When to expect violations of causal faithfulness and why it matters

Abstract: I present three reasons why philosophers of science should be more concerned about violations of causal faithfulness (CF). In complex evolved systems, mechanisms for maintaining various equilibrium states are highly likely to violate CF. Even when such systems do not precisely violate CF, they may nevertheless generate precisely the same problems for inferring causal structure from probabilistic relationships in data as do genuine CF-violations. Thus, potential CF-violations are particularly germane to experimental science when we rely on probabilistic information to uncover the DAG, rather than already knowing the DAG from which we could predict the right experiments to 'catch out' the hidden causal relationships.

1. Introduction

Several conditions must be met in order to apply contemporary causal modeling techniques to extract information about causal structure from probabilistic relationships in data. While there are slightly different ways of formalizing these requirements, three of

the most important ones are the causal Markov, causal modularity, and causal faithfulness conditions. Potential failures of the first two of these conditions have already been the subject of discussion in philosophy of science (Cartwright 1999, 2002, 2006; Hausman and Woodward 1999, 2004; Steel 2006; Mitchell 2008; Woodward 2003, 2010). I will address failures in the third condition, causal faithfulness, and argue that failures of this condition are likely to occur in certain kinds of systems, especially those studied in biology, and are likely to cause a characteristic sort of trouble in experimental settings when using probabilistic relationships between variables to find the causal structure of the system under investigation.

Faithfulness is the assumption that there are no precisely counterbalanced causal relationships in the system that would result in a probabilistic independence between two variables that are actually causally connected. While faithfulness failures have been discussed primarily in the formal epistemology literature, I will argue that violations of faithfulness can impact experimental techniques, inferential license, and issues concerning scientific practice that are not exhausted by the formal epistemology literature.

In particular, a formal methodological perspective might suggest a distinction between genuine and merely apparent failures of CF, such that many apparent examples of CF-violating systems are not 'really' CF-violating. But as I will argue, this distinction is not epistemically justifiable in experimental settings: we cannot distinguish between genuine and merely apparent CF violations unless we already know the underlying causal structure. Without this information, merely apparent and genuine CF violations will be indistinguishable. Violations of CF faithfulness are particularly germane to experimental

science where the underlying causal structure is the subject of investigation, since CF is required to use causal modeling techniques to find directed acyclic graphs compatible with a given set of probabilistic relationships, such that further interventions can determine which is the correct underlying causal structure. Going from relationships in the data to unknown underlying causal structure is the most common direction of inference from the epistemic vantage point of science, and one that will be disrupted equally by genuine and 'merely' apparent CF violations.

This means that failures of CF arguably have the most potential, compared to violations of modularity or causal Markov, for wreaking havoc in experimental settings. They also have interesting methodological consequences for the practice of science: we should expect to find epistemic practices that compensate for CF-violations in fields that study systems where faithfulness is likely to fail. Thus, these conditions are of interest not only to those working on formal modeling techniques, but also to broader discussions in philosophy of science, especially those that concern epistemic practices in the biological, cognitive, or medical sciences.

2. Violations of the Causal Faithfulness Condition

Violation of CF occurs when a system involves precisely counterbalanced causal relationships. These causal relationships are 'invisible' when information about conditional and unconditional probabilities is used to ascertain a set of possible causal directed acyclic graphs (DAGs) that are consistent with data from that system. More precisely:

Let *G* be a causal graph and *P* a probability distribution generated by *G*. *<G, P>* satisfies the Faithfulness Condition if and only if every conditional independence relation true in *P* is entailed by the Causal Markov Condition applied to *G*. (Spirtes, Glymour, and Scheines 2000, 31)

One can think of faithfulness as the converse of the Causal Markov condition: faithfulness says that given a graph and associated probability distribution, the only independence relations are those that follow from the Causal Markov condition alone and not from special parameter values… (Woodward 2003, 65)

Informally, CF ensures that variables are only probabilistically independent if they are causally independent in the true causal graph. When CF is violated, causal relationships cancel each other out by having precisely counterbalanced parameter values, and the variables involved in those balanced relationships are probabilistically independent even though they are not causally independent. Thus, in systems that have CF-violating causal relationships, the probabilistic relationships between variables include independencies that do not reflect the actual causal relationships between those variables.

Probabilistic relationships are used to generate possible causal graphs. There may be multiple distinct causal graphs which all imply the observed set of probabilistic relationships. The candidate graphs can then be used to generate further interventions in the system that will distinguish between the graphs; if two candidate graphs make different predictions for the consequences of an intervention on variable A, then

performing this intervention on A should return an answer as to which of the candidates graphs matches the observed results. The use of probabilistic data to generate candidate causal graphs that can then be used to suggest further interventions can save huge amounts of time and energy by highlighting a few candidates from an indefinitely large number of candidate causal structures.

DAGs of CF violations may take several forms. For example:

Figure 1a   **Figure 1b**
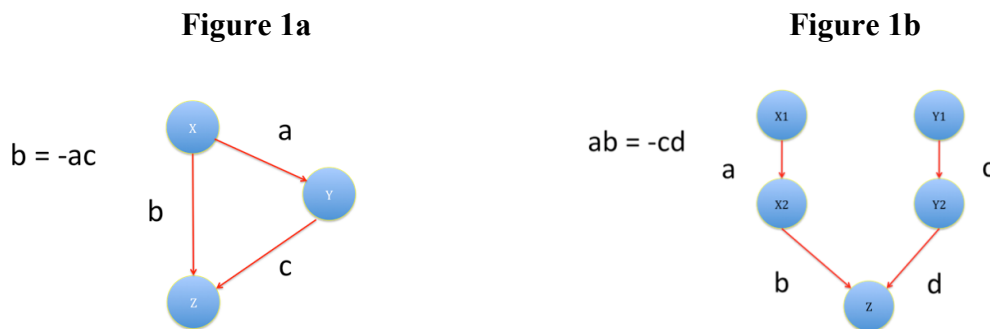
Some authors (Pearl 2000, Woodward 2010) rely on a stronger constraint, causal stability, which requires that probabilistic independence relationships be stable under perturbation of parameter values across some range, to eliminate "pathological" (i.e. CF-violating) parameter values.

Definition 2.4.1 Stability:

Let $I(P)$ denote the set of all conditional independence relationships embodies in P. A causal model $M = <D, \Theta>$ generates a stable distribution if and only if $P(<D, \Theta>)$ contains no extraneous independences – that is, if and only if $I(P(<D, \Theta>)) \subseteq I(P(<D, \Theta`>))$ for any set of parameters $\Theta`$. (Pearl 2000)

Violating causal stability would require a system to respond to changes in one parameter value with compensating changes in another parameter, so that the values remain exactly counterbalanced for some range of values.

The potential for CF-violations to reduce the reliability of methods for extracting causal structure from data is well-known in formal epistemology. However, I will argue that philosophers of science in general should pay more attention to such violations; understanding the difficulties that CF-violations pose will enhance our ability to accurately characterize features of experimental practice, and should be included in normative considerations regarding evidence and inference. This paper provides three main arguments in support of this:

(1) Even if CF-violating systems are measure 0 with respect to the set of causal systems with randomly distributed parameter values, this does not imply that we will only encounter them with vanishing probability. CF-violating systems may be of particular interest for modeling purposes compared to non-CF-violating systems, or because certain kinds of systems have features that render CF-violating parameter values more likely.

(2) As an example of point 1, structural considerations regarding dynamically stable systems that are the result of evolutionary processes should lead us to expect CF-violations in various biological systems. For systems that have evolved to maintain stable equilibrium states against external perturbation, we should also

expect violations of the stronger condition, causal stability. An example of this is briefly presented: mechanisms for salinity resistance in estuary nudibranchs.

(3) 'Apparent' CF-violations in equilibrium-maintaining systems can be generated in certain experimental conditions even though the actual causal relationships in question are not be exactly balanced. Some measurement circumstances will result in a data set that violates CF, even if the actual system being measured does not genuinely violate CF. We should be as concerned with merely apparent as with genuine CF-violations, since both kinds of violations lead to the same difficulties in moving from probabilistic relationships in data to accurate DAGs of systems.

These three points highlight why philosophers of science in general should be concerned: causal systems may not genuinely violate CF, but yet pose the same problems for experimental investigations as if they did. Apparent CF-violations occur when systems do not in principle violate CF but appear to due to measurement issues connected with data-gathering. In both genuine and merely apparent CF-violations, probabilistic relationships in the data will suggest a set of candidate causal graphs that are inaccurate; as a result, further interventions will yield conflicting answers. Scientists could in principle 'catch out' these merely apparent CF-violations *if they knew exactly how to test for them*. But to do this, they would need the DAG, and this is the information that they lack when proceeding from the data to underlying causal structure. When we have incomplete knowledge of the causal structure of the system under investigation, we lack this ability to distinguish between merely apparent and genuine CF-violations.

3. The measure of CF-violating systems

Spirtes, Glymour, and Scheines (2000) offer a proof that CF-violating systems are Lebesgue measure 0 with respect to possible causal systems, while non-CF-violating systems are measure 1. "The parameter values—values of the linear coefficients and exogenous variances of a structure—form a real space, and the set of points in this space that create vanishing partial correlations not implied by the Markov condition have Lebesgue measure 0" (41). From this, they conclude that we are vanishingly unlikely to encounter CF-violating systems, and so proceed with the presumption that any given causal system is not CF-violating. This proof may be part of the reason why comparatively little attention has been paid to causal faithfulness compared to the causal Markov and modularity conditions. However, the fact that CF-violating systems are measure 0 in this class does not imply that we will not encounter them with any frequency.

To motivate this, consider an analogy with rational numbers. They are also measure 0 with respect to the real numbers, while irrational numbers are measure 1. And, there are circumstances under which we are vanishingly unlikely to find them. If a random real number were to be chosen from the number line, the probability that we will draw an irrational number is so overwhelming as to warrant ignoring the presence of rational numbers. However, this does not imply that rational numbers are unlikely to be encountered *simpliciter*: bluntly put, we don't encounter numbers by randomly drawing them from the number line. Rational numbers are encountered, and used, overwhelmingly

more often than one would expect from considering only the proof that they are measure 0 with respect to real numbers.

The Spirtes, Glymour, and Scheines proof assumes that all parameter values within the range of a continuous variable are equally probable (Zhang and Spirtes 2008). Without this assumption, one can't presume that the CF-violating values are vanishingly unlikely. It is true that if causal systems took on parameter values randomly from their range, we would expect to encounter CF-violating systems with vanishingly small probability, and in that scenario, we could safely ignore CF-violations as a real possibility on any given occasion. However, some systems survive, and become scientifically interesting targets for investigation, precisely because they achieve long-term dynamic equilibrium via mechanisms that rely on balanced parameter values. In such systems, the parameter values are not equally probable over their range, but disproportionately likely to be centered around the balanced value(s). In fields like biology, neuroscience, medicine, etc., we are interested in modeling systems that involve equilibrium maintaining mechanisms. This suggests that our modeling interests are focused on CF-violating systems in a way that is disproportionate to their measure when considered against all possible causal systems, and that CF-violating parameter values are disproportionately probable in the first place. Thus, we cannot conclude from the fact that CF-violating parameter values have measure 0 with respect to all possible parameter values that we will not encounter such violations on a regular basis.

Zhang and Spirtes (2008) discuss some circumstances in which systems may violate CF. However, their discussion makes it seem like CF-violations occur primarily in artificial or constructed circumstances. One such example is homeostatic systems, which

maintain equilibrium against some range of perturbations, such as thermostats maintaining a constant temperature in a room. Zhang and Spirtes demonstrate that CF can be replaced with two distinct subconditions, that, taken together, provide almost the same inferential power as causal faithfulness. If systems violate only one of these subconditions, such violations can be empirically detected. This is an extremely useful result, and increases the power of Bayes' nets modeling to recover DAGs from data. However, this result should not be taken as resolving the problem.

In particular, their use of a thermostat as example of a homeostatic system does not do justice to the incredibly complex mechanisms for homeostasis that can be found in various biological systems. Considering these more sophisticated examples provides a clearer view of the potential problems involved in modeling such systems under the assumption of causal faithfulness.

4. Evolved dynamical systems and equilibrium-maintaining mechanisms

The tendency for evolved systems like populations, individual organisms, ecosystems, and the brain to involve precisely balanced causal relationships can be easily explained by the role these balanced relationships play in maintaining various equilibrium states (see, for instance, Mitchell 2003, 2008). Furthermore, the mechanisms by which organisms maintain internal equilibrium with respect to a huge variety of states need to be flexible. They need to not simply maintain a static equilibrium, but maintain it against dynamic perturbation from the outside. This means that many mechanisms for equilibrium maintenance can maintain a fixed internal state over some range of values in other variables. Thus, a system that survives because of its capacity to maintain stability

10

in the face of changing causal parameters or variable values will be likely to display CF-violating causal relationships, and will also violate the stronger condition of causal stability.

An intriguing example is nudibranchs, commonly known as sea slugs (see especially Berger and Kharazova 1997). Many nudibranchs live in ecosystems such as reefs, where salinity levels in the water change very little. In cases where salinity levels vary over narrow ranges, nudibranchs respond to changes in salinity levels by a cellular mechanism for osmoregulation, where cells excrete sodium ions or take in water through changes in cell ion content and volume. This mechanism provides tolerance, but not resistance, to salinity changes, because it maintains equilibrium by exchanging ions and water with the surrounding environment. In cases of extremely high or low salinity, this mechanism will cause the animal to extrude too much or take in too much.

Euryhaline nudibranchs, found in estuary environments where saline levels may vary dramatically between tides and over the course of a season or year, display a much higher level of resistance to salinity changes. There is a pay-off, in the form of increased food sources with reduced competition, to withstanding the wider variation in saline levels. But in these environments, the osmoregulatory mechanism for salinity tolerance is insufficient. Further mechanisms have evolved in nudibranchs (and in molluscs more generally) for maintaining constant internal salinity levels in conditions of extreme salinity variations in the external environment. The osmoregulation mechanism is supplemented with an additional mechanism which involves hermeticization of the mantle, which prevents water and ion exchange with the outside environment. Mantle hermeticization and osmoregulation are distinct mechanisms, but in contexts of extremely

high or low salinity, they both act such that the variables of external and internal salinity are rendered independent.

Further, there are two distinct mechanisms in muscle cells that work in coordination in extreme salinity to maintain a balance of sodium and potassium ions inside the muscle cell. There are two ion pumps in the cell that maintain overall ion concentration at equilibrium across a fairly substantial range of salinity variation in the external environment. Even though external salinity has several causal effects on the internal ion balance of a cell, these two variables will be probabilistically independent for a range of external salinity values (in particular, for the range in which the organisms are naturally found).

> The ion balance of muscle cells during adaptation to various salinities could not be achieved by virtue of the Na/K-pump alone, removing sodium and accumulating potassium. As it is clear from the data obtained, the concentration of both ions drops at low salinity and increases at high salinity. Therefore, the effective ion regulation in molluscan cells can be provided only by cooperative action of two pumps – the Na/K-pump and Na,Cl-pump, independent of potassium transport. (Berger and Karazova 1997, 123-4)

It's worth clarifying that not all variables in the nudibranch salinity regulation mechanisms will be independent: only two variables, internal and external salinity, will be rendered independent by the balanced parameter values. But, because those variables

are actually connected by a chain of causal relationships, this is a spurious independence, one that violates CF.

There are several points that this example illustrates. The first is that of the comparative probability that a complex system, such as an organism like a nudibranch, will display CF-violating causal relationships in the form of mechanisms that maintain equilibrium. We can see how the assumption that all parameter values are equally likely falls apart in the case of evolved systems. Let's grant that, in some imaginary past history, all the parameter values for causal relations in mechanisms such as these two ion pumps were equally likely. This would have resulted in a vast number of organisms that died rapidly with internal ion imbalances. The organisms that managed to stick around long enough to leave offspring were, disproportionately, those with mechanisms that were precisely counterbalanced to maintain this internal equilibrium. Having CF-violating mechanisms would be a distinct advantage. The same applies for other important equilibrium states –organisms with less closely matched values are less capable of maintaining that equilibrium state. Over time, those with the closest matches for parameter values will be more likely to survive. Thus, even if we grant the assumption that all parameter values start out as equally likely, we can see how rapidly the CF-violating ones would come to be vastly overrepresented in the population.

The second point it illustrates is how such sophisticated equilibrium-maintaining mechanisms can violate CF in a much more problematic way than the comparatively simplistic thermostat example considered by Zhang and Spirtes.[1] The two ion pump

---

[1] Note that a DAG representing the two mechanisms for the ion pumps is not of the triangular form that is potentially detectable using the methods in Zhang and Spirtes (2008).

mechanisms are not balanced merely for a single external salinity value: they are balanced for a range of values. Thus, this example violates not merely CF but also the stronger condition of causal stability. This example is interesting in that we know that salinity matters to slugs: finding a probabilistic independence between internal and external salinity is the cue to go looking for an explanation, since we know there is a causal connection between those variables. But we know this because of additional prior, mechanistic, knowledge that we have about those variables. If we did not already have this prior knowledge about the causal connection between internal and external salinity, and were relying on probabilistic relationships alone to find DAGs compatible with our data, we would be systematically misled by this independence.

I am not claiming that all causal relationships in such systems, or all such systems, will violate CF or causal stability. Rather, for any given system that involves equilibrium-maintaining mechanisms, and especially for those with sophisticated evolved equilibrium-maintaining mechanisms, there may be at least some causal relationships that violate either or both of these conditions. This changes the stance we take at the beginning of an investigation: rather than starting from the assumption that CF-violations are vanishingly unlikely, and only revisiting this assumption in the face of difficulties, we should start investigations of such systems with the assumption that it is likely that there will be at least one such spurious probabilistic independence.

5. Apparent CF-violations and their experimental consequences

Consider a possible response to the argument in the previous section. One might be concerned that the examples I offer do not involve genuine CF-violations–when

examined more closely, it may turn out that the causal relationships in questions are not exactly balanced, but merely close. This response might involve the claim that even in the case of biological systems, CF is not genuinely violated, because there are slight differences in parameter values that could be identified, especially if one performed the right interventions on the systems to 'catch out' the slight mismatch in parameter values. Or, by taking recourse to causal stability, one might say that while the equilibrium state of some systems involves precisely counterbalanced causal relationships, in the case of perturbation to that equilibrium, these relationships will be revealed. Perturbation of systems that return to equilibrium would thus be a strategy for eliminating many (or most) merely apparent CF-violations.

Answering this challenge brings us to the heart of why CF-violations deserve broader discussion. Considered from a formal perspective, there is a deep and important difference between systems that actually violate CF, or, causal stability, and those that do not. From a purely formal perspective, merely apparent CF-violations are not methodologically problematic in the same way that genuine ones are. But the ways in which merely apparent CF-violations can be 'caught out' generally will require information about the DAG for the system, in order to predict precisely which variables should be intervened on, within what parameter ranges, in order to uncover closely-but-not-exactly matched parameter values. While it is in principle possible to do this, it requires knowing precisely which intervention to perform, and it is this information that will be lacking in a large number of experimental situations where we are looking fir but don't already have the DAG for the system.

A particular data set drawn from a target system for which investigators are seeking the DAG may have spurious independencies between variables (i.e. violate CF) even though in the true DAG, those parameters are not precisely balanced. In other words, depending on how the data is obtained from the system, the data set may violate CF even though the system itself doesn't. How could this happen? There are a soberingly large number of ways in which a data set can be generated such that a merely apparent CF-violation occurs. The point to note here is that *merely apparent violations will cause exactly the same problems for researchers looking for an unknown DAG as would genuine CF-violations*. Here are some ways in which a dynamically complex non-CF-violating system may nevertheless result in a dataset that is CF-violating.

The first is quite obvious: parameter values that are not exactly counterbalanced may nevertheless be close enough that their true values differ by less than the margin of error of measurements. Consider the parameter values in diagram 1a. A genuine CF-violation will occur if $a=-bc$. However, an apparent CF-violation will occur if $a\pm\varepsilon_1=-bc\pm\varepsilon_2$. Concerns about the precision of measurements and error ranges are well-known, but it is useful to consider them here with respect to the issue of causal faithfulness as another way to flesh out their role in investigatory practices.

Other ways in which apparent CF-violations may occur stem from temporal factors which play a role in the 'catching' of equilibrium-balanced causal relationships. Consider the time scale of a system that involves balanced causal relationships for the purposes of restoring and maintaining some equilibrium state: this may be on the order of milliseconds for some cellular processes, tens to hundreds of milliseconds for many neurological processes, minutes to days for individual organisms. After a perturbation

takes place, the system will re-establish equilibrium during that range of time. In order to successfully 'catch' the counterbalanced causal relationships in the act of re-equilibrating, the time scale of the measurements must be on a similar or shorter time scale. If the time scale of measurements is long with respect to the time scale for re-establishing equilibrium, these balanced causal relationships will not be caught.

This basic point about taking state change data from dynamic processes has particular implications for CF-violations. For processes that re-equilibrate after 50 ms, for instance, a measurement device that samples the process at higher time scales, such as 500ms, will miss the re-equilibration. Thus, even though the system does not violate causal stability, it will behave as if it does, as it will appear that there is a conditional independence between two variables across some range of values, namely, the range between the initial state and the state to which the system was perturbed. In particular, if we do not know what the time scale is, or is likely to be, for re-equilibration, we cannot ensure that a persisting probabilistic independence between two variables in question is genuine and not a consequence of an overly fast re-equilibration timescale.

There are also possibilities for phase-matched cycles that that will make a non-CF-violating oscillating system appear to violate CF. Some systems develop equilibrium mechanisms that result in slight oscillations above and below a target state. If the measurements from this system are taken with a frequency that closely matches that of the rate of oscillation, then the measurements will pick out the same positions in the cycle, essentially rendering the oscillation invisible. This would constitute an apparent CF-violation as well.

Predicting possible CF-violations, real or apparent, requires information about the dynamic and evolved complexity of the systems in question, the particular equilibrium states they display, the time scale for re-establishment of equilibrium compared with the time scale of measurement, and/or the cycle length for cyclical processes.

6. Conclusion

To summarize briefly: some kinds of systems, especially those studied in the so-called 'special sciences', are likely to display the kinds of structural features that lead to CF-violations, such as mechanisms for equilibrium maintenance across a range of variable values. Some systems that do not have CF-violating DAGs may nevertheless generate CF-violating data sets. When we are considering the inferences made from probabilistic relationships in data to a DAG for the underlying system when we do not already have the DAG in hand, we cannot distinguish between genuine and merely apparent CF-violations. Both will cause the same epistemic difficulties for scientists, which is why merely apparent CF-violations deserve broader attention.

It's important to note that I am not discounting the extraordinary achievements in formal epistemology and causal modeling that have marked the last two decades of research on this topic. The steps forward in this field have been monumental, including the development of methods by which to reduce some of the issues arising from CF-violations (such as Zhang and Spirtes 2008). Rather, my goal is to clarify the ways in which apparent CF-violations can arise, the kinds of structural features a system might display that would increase the likelihood of CF-violation, and to bring this issue from discussion in formal epistemology into consideration of scientific practice more broadly.

References

Berger, V.J., and A.D. Kharazova. 1997. "Mechanisms of Salinity Adaptations in Marine

Mollusks." *Hydrobiologia* 355 (1-3): 115-126.

Cartwright, Nancy. 1999. "Causal Diversity and the Markov Condition." *Synthese* 121

(1-2): 3-27.

Cartwright, Nancy. 2002. "Against Modularity, the Causal Markov Condition, and Any

Link between the Two: Comments on Hausman and Woodward." *The British*

Journal for the Philosophy of Science 53 (3): 411-453.

Cartwright, Nancy. 2006. "From Metaphysics to Method: Comments on Manipulability

and the Causal Markov Condition." *The British Journal for the Philosophy of*

Science 57(1): 197-218.

Hausman, Daniel M. and James Woodward. 1999. "Independence, Invariance and the

Causal Markov Condition." *The British Journal for the Philosophy of Science*

50 (4): 521-583.

Hausman, Daniel M. and James Woodward. 2004. "Modularity and the Causal Markov

Condition: A Restatement." *The British Journal for the Philosophy of Science*

55 (1): 147-161.

Mitchell, Sandra D. 2003. *Biological Complexity and Integrative Pluralism*. Cambridge

Studies in Philosophy and Biology: Cambridge University Press.

Mitchell, Sandra D. 2008. "Exporting Causal Knowledge in Evolutionary and

Developmental Biology." *Philosophy of Science* 75 (5): 697-706.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University

Press.

Russo, Federica and Jon Williamson. 2007. "Interpreting Causality in the Health

   Sciences." *International Studies in the Philosophy of Science* 21 (2): 157-170.

Steel, Daniel. 2006. "Indeterminism and the Causal Markov Condition." *The British*

   Journal for the Philosophy of Science 56 (1): 3-26.

Spirtes, Peter, Clark Glymour, and Richard Scheines. 2000. *Causation, Prediction, and*

   Search. Cambridge, MA: The MIT Press.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*.

   Oxford University Press.

Woodward, James. 2010. "Causation in Biology: Stability, Specificity, and the Choice of

   Levels of Explanation." *Biology and Philosophy*. 25 (3): 287-318.

Zhang, Jiji and Peter Spirtes. 2008. "Detection of Unfaithfulness and Robust Causal

   Inference." *Minds and Machines* 18 (2): 239-271.