Perspectives on Causal Specificity

Abstract

Causal specificity is a measure of how important a cause is relative to another. Waters (2007) has developed a theory of causation that deals with specificity. Weber (2006, 2017a, 2017b) has thoroughly criticized it. I defend Waters's theory by showing that non-systematicity is unproblematic. I also argue that Weber's desiderata for theories of causation are too restrictive and insensitive to developments in biological technology. I finally challenge the most fundamental assumption in the framework of causal specificity—that bijective functions are most specific—thus calling for its reassessment.

Word count: 4,768

Introduction

Philosophers of biology debate whether some causes are more important than others in an explanation. The proponents of causal parity maintain that all causes of an effect are equally important. By contrast, the advocates of causal privilege argue that some causes are more important than others. According to the latter, a number of causes may be privileged and the degree to which each is privileged relative to others can differ (Waters, 2007; Weber, 2006; Woodward, 2003). I will explain the important aspects of this discussion using Waters's (2007) terminology.

Waters specifies conditions that a cause must meet in order to 'make a difference' in an effect (e.g., protein synthesis). The basic idea is that if a cause accounts for the variation in an effect (which occurs in a real population), then it is a difference maker with respect to that effect. Causes that counterfactually could have made a difference but that do not actually do so are called potential difference makers. Causes that *do* make a difference are called *actual* difference makers. If only one cause makes the difference, it is called *the* actual difference maker. If a number of causes make the difference, then each is *an* actual difference maker. Causes that actually make a difference are clearly more important than those that do so only potentially, since they account for the actual variation in the effect. The particularly difficult issue concerns the *degree* to which some given *an* actual difference maker is privileged relative to another. More precisely, let A:{x, y, z} denote the set of causes that actually make a difference in effect B. The crucial question is whether, say, x is privileged over y and z by degrees p and q, respectively. If so, we need to understand what this 'degree' is.

The degree of causal privilege is measured in terms of *causal specificity* (defined in section I), a concept developed by Weber (2006) and Woodward (2010). The central idea is that the more 'closely' the values of the cause variable map to those of the effect variable, the more specific the

said cause is with respect to the effect (as compared with its other causes). Waters (2007) claims that the causal specificity of the DNA is greater than that of the splicing agents vis-à-vis their common effect, mature mRNA. As a result, the DNA is the most privileged cause in this context. Although he largely supports the general framework of causal specificity, Marcel Weber (2006, 2017a, 2017b) has systematically criticized Waters's particular theory of causation. Specifically, he has argued that the theory's focus on actual populations often prevents it from being systematic.

I aim to accomplish three tasks in this paper. The first is to defend Waters's theory of causation against Weber's criticisms. The second is to examine the desiderata of theories of causation and argue that Weber's conditions seem unreasonable. The third is to criticize a fundamental assumption about causal specificity that its proponents share. In the first section, I outline Weber's (and Woodward's) account of causal specificity and provide some empirical details from molecular biology to contextualize the discussion. In section two, I explain Weber's criticism that Waters's theory is not systematic. I defend Waters by showing that this problem is not unique to his theory but is a feature of the explananda that all theories of biological causation need to contend with. [1,2] In the third section, I discuss some general issues around the desiderata for theories of causation and argue that Weber's conditions are unreasonable. The fourth section is concerned with criticizing the widespread assumption that bijective functions are causally most specific; I demonstrate that a type of non-bijective function can be more causally specific, thus

[1] I am focusing on Waters's (2007) theory in part because it presupposes Woodward's (2003), so a successful defense of the former entails a defense of the latter (cf. Weber, 2006).

[2] Despite Weber's critique over the last decade or so, neither Waters nor Woodward has responded to him except for Woodward's quick mention of Weber (2006) in a footnote (2010, 305, footnote 17). At present, none of Weber's papers on 'causal specificity' is cited by either Waters or Woodward.

suggesting that the framework's core assumption be reassessed. In the final section, I briefly highlight some implications of my arguments for debates in philosophy of biology.

I. Causal specificity and its relation to molecular biology

Causal specificity is defined in terms of 'functional mapping.' Suppose cause *C* and effect *E* are discrete variables, each ranging over finite sets of values. The causal specificity of *C* corresponds with how 'closely' its values map to those of *E*. More precisely, suppose that *C* and *E* range over $\{c_1, c_2, \ldots, c_m\}$ and $\{e_1, e_2, \ldots, e_n\}$, respectively, and the function *f* maps *C*-values to *E*-values (Weber, 2006; cf. Woodward, 2010).[3] The larger the number of *C*-values that map to their corresponding *E*-values, the more specific *C* is as a cause of *E* (as compared to its other causes). If each and every *C*-value maps to one and only one *E*-value, then the mapping is bijective, making *C* the most specific cause of *E* (Woodward, 2010).[4]

The debate about the relative specificities of various causes emerged in the context of molecular biology. In eukaryotes, sequences of nucleotides on the DNA are transcribed into the pre-mRNA using certain enzymes, such as the RNA polymerase. The pre-mRNA molecule is broken down and reconstituted using 'splicing agents,' which work in conjunction with other enzymes and background 'cellular machinery.' The parts of the pre-mRNA that are excised are known as introns, while its remaining parts are known as exons. The exons are combined to constitute what is then called the mature mRNA, a molecule used in protein synthesis.

[3] The function presupposes Woodward's (2003) manipulability theory of causation, according to which it should answer counterfactual questions such as, what would happen if *C*-value changes from c_1 to c_{25}. If *f* is counterfactually robust—as causal generalizations in biology should reasonably be—then *E*-value would change to e_{25}.

[4] I am assuming that the sizes of the two sets are identical (i.e., $m = n$).

Waters (2007) claims that the DNA and splicing agents both make a difference in the mature mRNA molecule, but the former is causally *more* specific. That is, the number of DNA-values mapping to the mature mRNA values is greater than the number of splicing agents' values mapping to those of the mature mRNA. Therefore, genes are the most specific or privileged cause of the mature mRNA.

II. Systematicity of Waters's theory of causation

Weber (2017b) argues that Waters's theory is not systematic, because it focuses on *actual* populations. He points out that the frequency of a given causal variable differs radically from one biological context to the next. For example, the frequency of splicing agents is much higher in eukaryotes than in prokaryotes. As a result, their causal specificity is significantly greater in the former than in the latter. Waters's theory purports to answer *general* questions, such as whether the causal specificity of the DNA is higher than that of the splicing agents. However, its focus on actual populations precludes it from treating a causal variable in a systematic fashion, thereby preventing it from answering general questions. Weber writes, "The main problem is that [Waters's theory] is very sensitive to the relative abundance of a causal factor in some defined population. Thus, [the causal variable's] values will be highly context dependent, to such an extent as to make any kind of systematic comparison across contexts difficult" (2017b, 578). The theory cannot therefore be used to answer general questions about a variable's causal specificity, and consequently, it does not allow systematic, cross-contextual comparison between the specificities of two (or more) causal variables. No doubt, variables whose frequencies *are* relatively uniform across biological contexts (e.g., RNA) do not pose problems for Waters's theory. So, to be precise, Weber's claim is *not* that Waters's theory can never treat a variable systematically. Instead, his

argument is that it cannot systematically analyse a causal variable whose frequency radically differs between contexts.

According to the advocates of causal privilege, the central purpose of developing theories of causation is to explain why biologists choose certain causes over others when explaining some phenomena that occurs in *real* biological populations (Waters, 2007; Woodward, 2003). Now, it is *not* a feature of Waters's theory that the frequency of splicing agents differs between eukaryotes and prokaryotes. Rather, this is a biological fact that all theories of causation must contend with. Consequently, there is nothing distinctive about Waters's theory that warrants the charge of non-systematicity. Consider a simple analogy. All empirical scientific theories face the problem of induction, the drawing of universal generalizations on the basis of finite evidence. This problem makes scientific theories fallible in principle, because there could exist some evidence that refutes the theory (Nola & Sankey, 2014). However, it would be peculiar to exclude some of the theories on this basis but spare others.

If Weber's criticism is to hold water, he needs to show that either (i) Waters's account fails to be systematic as a result of its own theoretical shortcomings, or (ii) there is at least one other theory that systematically analyses causal specificity. If he attempts to show (i), he must do so by referring solely to Waters's theory, not to its explananda. As I have argued, he relies on the intractability of the explananda to criticize the theory. Alternatively, Weber may demonstrate (i) by showing (ii), because the latter entails the former: if a theory can systematically analyze causal specificity, then the inability of Waters's theory to do the same must be the result of its internal features (or of its application). In other words, if another theory can provide a systematic analysis, then this gives good reason for thinking that the explananda are tractable after all. Consequently, the failure of Waters's theory to do the same cannot be the result of its explananda. Its failure

would arguably be the result of its theoretical apparatus. However, Weber's criticism satisfies neither (i) nor (ii). Therefore, it does not provide reasonable grounds for deeming Waters's theory uniquely problematic.

III. Biological normality and desiderata for theories of causation

My purpose here is to extend the foregoing discussion by analyzing the conditions Weber thinks theories of biological causation must meet. I will argue that Waters's theory meets these conditions. I will also suggest that the conditions themselves are quite misplaced.

Weber (2017b) claims that any plausible theory of causation must meet the conditions of 'biological normality.' A causal intervention is biologically normal if it (a) results from natural processes with non-negligible probabilities and (b) is compatible with the ordinary biological functions of the organism. For instance, DNA transcription is a natural process with non-negligible probability and is compatible with the organism's functions. I think Waters's (2007) theory meets these conditions. The causal interventions in actual biological populations are by definition natural and compatible with the organism's functioning. It is impossible to consider normal biological populations without also thinking about their causal relations as natural and compatible. Alternatively, the concept of actual populations would be vacuous, if not self-contradictory. Accordingly, Waters's theory cannot fail to meet these conditions, because it explicitly focuses on and restricts itself to actual populations. Hence, the theory satisfies the conditions its critic thinks any acceptable theory of biological causation should meet.

A particular reading of Waters (2007) makes it even more difficult for his theory to fail to meet Weber's conditions. The *conditional* reading begins with the fact that biologists consider certain causes as more important than others when explaining some phenomena. The task of

Waters's theory, on this reading, is to provide a principled account that explains their choices. In particular, its task is to answer questions of the following kind: given that a number of causes account for the variation in the effect, which of these is causally most specific? All of the ordinary causes a biologist invokes when explaining some phenomena are natural causes compatible with the organism's functioning, thus satisfying Weber's conditions (a) and (b), respectively. The conditional reading takes this as a given, and the purpose of the theory, on this reading, is to provide a principled account of the relative importance of various causes. Hence, on the conditional reading, Waters's theory will always meet Weber's conditions.

Let us turn to a more general discussion of Weber's conditions and examine whether these demands capture the intuitions about theories of causation that philosophers of biology have in mind. I will focus on Weber's first condition, which states that a causal intervention is biologically normal if it (a) results from natural processes with non-negligible probabilities. This requirement is problematic for at least three reasons. First, natural processes with *negligible* probabilities remain philosophically unanalyzable even though they are ordinarily thought of as biologically normal. For instance, (successful) genetic mutations have very low probabilities. However, they are causally significant for explaining a wide variety of biological phenomena, such as phenotypic variation. In Waters's terminology, genetic mutations are causes that make a difference, and biologists no doubt invoke them in their explanations. Yet, if condition (a) is accepted, genetic mutations (and other natural processes with negligible probabilities) would remain unanalyzable from the perspective of a philosophical theory of causation. In short, this condition wrongly excludes improbable factors that are nonetheless causally relevant.

The second reason that this condition is problematic is that theories of causation which satisfy it will not analyze non-natural interventions. Interventions are usually thought of as "non-

natural" if they cannot occur without the use of technology. Nonetheless, some non-natural interventions, such as *in vitro* fertilization or genomic editing, are often highly relevant for explaining biological phenomena. Indeed, the advent of technology allows biologists to intervene in very specific ways. A philosophical theory that fails to analyze this excludes important aspects of biological phenomena. Weber's condition has precisely this effect: it deems as unacceptable philosophical theories that analyze non-natural but causally relevant interventions. Once again, the condition is too restrictive and, importantly, it is insensitive to developments in biological technology.

Third, Weber's condition is problematic because the boundaries between 'natural' and 'artificial' are more difficult to define than he assumes. For example, gene editing is arguably non-natural because it is carried out using technology. However, edited genes are transcribed and translated into proteins using 'natural' processes. Only the initial cause in this chain of events is supposedly non-natural. The subsequent causes are perfectly natural. Indeed, gene editing technology meets the second condition that (b) the interventions be compatible with the rest of the organism's functioning. However, by requiring that interventions be natural, condition (a) precludes theories of causation from selecting causally relevant factors in technologically altered populations. In other words, on this condition, theories of causation will altogether ignore populations with 'non-natural' interventions even if these interventions are causally fundamental (e.g., gene editing). Consequently, the condition is, once again, unduly restrictive and insensitive to developments in biological technology.

Furthermore, Weber's conditions are in tension with his criticism (presented in the previous section) that Waters's theory fails to treat a causal variable systematically. On the one hand, his condition (a) requires theories of causation to focus only on *actual* populations. On the other hand,

his criticism of Waters's account suggest that he requires theories of causation to be systematic, implying that they should *not* be susceptible to the changing reality of actual populations. A more charitable interpretation is that he requires the theories to be systematic despite the mutability of biological phenomena. However, as I argued in the previous section, it is difficult to see how a theory can be systematic when its explananda is highly mutable. This is surely the case with biological populations, as Weber himself claims. As a result, it is difficult to see how theories can restrict themselves to actual populations *and* be systematic. It seems, then, that these two requirements are in tension with one another. Minimally, the desiderata for theories of causation must be mutually agreeable. They would otherwise require theories to perform incompatible tasks. Weber's requirements (of biological normality and systematicity) fail to meet even this demand.

In light of these considerations, it is natural to ask Weber to explain how these two requirements are compatible. There are at least two courses of action available to him. First, he may argue that while no actual theory of causation has succeeded in satisfying these requirements, it is possible that some theory *could* meet them. Second, he may develop a theory that satisfies both requirements. However, no actual theory of causation (that I am aware of) meets these requirements, suggesting that they are overly restrictive. This rules out the second course of action. As for the first course, it needs to be shown how exactly a theory could be systematic and focus on actual populations. It is insufficient to stipulate conditions without providing at least a sketch of a possible solution. Weber has not provided a sketch of any kind.

Finally, I want to return to the requirement of systematicity. I have already suggested that this requirement is overly demanding. Nevertheless, even if we assume for argument's sake that it could be satisfied, the new insights gained from a systematic analysis of causal specificity do not sufficiently advance our understanding of biological causation. Generally speaking, philosophers

of biology no longer maintain that biological generalizations are universal. They recognize that the generalizations only distribute over well-defined domains (Waters, 2007). In my view, the case of causal specificity should be understood within this framework; there is no need for cross-contextual, domain-general comparisons when domain-specific analyses suffice. For instance, it can be meaningfully asked whether the causal specificity of the DNA with respect to some effect is greater than that of the splicing agents with respect to the same effect within a *well-defined population*. If 'yes,' then we have reason for thinking that, in this particular context, the DNA is more causally important than the splicing agents with respect to a given effect.

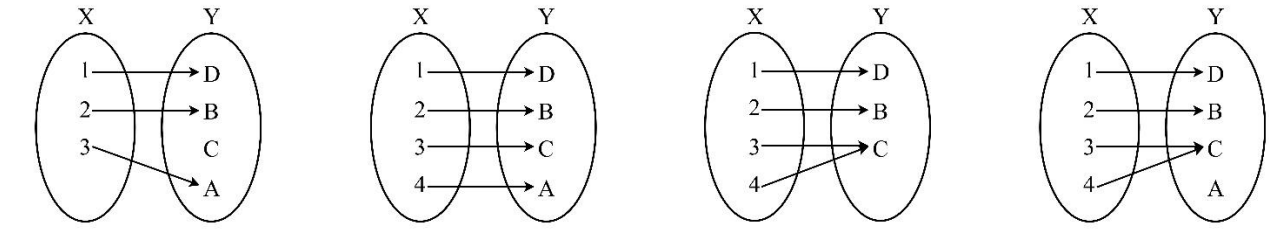
However, systematic analyses do not provide insights of this type. In particular, it is not very informative to compare the specificities of two (or more) causal variables with respect to different effects (in same or different contexts) or same effects (in different contexts). The former are uninformative because there is no relevant commonality based on which the differences could be meaningfully compared; the latter only tell us that a certain cause is more specific than its counterpart with respect to the same effect in a number of contexts. Yet, as Weber rightly points out, the frequencies of variables radically differ between biological contexts. So even if a systematic analysis generated new insights, it would remain largely uninformative when applied to genuine populations. In light this, it is clear that domain-general or systematic analysis is usually uninformative. Yet, that is precisely what the requirement of systematicity demands from theories of causation. Because this requirement is not conducive to advancing our understanding of causation in biology, it is best to altogether eliminate it.

IV. Are bijective functions (causally) most specific?

The aim of this section is to challenge the fundamental idea in the framework of causal specificity: that bijective functions are causally most specific. I will introduce some technical terminology before explaining that a type of non-bijective function can be more specific.

In the first section, I explained specificity in terms of mapping between variables. Let *f* be a function that maps *C*-values to *E*-values. *f* is considered a function if and only it meets the following conditions. First, all *C*-values must map to some *E*-value; there cannot be unmapped *C*-values. Second, no *C*-value can map to more than one *E*-value; each *C*-value must map to at most one *E*-value. Nothing about *E*-values is relevant when determining whether *f* is a function.

A *surjective* ('onto') function is one in which all *E*-values are the image of some *C*-value under *f*, meaning that there are no unmapped *E*-values. A function is *injective* ('one-to-one') when each *E*-value is the image of at most one *C*-value under *f*. That is, given that some *E*-values are mapped, each of these is mapped by at most one *C*-value. A function is *bijective* if and only if it is surjective and injective. The following diagram summarizes these concepts:



An injective non-surjective function (injection, not a bijection)

An injective surjective function (bijection)

A non-injective surjective function (surjection, not a bijection)

A non-injective non-surjective function (also not a bijection)

This framework can be used to map genuine causal relations in biology and to determine their relative specificities. The table below presents some *bona fide* causal relations alongside their respective mappings (Weber, 2017a, 17-8):

Stage of gene expression (C→E)	Mapping *f*(C)=E
1: DNA → DNA (replication)	Bijective
2: DNA → RNA (transcription in prokaryotes)	Bijective
3: RNA → DNA (reverse transcription)	Bijective
4: DNA → pre-mRNA (transcription in eukaryotes)	Bijective
5: pre-mRNA → mature mRNA (in eukaryotes)	Not a function[5]
6: exon parts → protein domains	Surjective non-injective
7: mature mRNA → proteins (translation in eukaryotes)	Surjective non-injective

In this framework, Weber favors a numerical interpretation of the specificity of these mappings. He writes, "Depending on the range of invariance and the *number of values* that the independent [*C*-values] and dependent [*E*-values] variables can take, we can speak of a relation being more or less causally specific" (2006, 606; my emphasis). He claims that "[t]he elements in the codomain [*E*-values] may be mapped onto by different *number* of arguments [values] from the domain [*C*-values] (in the surjective and non-injective cases), or different *proportions* of elements in the codomain may be mapped onto by an argument from the domain (in the injective and non-surjective cases)" (2017a, 16; my emphasis). But in the same paper he distances himself from a proportional notion when he writes, "By "causally most specific" I mean that genes bear Woodward's relation INF [influence] to proteins in the highest degree. By "the highest degree" I mean that the *number of values* that the variables on both sides of the INF relation can take is vastly higher (i.e., many orders of magnitude) than that of any other causal variables that bear the

[5] Because parts of the pre-mRNA are excised (introns), some of its values cannot map to those of the mature mRNA. Consequently, pre-mRNA to mature mRNA mapping does not constitute a function.

relation INF to protein sequences (e.g., splicing agents)" (2017a, 32; my emphasis). Weber clearly favors a numerical conception of causal specificity, according to which one ought to count the number of mappings between *C*-values and *E*-values. The greater the number of mappings between *C* and *E*, the more specific the former is as a cause of the latter (as compared with its other causes). The functions that arguably exhibit the highest number of mappings are bijective, leading to the consensus view that they are causally most specific (Weber 2017b; Woodward, 2010).

I want to argue that surjective non-injective functions can be more specific than bijective functions. If this is true, then the assumption that the DNA is causally most specific may need to be reevaluated. To be sure, I am not claiming that the DNA *actually* fails to be the most specific cause. Instead, my argument will try to show that the consensus view—that bijective functions are always causally most specific—is not correct.

Consider a cause F:{1, 2} and its effect Z:{a, b}. Suppose the function m:$F \rightarrow Z$ bijectively maps F to Z. Consider another cause G:{1, 2, 3} and its effect Z:{a, b}. Suppose the function n:$G \rightarrow Z$ surjectively non-injectively maps G to Z. Let $G(x)=F(x)$ but let $G(3)=G(1)=F(1)$, meaning that *G*-values '1' and '3' and *F*-value '1' map to *Z*-value 'a.' Calculating causal specificities using Weber's approach generates values 2 and 3 for *m* and *n*, respectively. That is, two values of *F* and three of *G* map to Z, meaning that the latter (surjective non-injective) is more specific than the former (bijective). Thus, the same effect can have two causes such that the one which maps to it bijectively—which is presumably the most specific mapping—is *less* specific than the one that maps to it surjectively non-injectively. This is particularly problematic for the proponents of causal privilege, because they regard some of the higher-order causal relations in biology (e.g., mature mRNA → proteins) as involving surjective non-injective mappings. They also consider the most fundamental and specific causes (e.g., DNA → primary transcript) as bijective.

Bijective functions being less causally specific than surjective non-injective functions provides a reason for re-examining the framework. The important intuition behind the idea that bijective functions are supposedly most specific is that nothing seems more specific than one cause giving rise to one and only one effect. This intuition seems *prima facie* correct. Nonetheless, the framework allows non-bijective functions to be more specific. Consequently, the issue is likely with the "number of values" conception used here. As a result, alterations may be required to make this approach work. One suggestion is that the mappings could be used without requiring that they be functions. This would provide more room for theoretical development. More radically, an altogether different approach may be developed that (quantitatively) captures the idea of causal specificity (cf. Griffiths et al., 2015).

V. General implications and conclusion

I will briefly highlight some implications of my arguments. First, the proponents of causal privilege (Waters, 2007; Weber, 2006; Woodward, 2003, 2010) agree that some causes are more important than others when explaining some phenomena. While there are differences about which cause matters to what degree in some context, the authors agree that some causes are definitely privileged over others. Waters's (2007) theory, which relies on Woodward's (2003) manipulability theory of causation, is one of the most thoroughgoing attempts at developing a theory of causal selection in biology. Its central purpose is to provide an alternative to causal parity, according to which all causes in an explanation are equally important. In this paper, I explained the criticisms of Weber, a proponent of causal privilege, against Waters's theory. If my defense of the latter is successful, however, it may go some way in resolving issues internal to the framework of causal privilege. More optimistically, this defense could unify the various proponents of causal privilege, and a robust account of causal privilege as an alternative to causal parity may be developed.

Second, I argued that Weber's conditions for theories of biological causation are overly restrictive. I suggested that these conditions should not prevent theories of causation from analysing causal interventions made using technology. Given the widespread increase in genetic technology, it seems misplaced to regard as irrelevant factors that are causally significant in biological explanations. As such, the desiderata for theories of biological causation may need to be reassessed.

To conclude, I explained what causal specificity is and how it relates to molecular biology. I tried to rebut Weber's argument that Waters's theory fails to be systematic. I also showed that the theory meets Weber's conditions of biological normality, and I argued that his conditions are quite unreasonable. Finally, I showed that surjective non-injective functions can be more specific than bijective functions, thus suggesting the need for re-evaluating at least this tenet in the framework of causal specificity.

References

Griffiths, P. E., Pocheville, A., Calcott, B., Stotz, K., Kim, H., & Knight, R. (2015). "Measuring Causal Specificity." *Philosophy of Science*, 82 (4), 529–555.

Nola, R., & Sankey, H. (2014). *Theories of Scientific Method: An Introduction*. Routledge.

Waters, C. K. (2007). "Causes that make a difference." *The Journal of Philosophy* 104 (11): 551-79.

Weber, M. (2006). "The Central Dogma as a Thesis of Causal Specificity." *History and Philosophy of the Life Sciences* 28 (4), 595–609.

---. (2017a). "Causal Selection versus Causal Parity in Biology: Relevant Counterfactuals and Biologically Normal Interventions." Forthcoming in: C. Kenneth Waters and James Woodward (eds.), *Philosophical Perspectives on Causal Reasoning in Biology*.

---. (2017b). “Which Kind of Causal Specificity Matters Biologically?” *Philosophy of Science*, 84 (3), 574–585.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

---. (2010). “Causation in biology: Stability, Specificity, and the Choice of Levels of Explanation.” *Biology & Philosophy*, 25 (3), 287–318.