniel Steel†

Critics of functional explanations in social science maintain that such explanations are illegitimate unless a mechanism is specified. Others argue that mechanisms are not necessary for causal inference and that functional explanations are a type of causal claim that raise no special difficulties for testing. I show that there is indeed a special problem that confronts testing functional explanations resulting from their connection to second-order causal claims. I explain how mechanisms can resolve this difficulty, but argue that this does not provide support for methodological individualism since it is not necessary that the mechanisms be described in terms of individual interactions.

1. Introduction. Functional explanations are usually understood to account for the presence or prevalence of a feature by reference to its effects. Although functional explanations have long been an accepted part of evolutionary biology, in social science they are often objects of suspicion. One common claim is that strong evidence for functional explanations can be provided only if the mechanism through which the allegedly beneficial effect leads to the prevalence of the cause is described, and that in social science such mechanisms are frequently left unspecified (cf. Elster 1983, 61; Little 1998, 6–7). In contrast, others deny that mechanisms are necessary for causal inference and argue that functional explanations, properly understood, are a type of causal hypothesis and as such present no special problems for testing that are not also present with regard to ordinary (i.e. non-functional) causal claims (cf. Cohen 1978; Kincaid 1990, 1996). Since the mechanisms in question are generally understood to involve the interaction of individuals, this dispute is one facet of the methodological individualism-versus-holism debate in the philosophy of social science. That is, advocates of mechanisms see their arguments concerning functional explanations as a buttress for methodological individualism, while the critics of these arguments are motivated in large measure by a

†To contact the author, please write to: Department of Philosophy, 503 S. Kedzie Hall, Michigan State University, East Lansing, MI 48824-1032; e-mail: steel@msu.edu.

Philosophy of Science, 72 (December 2005) pp. 941–952. 0031-8248/2005/7205-0024$10.00

desire to undermine this position. In this essay, I inquire into whether testing functional hypotheses in social science presents special difficulties that do not arise for causal hypotheses in that domain generally, and, if so, whether mechanisms are useful for ameliorating these difficulties.

I begin by describing the relevant interpretation of functional explanations, according to which they are distinguished by the possession of a second-order causal claim.[1] A second-order causal claim is one in which the presence of a causal relationship is itself said to be a cause of some thing else. To take a classic example, blackness is a cause of moths evading predation in soot-encrusted woods, and this relationship in turn is a cause of the prevalence of black moths in such circumstances. This case illustrates the type of second-order causal claim that will concern us here: *A*'s causing *B* is itself a cause of *A*'s prevalence or persistence. I show there is indeed a difficulty that confronts efforts to test second-order causal claims, and hence functional explanations, that does not arise in the case of ordinary causal claims. I then inquire into whether, and if so how, attention to underlying mechanisms can mitigate this problem. I argue that mechanisms can be useful for this purpose, but I suggest that this does not provide support for methodological individualism, since the mechanisms in question need not be stated in terms of interactions among individual agents.

2. Functional Explanations and Causes. In his essay "The Logic of Functional Analysis" Carl Hempel considered whether functional explanations can be interpreted in terms of the covering law model, and if not, whether they are a legitimate alternative form of scientific explanation at all. Functional explanations resisted accommodation in Hempel's model due to possessing the following feature:

> (F) Functional explanations explain the presence of an entity by reference to its effects.

Given (F), functional explanations, if understood as causally, seem to run afoul of the principle that a cause must precede its effect. Moreover, as Hempel argued, it is difficult to interpret functional explanations by means of the covering law model, since it generally possible that several different entities could have brought about the beneficial effect in question (1965, 310–311). Thus, the presence of the entity cannot be inferred from the supposed need for the beneficial effect.

Probably the most common response to Hempel's objection has been to accept (F) while arguing that this feature of functional explanations

1. I borrow the expression "second-order causal claim" from Hitchcock (1996).

can nevertheless be made consistent with a causal interpretation (cf. Wright 1976; Little 1991; Kincaid 1990, 1996; Hitchcock 1996).[2] The shared idea of these proposals is that, "*A* functions to do *B*," means that *A* is a cause of *B*, and that *A* is present or persists because it has this effect. For example, according to Kincaid, "The function of *A* is *B*," should be understood to mean:

(1) *A* causes *B*.
(2) *A* persists because it causes *B*. (1990, 344).

Since the purpose of this proposal is to show how functional explanations can be understood as a special type of causal explanation, the "because" in (2) must be understood to mean "causes" or "is a cause of." Hence, (2) is a second-order causal claim: *A*'s being a cause of *B* is a cause of *A*'s persistence.

Kincaid's schema encompasses several sorts of explanations. For example, it covers intentional explanations like, "Physicians recommend penicillin for pneumonia because of its curative effects against this bacterial infection." Here treatment with penicillin causes recovery from pneumonia, and this fact causes physicians to persist in recommending penicillin as a treatment for this ailment. Explanations that invoke natural selection, or some analogue, can also be encompassed by Kincaid's schema. For example, when trees in forests near Manchester became blackened with factory soot, blackness among moths was a cause of evading predators, and black moths came to predominate in the population because of this effect. Some causal feedback loops that one would hesitate to call "functional" are also covered by Kincaid's schema. For instance, heroin use causes addiction, and heroin use persists because it causes addiction.[3] Such examples suggest that one aspect of the commonsense understanding of functional explanations is that the allegedly functional item or property is beneficial in some sense (cf. Elster 1983, 57). However, the question of whether functional explanations are inherently bound up with a concept of beneficial effects is independent of the issue of concern here, which is whether functional explanations pose special difficulties for causal inference.

Given Kincaid's proposal, a functional explanation consists of a conjunction of an ordinary causal claim (1) and a second-order causal claim (2). Clearly, the inclusion of (1) entails that functional explanations inherit

2. But see Cummins (1975) for a different approach.

3. Kincaid (1996, 111) adds an additional condition to his schema requiring that "*A* is causally prior to *B*." This means that the causal loop connecting *A* and *B* cannot be initiated with *B*. However, this condition is fulfilled in the heroin example, since heroin addiction must be preceded by heroin use.

all of the difficulties surrounding the underdetermination of ordinary causal hypotheses. It will be useful to briefly indicate one of the most important difficulties of this sort. It is a familiar slogan that correlation does not imply causation: a probabilistic dependency between two variables A and B might result from A being a cause of B, B being a cause of A, or the presence of one or more common causes of A and B. Time ordering of variables can sometimes eliminate one of the first two possibilities, but does not rule out the third. Let us use the expression the *problem of confounders* to refer to the possibility that a probabilistic dependence between two variables is explained by a common cause rather than by one of the variables being a cause of the other.

Two common strategies for dealing with the problem of confounders are controlled experiments and conditioning on potential common causes of the variables in question. However, controlled experiments are rarely a possibility in social science, and it is also rare that all potential confounders can be measured.[4] In light of such difficulties, it is sometimes claimed that causal inference in social science is possible *only* given detailed knowledge of the underlying micro-mechanisms (cf. Elster 1983; Little 1991, 1998). If this claim were true, then it would follow that functional explanations cannot be firmly established except when knowledge of mechanisms is available. There are reasons to be skeptical that causal inference in social science is *never* possible without mechanisms.[5] However, the important point for our purposes is that the problem of confounders is not a special problem for functional explanations; rather it is a challenge that confronts causal inference generally. Consequently, if there is some especially problematic characteristic of functional explanations that does not face all causal hypotheses in social science, then it must arise from the second-order causal claim.

3. Testing Second-Order Causal Hypotheses. Consider the following claim: racial discrimination functions to maintain a pool of unskilled, cheap labor. The thought behind this hypothesis is that one consequence of racism is that the ethnic minority in question is generally denied access

4. There are other drawbacks to the method of conditioning on all potential confounders. In particular, this method can lead to incorrect conclusions if some potential confounders are in fact intermediate causes or common effects (cf. Spirtes, Glymour, and Scheines 2000, chapter 8).

5. Most significantly, Pearl (2000) and Spirtes, Glymour, and Schienes (2000) describe a theory of causal inference that identifies conditions in which one could reliably infer that one variable is a cause of another even when experiment is not possible and one does begin with the assumption that all potential common causes have been measured. For a discussion of mechanisms and the problem of confounders in social science see Steel (2004).

to adequate education as well as to stable, well-paying jobs, which has the consequence that members of the minority group suffer high rates of unemployment and underemployment. These pools of unemployed people, then, are said to be advantageous from the point of view of agricultural or industrial corporations, for whom a ready supply of cheap labor is an essential ingredient of profitability.

Rendered in Kincaid's schema, the explanation would take a form like the following:

(3) Racial discrimination is a cause of a ready supply of cheap labor.
(4) The fact that racial discrimination is a cause of a ready supply of cheap labor is a cause of the persistence of racial discrimination.

There would clearly be significant challenges to assessing either of these two causal claims. In addition to the problem of confounders, there is the difficulty of finding an objective and accurate procedure for measuring racial discrimination. But these are typical examples of difficulties that confront causal inference in social science generally. Is there any further problem arising specifically from the second-order causal claim (4)? I suggest that there is.

To see the problem, consider the first step that is often taken to test a claim of the form A is a cause of B, namely, to collect data to decide if A and B are probabilistically dependent. In the present example, let R represent racial discrimination and L the supply of cheap labor, and let us suppose that an acceptable procedure for measuring these variables has been devised. For example, in the present case it might be reasonable to use educational and employment opportunities of the minority group as a proxy for discrimination. The causal claim in (3), then, predicts that there is a positive association between R and L. Of course, even if the data confirms this prediction, it cannot be immediately inferred that (3) is true, since it might be instead that L is a cause of R, or that there is a common cause of both. However, with respect to (3) the initial steps of the testing process are relatively straightforward, even if it is unclear, for instance, how the problem of confounders will be dealt with.

In contrast, consider (4). Following the approach taken with regard to (3), we would first inquire into whether there is a positive association between racial discrimination being a cause of cheap labor and the persistence of racial discrimination. Let C be a binary variable that indicates whether racial discrimination is a cause of cheap labor ($C = 1$ if the causal relationship holds, $C = 0$ otherwise)[6] and P indicate whether racial discrimination persists (P could be interpreted as representing the change in

6. Compare with Kincaid (1990, 345).

R over some specified period of time). Thus, (4) predicts that there is a positive association between C and P. The process of collecting data to decide whether this is indeed the case would presumably work something like the following. Within a sample of communities, one would distinguish those communities in which $C = 1$ from those in which $C = 0$; then one would collect data on P to decide whether racial discrimination was more persistent in the former communities than in the latter.

It is easily seen that this scenario raises a problem not encountered in the case of (3): in each instance, the measurement of C is an inference concerning the claim that R is a cause of L. As such, each of these measurements will face all of the problems that confront causal inferences, such as the problem of confounders. Moreover, making these inferences in a reliable fashion will be made more difficult by the necessity in each case to confine attention to a smaller sample of data (a specific community). In short, deciding whether there is a positive association between C and P is not a relatively straightforward process of collecting data and performing a statistical test; rather, it requires a causal inference for each community in the sample. In contrast, testing an ordinary first order causal claim like (3) need not contend with this problem for the simple reason that the variables in first order causal hypotheses do not represent whether or not a causal relationship obtains.

The above considerations, then, reveal that there is indeed a challenge confronting the testing of functional explanations that does not arise in the case of ordinary causal hypotheses. Functional explanations include a second-order causal claim, and any second-order causal claim contains a variable whose measurement requires a causal inference concerning a first-order causal claim. For convenience, let us refer to this as the *causal measurement problem*. The causal measurement problem, then, immensely complicates the process of collecting statistical data to decide whether probabilistic dependencies predicted by the second-order causal claim obtain.

4. Do Mechanisms Help? Critics of functional explanations in social science who emphasize the importance of mechanisms have not, to my knowledge, been motivated by the causal measurement problem. Rather, their concern is that unless details concerning underlying mechanisms are provided, functional explanations are little more than just-so stories. As Daniel Little puts it:

> It is almost always possible to come up with some beneficial consequences of a given institution; so in order to justify the judgment that the institution exists because of its beneficial consequences we need to have an account of the mechanisms which created and re-

> produced the institution which shows how the needs of the system as a whole influenced the development of the institution. (1998, 6)[7]

However, if it is easy to specify some supposedly beneficial consequence of a social practice, it is not much harder to invent hypotheses concerning underlying mechanisms. For example, in the racism example described above, one could suppose that the agricultural or industrial corporations who allegedly benefit from racial discrimination actively lobby local, state, and federal governments to support policies that perpetuate discrimination and oppose policies that would have the opposite effect (though they publicly deny that their lobbying efforts have this goal).

Of course, an advocate of mechanisms would likely respond that it is not merely sufficient to propose a potential micro-mechanism; it is also necessary to provide evidence for it. But one who disputes the necessity of mechanisms for establishing functional hypotheses would point out that the same is true for claims about beneficial consequences (cf. Kincaid 1996, 116–118). It is not enough to say that the institution or practice in question has some effect, and the fact that it has that effect causes it to persist. Evidence must be provided for these claims. The issue, then, is whether the requisite evidence must involve an account of the underlying mechanisms: is there any problem specific to testing functional explanations for which mechanisms provide a particularly helpful solution? Given the discussion in the foregoing section, this question devolves into a consideration of whether attention to mechanisms can help to resolve the causal measurement problem. Although advocates of the value of mechanisms in social science have not addressed this question, I suggest that the answer is 'yes'.

One strategy for dealing with the causal measurement problem is to reformulate the functional explanation in such a way that the troublesome second-order causal claim is avoided. In particular, specifying a mechanism can enable one to transform the second-order causal claim into a series of first-order claims. Consider again the case of the black and grey moths. In Kincaid's format, the relevant causal claims could be put as follows (where these claims are understood in reference to the population of moths in the soot darkened Manchester forests):

(5) Blackness is a cause of survival to sexual maturity.
(6) The fact that blackness is a cause of survival to sexual maturity is a cause of the increasing prevalence of black moths.

Instead of attempting to test the second-order causal claim (6) directly, the commonsense way to proceed is by considering the following claims.

7. See Elster (1983, 55–68) for an expression of the same sentiment.

In the environment in question, the grey moths are more likely to be eaten by birds than the black moths, and hence black moths are more likely than grey ones to survive to sexual maturity. Given that the proportion of black moths in the descendent generation is a function of the gene pool of moths surviving to sexual maturity in the parent generation and that black moths are more likely to pass along genes for black coloration to their offspring, the survival of each black moth positively influences the relative frequency of black moths in the next generation. If the causal claims in this natural selection scenario are true, then the second-order causal hypothesis (6) seems to follow straightaway.

The key to understanding more generally and precisely how the second-order causal claim can be avoided lies in explicating the claim that a given property persists or increases in prevalence. Let B indicate blackness. The persistence or prevalence of B can be understood as a rough way of describing the change in the proportion of B in the population over time. Let us think of the "proportion of B" in probabilistic terms, so that the proportion of B is interpreted as the probability that a randomly chosen individual of the population is B. Suppose that we are concerned with the change in the proportion of B from generations n to $n + 1$. Let P_n and P_{n+1} be the probability functions for the populations at n and $n + 1$, respectively. Then the change in B from n to $n + 1$ can be defined straightforwardly as follows: $\Delta P(B) =_{\text{def}} P_{n+1}(B) - P_n(B)$. The claim "$B$ persists" could then be interpreted to mean that $\Delta P(B) \geq 0$; the claim "B becomes more prevalent" could be understood to mean that $\Delta P(B) > 0$, and so forth.

The initial premise in the moth example was that the color of the moth was a cause of avoiding predation. Moreover, this causal relationship was assumed to produce a positive probabilistic dependence between dark coloring and survival. Let S indicate survival to reproductive maturity. Then in the abstract, the reasoning is the following:

(a) B is a cause of S in n; therefore,
(b) B and S are positively associated in n (i.e., $P_n(B|S) > P_n(B)$).

Suppose that we desire to explain why B increased in prevalence. Hence, from $P_n(B|S) > P_n(B)$ we want to derive the conclusion $\Delta P(B) > 0$, that is, $P_{n+1}(B) - P_n(B) > 0$. There are several potential bases for this inference, depending on the specifics of the case at hand. One possibility would be to provide evidence for the assumption that $P_{n+1}(B) = P_n(B|S)$. For instance, this assumption would be appropriate if the members of generation $n + 1$ consist exactly of descendants of members of generation n that survived to sexual maturity, B is related to fitness only in virtue of its effect on S, and S is perfectly heritable.

In the scenario just described, the fact that B is a cause of S makes a

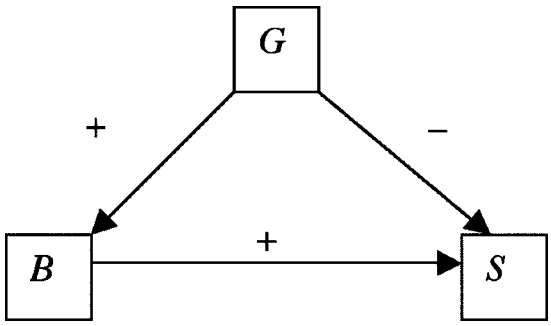

Figure 1.

difference to $\Delta P(B)$; were B not a cause of S, $\Delta P(B)$ would have been diminished. It is on the basis of such considerations that it is reasonable to encapsulate the natural selection account with the claim that B's being a cause of S is itself a cause of $\Delta P(B)$. Thus, a second-order causal claim is sometimes an abbreviated way of describing a more detailed explanation that does not involve second-order causal claims.

Let us consider the role that mechanisms play in the abstract scenario described above. Most apparently, an inquiry into the workings of the selective mechanism in the given case is necessary for the step from $P_n(B|S) > P_n(B)$ to $\Delta P(B) > 0$. The inference from (a) to (b), on the other hand, seems rather different. Given some probabilistic analyses of causation (e.g., Suppes 1970; Eells 1991), it might be taken for a conceptual truth. However, it is possible for B to be a (contributing) cause of S, and yet for B and S to not be positively associated. For instance, suppose that the gene G for blackness in the moths also is a cause of some deleterious developmental abnormality. This situation could be represented in diagram form as in Figure 1. In the diagram, the arrows indicate the relationship "is a cause of," and the plus and minus signs, respectively, indicate whether the cause promotes or inhibits the effect. In such a case as this, whether there is a positive or negative association between B and S depends on the strengths of the causal relationships. A similar situation would arise if A influenced B through multiple, conflicting paths, for instance, if the trait contributed to the avoidance of predators but inhibited food gathering abilities. Given the complex relationships among genes, phenotypes, and reproductive success, such scenarios cannot be dismissed out of hand as merely implausible, theoretical possibilities. Examination of mechanisms, then, may be relevant to the justification of the inference from the claim that B is a cause of S to the conclusion that B and S are positively associated.

The strategy of transforming the second-order causal claim into a set of first-order hypotheses can also be employed in social science. The most straightforward means of accomplishing this is to hypothesize a mechanism analogous to natural selection. For example, one area of research in sociology known as organizational ecology does exactly this. Organizational ecology attempts to explain certain features of particular types of organizations—labor unions, corporations, churches, etc.—in distinct contexts on the basis of differential mortality and founding rates. One important thread in this literature examines the distinct environments to which generalist and specialist organizations are best suited, for instance, inquiring into the conditions in which consolidation among generalist organizations creates resource opportunities for specialists.

It is less clear that a mechanism that appeals to intentional action would succeed in avoiding second-order causal claims. For instance, consider the hypothesized mechanism in the racism example. This proposed that the fact that racism is a cause of cheap labor is a cause of a belief to this effect on the part of the capitalists, which in turn led them to take actions that contribute to the persistence of racism. Yet this involves a second-order causal claim: the fact that racism is a cause of cheap labor is hypothesized to be a cause of the belief that this is indeed the case. However, the second-order causal claim in this case might be avoided by supposing that the direct cause of the belief in the causal relationship is not the causal relationship itself but the probabilistic dependency that results from it. Given this maneuver, the second-order causal claim in the racism example might be avoided in the manner described above.

5. Mechanisms and Methodological Individualism. The foregoing discussion proceeded with no explication of what precisely was meant by the term 'mechanism'. For present purposes, mechanisms can be understood as the process (or processes) through which the cause influences its effect, as was illustrated in the moth and the racism examples. In the context of the philosophy of social science, an emphasis on mechanisms is often closely tied to methodological individualism, the position that explanations in social science should be stated in terms of interactions of individual agents (cf. Elster 1989; Little 1998). Does the usefulness of mechanisms with regard to the causal measurement problem, then, provide support for methodological individualism? I suggest that it does not.

Everyone will agree that social mechanisms inevitably do involve individual interactions. But the fact that societies are composed of interacting agents, and hence that whatever happens ultimately depends on these interactions, does not entail that every adequate *description* of a social mechanism must be phrased in individualist terms. Whether a description of a mechanism is adequate depends on the purpose for which

it is intended. In the present context, we can suppose that the aim is an amelioration of the causal measurement problem: recasting the second-order causal claim so that the functional explanation is made more readily testable. The question, therefore, is whether an individualist description of the mechanism is always necessary for the achievement of this purpose. It is not difficult to argue that the answer to this question is "no."

To see why this is so, consider the example of organizational ecology described above, wherein explanations modeled after those found in evolutionary biology are used to account for prevalent features of organizations in various types of environments. This approach makes a point of deemphasizing the importance of the strategic decision making of organizational managers and leaders on the grounds that organizations typically exhibit a high degree of inertia and that it is often highly uncertain which actions are likely to lead to desired outcomes (cf. Hannan and Freeman 1989, 3–7, Chapter 4). Hence, these authors maintain that changes in prevalence of distinct organizational types is best understood in terms of differential survival rates rather than in terms of adaptive changes to existing organizations instigated by strategic managerial action.

When organizational inertia is strong and the effects of change uncertain, the mechanisms underlying changes in prevalence of distinct organizational types might be adequately described in a way that places little emphasis on individual agency.[8] Moreover, as was explained in the foregoing section, selective models can be used to reformulate functional explanations in a way that avoids the causal measurement problem. The end result is that, although attention to mechanisms can help resolve the causal measurement problem in social science, it is not always necessary that those mechanisms be described in terms of the interactions of individual agents. Hence, mechanisms are useful for ameliorating a challenge that specifically confronts functional explanations, but this fact does not provide an argument for methodological individualism.

6. Conclusion. A common strand in discussions of functional explanations interprets them as containing a second-order causal claim, that is, a hypothesis that asserts that the fact that one thing is a cause of another is itself a cause of something else. Yet there has been little examination of the special difficulties that would arise for testing such causal claims. I have argued that testing functional explanations so understood does indeed raise special challenges not encountered in the evaluation of other types of causal claims, namely, what I called the causal measurement

8. Indeed, Kincaid (1996, 163–166) cites the organizational ecology literature, and Hannan and Freeman in particular, as a case of quality social science that explicitly eschews methodological individualism.

problem. Moreover, I proposed that inquiries into mechanisms can alleviate this difficulty. However, I argued that this conclusion does not provide support for methodological individualism, since it is not always necessary that the mechanisms be described at the level of individual interactions.

REFERENCES

Cohen, G. A. (1978), *Karl Marx's Theory of History: A Defense*. Princeton, NJ: Princeton University Press.

Cummins, Robert (1975), "Functional Analysis", *Journal of Philosophy* 72: 741–764.

Eells, Ellery (1991), *Probabilistic Causality*. Cambridge: Cambridge University Press.

Elster, Jon (1983), *Explaining Technical Change: A Case Study in the Philosophy of Science*. Cambridge: Cambridge University Press.

——— (1989), *Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

Hannan, Michael, and John Freeman (1989), *Organizational Ecology*. Cambridge, MA: Harvard University Press.

Hempel, Carl (1965), *Aspects of Scientific Explanation*. New York: Free Press.

Hitchcock, Christopher (1996), "A Probabilistic Theory of Second-Order Causation", *Erkenntnis* 44: 369–377.

Kincaid, Harold (1990), "Assessing Functional Explanations in the Social Sciences", in A. Fine, M. Forbes, and L. Wessels (eds.), *PSA 1990: Proceedings of the 1990 Biennial Meeting of the Philosophy of Science Association*, vol. 1. East Lansing, MI: Philosophy of Science Association, 341–354.

——— (1996), *Philosophical Foundations of the Social Sciences*. Cambridge: Cambridge University Press.

Little, Daniel (1991), *Varieties of Social Explanation: An Introduction to the Philosophy of Social Science*. Boulder, CO: Westview.

——— (1998), *Microfoundations, Method, and Causation: On the Philosophy of the Social Sciences*. New Brunswick, NJ: Transaction Publishers.

Pearl, Judea (2000), *Causality: Models, Reason, and Inference*. Cambridge: Cambridge University Press.

Spirtes, Peter, Clark Glymour, and Richard Scheines (2000), *Causation, Prediction, and Search*. Cambridge, MA: MIT Press.

Steel, Daniel (2004), "Social Mechanisms and Causal Inference", *Philosophy of the Social Sciences* 34: 55–78.

Suppes, Patrick (1970), *A Probabilistic Theory of Causality*. Amsterdam: North-Holland.

Wright, Larry (1976), *Teleological Explanations: An Etiological Analysis of Goals and Functions*. Berkeley: University of California Press.