Biol Philos (2010) 25:287–318
DOI 10.1007/s10539-010-9200-z

Causation in biology: stability, specificity, and the choice of levels of explanation

James Woodward

Received: 22 July 2009 / Accepted: 27 January 2010 / Published online: 6 February 2010

Abstract This paper attempts to elucidate three characteristics of causal relationships that are important in biological contexts. Stability has to do with whether a causal relationship continues to hold under changes in background conditions. Proportionality has to do with whether changes in the state of the cause "line up" in the right way with changes in the state of the effect and with whether the cause and effect are characterized in a way that contains irrelevant detail. Specificity is connected both to David Lewis' notion of "influence" and also with the extent to which a causal relation approximates to the ideal of one cause–one effect. Interrelations among these notions and their possible biological significance are also discussed.

Keywords Cause · Stability · Levels of explanation · Specificity

Introduction

Philosophical discussion of causation has tended to focus, understandably enough, on finding criteria that distinguish causal from non-causal relationships. There is, however, another important project, also belonging to the philosophy of causation that has received less attention, at least among philosophers. This is the project of elucidating and understanding the basis for various distinctions that we (both ordinary folk and scientists) make *among* casual relationships. This essay attempts to contribute to this second project. In particular, I focus on certain causal concepts (used to mark distinctions among causal relationships) that are employed in biological contexts; these include the notions of *stability* or *non-contingency of association*, appropriate choice of *level of causal description or explanation*, and *causal specificity*. These notions turn out to be interrelated in various complex ways.

J. Woodward (✉)
California Institute of Technology (Caltech), Pasadena, CA, USA
e-mail: jfw@hss.caltech.edu

In saying that attention has tended to focus on the first of the two projects distinguished above, I do not mean that the second project has received no attention at all. One does find self-conscious discussion of notions like causal specificity among researchers in many different areas of biology, including epidemiologists, geneticists, and molecular biologists. Moreover, in the philosophical literature there are discussions of closely related ideas, although the connections with causal notions of biological interest are rarely explicitly recognized. In particular, as I discuss below, the notion of non-contingency of association is closely related to the notion of the stability, insensitivity or invariance of a causal relationship, as discussed by, e.g., Mitchell (2000) and by me (Woodward 2003, 2006), the notion choosing an appropriate level of explanation is related to Yablo's idea (1992) that causes should be "proportional" to their effects, and the notion of causal specificity has interesting relations both to the notion of proportionality and to Lewis' (2000) notion of influence. However, recognition of these connections is complicated by the fact that both the biological and philosophical literatures sometimes fail to distinguish between the two projects described above. More specifically, the features under discussion (non-contingency, specificity etc.) are not infrequently treated (e.g., in Susser 1977) as conditions that can be used to distinguish between causal and non-causal relationships, rather than (as I would urge) features that should be used to distinguish *among* causal relationships. In particular, it is common in the biological literature (particularly in epidemiology—e.g., Hill 1965) to refer to these features as "criteria for causation"; this has suggested both to biologists and others that the features are proposed as necessary conditions for a relationship to be causal. This in turn prompts the response that relationships can qualify as causal even if they lack some or all of the features of stability, specificity and so on. I agree, but urge that it does not follow that the features are unimportant for theorizing about causation or that they do not play important roles in particular scientific contexts.

My aim in this essay is to elucidate what is meant when causal relationships are described as more or less contingent, specific, or framed at an appropriate or inappropriate level, to explore some of the interrelationships among these notions, and locate them notions within a larger framework for discussing causation and explanation. I will also try to illustrate how a concern with whether causal relationships are specific, stable and so on arises in a very natural way in many biological contexts. I will add that although I have attempted to provide biological illustrations of these causal notions, my primary interest is in the content of the notions themselves and less in the empirical details of the illustrations. For example, it is commonly claimed that the causal relationship between DNA sequence and the proteins for which it "codes" is "specific". My concern is with what this claim means—with the empirical features that biologists believe this relationship to possess which leads them to think of it as specific—and only secondarily with the complicated and controversial question of whether the relationship in fact possesses these features. For example, some will hold that it is more accurate to think of the causal specificity achieved in protein synthesis as not due to DNA sequence alone but instead as the result of the interaction of DNA with many other transcriptional

factors.[1] Others may think that many causal relationships in biology—e.g., those having to do with gene action—are less "specific" than commonly supposed. But even in these cases, we still face the questions of what features a relation must possess in order to count as specific and what the contrast between specificity and non-specificity amounts to. It is these sorts of question that I will be exploring.[2]

Apart from its intrinsic interest, the contrast between those causal relationships that are stable, proportional, and specific and those that are not is important for another reason. A number of influential arguments within philosophy of biology turn on claims about "causal parity". Suppose that two different factors, C_1 and C_2, are both causally relevant to some outcome E. Defenders of causal parity theses claim that in at least many of these cases, there is no principled basis for distinguishing between the causal role played by C_1 and by C_2 with respect to E and that instead we must regard both as playing a "symmetric" causal role. For example, developmental systems theorists like Griffiths and Gray (1994) argue that since both genetic factors and many extra-genetic factors are relevant to developmental and evolutionary outcomes, there is no causal role that genes play in development and evolution that is not also played by these other factors. One possible response to such arguments is that although both genetic and extra-genetic factors are causally relevant to outcomes, they nonetheless may not be related to them in a symmetric way. Instead it may be that the relation of one set of factors to the outcomes of interest is more stable, proportional and/or specific than the other. In other words, an understanding of notions like stability, proportionality and specificity give us the resources to distinguish among the different roles or relations that causally relevant factors may bear to an outcome. This theme is explored in more detail in Sect. 7.

Causation and explanation

My strategy in what follows will be to introduce a very undemanding or minimalist notion of causation, based on the interventionist framework described in Woodward (2003). I will then use this as a basis on which to explore the various other distinctions, having to do with stability, specificity and so on, that might be made among causal relationships satisfying this minimalist conception.

Consider the following characterization of what it is for X to cause Y (where "cause" here means something like "X is causally relevant to Y at the type-level"):

[1] A related point is that for ease of exposition, I generally discus what it is for a causal relationship linking a (single) factor C to an effect E to be stable, specific etc. But my discussion should be understood as applying also to the stability, specificity etc. of relationships linking *combinations* of causal factors, C_1, C_2 etc. to effects—these too can be more or less stable etc. In particular, it should be kept in mind that even if the individual relationships between C_1 and E and between C_2 and E are by themselves relatively unstable, non-specific etc., it is entirely possible for relationships linking different combinations of values of C_1 and C_2 to E, to be much more stable and specific.

[2] Another way of describing the project is in terms of the development of a vocabulary and framework for describing features of causal relationships that are often of biological interest; a framework that (I would claim) is more nuanced and illuminating than more traditional treatments of causation in terms of laws, necessary and sufficient conditions and so on.

(**M**) *X* causes *Y* if and only if there are background circumstances *B* such that if some (single) intervention that changes the value of *X* (and no other variable) were to occur in *B*, then *Y* or the probability distribution of *Y* would change.

Here *X* and *Y variables*, which as Woodward (2003) explains, are the natural candidates for the relata of causal claims within an interventionist framework. A variable is simply a property, quantity etc., which is capable of at least two different "values".[3] *Background circumstances* are circumstances that are not explicitly represented in the *X*–*Y* relationship, including both circumstances that are causally relevant to *Y* and those that are not. An *intervention* on *X* with respect to *Y* as an idealized experimental manipulation of *X* which causes a change in *Y* that is of such a character that any change in *Y* occurs only through this change in *X* and not in any other way.[4]

As an illustration, according to **M**, short circuits cause fires because there are background circumstances (including, e.g., the presence of oxygen) such that in these circumstances, intervening to change whether a short circuit is present or absent will change whether a fire occurs (or the probability of whether a fire occurs). Similarly, consider Richard Dawkins' (1982) hypothetical example of a gene *R*, such that those with some abnormal variant *r* of this gene do not learn to read (because they have dyslexia) while those with the normal form *r** do learn to read (given appropriate background conditions, including the right sort of schooling etc.) Assuming that intervening to change the normal form *r* to the variant *r** (or vice versa) is associated (again in appropriate background circumstances) with changes in whether its possessor learns to read, *R* will count as a gene that causes reading, according to **M**.

This last example emphasizes what I meant in saying that M characterizes a weak and undemanding notion of "cause"; undemanding in the sense that it allows a relationship to qualify as causal even if it lacks features thought by some to be characteristic of paradigmatic causal relationships. Thus, a not uncommon reaction to Dawkins' example is that, if the facts are as he describes them, it is in some way misguided or misleading or perhaps just false to describe *R* as causing reading—hence that **M** is in need of emendation since it supports this description. For those who are worried about **M** for this reason, I emphasize again that my strategy in what follows is to use **M** as a foil or baseline to which other more demanding conditions on causation (having to do with stability, specificity etc.) may be added. These additional conditions (particularly, in this case, stability—see Sect. 2) may be used to capture what it is misleading or defective about Dawkins' causal claim and more generally to characterize "richer" notions of causation.

Note that according to **M**, the claim that *X* causes *Y* in itself commits us to nothing specific about *which* changes in *X* (produced by interventions) are associated with changes in *Y* and also says nothing about the particular background

[3] Philosophers often focus on causal claims relating types of *events*. We can represent this with a framework employing variables, by thinking of *X* and *Y* as two-valued, with the values in question corresponding to the presence or absence of instances of the event types.

[4] A more precise and detailed characterization of this notion is given in Woodward (2003, p. 98).

conditions *B* under which this association will occur. (It is enough that there exists such *B*.) Within the interventionist framework, information of these latter sorts is spelled out in terms of more detailed and specific interventionist *counterfactuals* specifying in a more detailed way just how *Y* changes under various possible interventions on *X* and under what background conditions such changes will occur. It is this more detailed information which is related to the considerations having to do with stability, specificity, and appropriateness of level which are the focus of this essay. One way (but by no means the only way) of spelling out this more detailed information is to describe mathematical or logical relationships (e.g., equations) connecting changes in one variable or set of variables to changes in another.

So far my focus has been on causation rather than causal explanation. However, unlike some philosophers, I draw no sharp distinction between providing a casual explanation of an outcome (hereafter the *explanandum-outcome*) and providing information about the causes of that outcome. According to the interventionist conception, when we provide such causal information we provide information that can be used to answer a *what—if—things—had—been—different question*: we identify conditions under which the *explanandum-outcome* would have been different, that is, information about changes that (in principle, and assuming we were able to perform them) might be used to manipulate or control the outcome. More generally, successful causal explanation consists in the exhibition of patterns of dependency (as expressed by interventionist counterfactuals) between the factors cited in the *explanans* and *explanandum*—factors that are such that changes in them produced by interventions are systematically associated with changes in the *explanandum-outcome*. Other things being equal, causal explanations will be better to the extent that the cited patterns of dependency are detailed, complete, and accurate in the sense of identifying the full range of changes in all those factors (and only those factors) such that, if these were to be changed by interventions, such changes would be associated with changes in the *explanandum-outcome*. In other words, good explanations should both *include* information about all factors which are such that changes in them are associated with some change in the *explanandum-outcome* of interest and *not include* factors such that no changes in them are associated with changes in the *explanandum—outcome*. (As we will see below, satisfaction of this feature is related to the notion of proportionality). Moreover, the patterns relating *explanans* and *explanandum* should be (in a sense to be described below) stable or invariant under changes in background conditions.

Stability and non-contingency of association

With this as background, I turn first to the notion of *stability* (also called non-contingency, insensitivity, invariance). Suppose that a relationship qualifies as causal according to **M**: there is a change in the value of *X* that when produced by an intervention in background circumstances B_i is associated with a change in the value of *Y*: in this sense there is a relationship of counterfactual dependence between the effect and the cause in circumstances B_i. The *stability* of this relationship of counterfactual dependence has to do with whether it would continue to hold in a

range of other background circumstances B_k different from the circumstances B_i. To the extent that the relationship of counterfactual dependence would continue to hold under a "large" range of changes in background circumstances or under background circumstances that are judged "important" on the basis of subject mater specific considerations (see below for more on both these notions), that relationship is relatively more stable; to the extent that the relationship would be disrupted by changes in background circumstances, it is less stable. Stability thus comes in degrees—rather than trying to identify some privileged set of background changes that we can use to classify relationships on one side or another of a "stable versus unstable" dichotomy, it is more plausible and better motivated to simply recognize that relationships can be more or less stable or stable under one set of background changes and not another. To the extent that the stability range of a generalization is known, we may help to spell out the content of the generalization by providing such details.

David Lewis (1986) provides an illustration of a relatively unstable (or, as he calls it, "sensitive") causal relationship which (slightly modified by me) is this: Lewis writes a letter of recommendation L that causes *X* to get a job she would not otherwise have got. This in turn has various other effects: *X* meets and marries a colleague she would not have married if she had not taken the job, they have children and grandchildren that would not exist in the absence of Lewis' letter, these grandchildren do various things A and so on.

Now consider the following claim:

(2.1) Lewis' writing the letter L caused *X*'s grandchildren to exist and to do A.

Given the facts just specified, it follows from Lewis' own theory of causation, as well as the account specified in **M**, that (2.1) is true. Whether or not we accept this judgment, virtually everyone will agree that there is something non-standard, or misleading about (2.1). Lewis traces this to the fact that (2.1) is highly sensitive or unstable. The counterfactual dependence associated with (2.1) may hold in the actual background circumstances but if these had been different, in a variety of "small" ways, then if Lewis had written the letter, *X*'s actual grandchildren would not have existed and would not have done A. This might have happened if, for example, *X*'s future spouse *Y* had not also taken a job at the same school as *X*, if other contingencies had led *X* not to marry *Y* and so on.

In characterizing the notion of stability, I said that what matters is whether some relationship of counterfactual dependence would continue to hold under a "large" or "important" range of background circumstances. Application of the quoted words depends on several considerations.[5] One straightforward possibility is that the range of background circumstances under which generalization G' is stable is a proper subset of the circumstances under which generalization G is stable; in this case, we can at least say that G is more stable than G' or stable under a larger range of background circumstances. In other cases, we rely on (i) subject matter specific information to tell us which sorts of changes in background circumstances are most "important" for the assessment of stability and/or (ii) attach particular importance

[5] For more detailed discussion, see Woodward (2006).

to stability under background circumstances that (again perhaps on the basis of subject matter considerations) are regarded as "usual" or "normal". As an example of (i), in assessing the stability of gene → phenotype relationships, we may attach particular importance to whether the relationship is stable under changes in environmental conditions that are "external" to the organism. More ambitiously and demandingly, we may also ask whether the relationship is stable under various changes that might occur elsewhere in the genome.

As a biological illustration, return to Dawkins' example of the gene *R*, which is such that when variant *r* is present, subjects have dyslexia and fail to learn to read (even if the "right" background circumstances are present) but also such that when variant *r*' is present, subjects do learn to read, given the "right" background circumstances. Although **M** agrees with Dawkins' assessment that *R* is a gene "for" (i.e., that causes) reading, the relationship of counterfactual dependence between *R* and whether or not subjects learn to read is relatively unstable under various changes in background conditions: change whether primary education is available (or even more dramatically, whether the culture is one in which there is a written language) and whether the subject learns to read will no longer be dependent on whether she possesses *r* or *r*'.

Contrast this case with claims about the genes that cause, e.g., eye color or external sexual characteristics. Of course the relationship of counterfactual dependence between possession of a *Y* chromosome and external sex characteristics depends upon many additional "background conditions" that are involved in sex determination. But although this relationship is not stable under all possible changes elsewhere in the genome or under suitable changes in various other processes involved in development, it is plausible that it is *more* stable under relevant environmental changes than the *R* → reading relationship ("More" in the sense that to a first approximation, the range of changes in background circumstances in which the *R* → reading relationship is stable is a proper subset of changes under which the relationship between possession of a *Y* chromosome and external sex characteristics.) Moreover, even though the gene → eye color or gene → sex characteristics relation requires the operation of many other factors that are internal to the organism and involved in development and gene expression, it is plausible that as long as these remain within some biologically "normal" range, the above relationships will hold; not so for the *R* → reading relationship.

Some readers may balk even at the suggestion that external sex characteristics or eye color are "genetically caused". My interest is not in arguing about these claims, but simply in observing that they at least seem more natural and less misleading than Dawkins' claims about the genetic causation of the ability to read. I suggest that differences in the relative stability is one important consideration (but not the only consideration—see below) that leads us to have this reaction. Put slightly differently, my suggestion is that part of whatever resistance we may feel to the claim that *R* causes reading has the same source as our resistance to the claim that Lewis' letter causes the existence of *X*'s grandchildren.

What I take to be a very similar idea is developed by the psychiatric geneticist Kenneth Kendler under the heading of "non-contingency of association". Kendler (2005) describes a number of different "criteria" (of which non-contingency of

association is one, along with "causal specificity" and choice of the appropriate level of explanation) that (he holds) should be satisfied for it to be appropriate to characterize a gene as a "gene for" a phenotypic trait or psychiatric disorder. According to Kendler,

> Noncontingent association means that the relationship between gene X and disorder Y is not dependent on other factors, particularly exposure to a specific environment or on the presence of other genes. (2005, 397)

Kendler's non-contingency condition is a stability or insensitivity condition: a gene → disorder relationship is stable or "non-contingent" to the extent that its holding does not depend on the presence of some specific or special environment (with the relationships not holding in other environments) or on whether particular forms of certain other genes are present.

Kendler claims that satisfaction of this criterion of non-contingency is

> … a typical (albeit not uniform) feature of genes that cause classical Mendelian disorders in humans (2005, 397)

In contrast, according to Kendler, there is considerable evidence that the effects of specific genes on psychiatric disorders are influenced both by environmental events and by other genes; hence that such relationships are less stable than "classical Mendelian" gene → phenotype relationships. To the extent this so, it becomes less appropriate to describe these genes as genes for the disorders in question. Although (as indicated above) I find this claim plausible, my primary interest is not in defending it. Rather I put it forward as an illustration of how the notion of stability captures something of biological interest.[6]

To further explore this notion, consider the connection between stability of a relationship and how proximate or distal it is. Obviously, as a general rule, more distal causal relationships with many intermediate links will be less stable than the individual links themselves. Suppose that we have a chain of causal relationships $X_1 \rightarrow X_2$, $X_2 \rightarrow X_3 \ldots X_{n-1} \rightarrow X_n$ which holds in the actual circumstances B in the sense that each individual link satisfies **M** in circumstances B and that in addition there is an overall relation of counterfactual dependence in the sense of **M** between X_n and X_1.[7] Suppose that $X_1 \rightarrow X_2$ would fail to hold in some set of circumstances B_1, $X_2 \rightarrow X_3$ would fail to hold in set of circumstances B_2, and so on. Then (assuming no additional complications such as backup mechanisms) the overall dependence from X_1 to X_n will be disrupted if any one of the circumstances in B_1 or B_2 or B_{n-1} holds. So unless $B_1 \ldots B_{n-1}$ are strongly overlapping (e.g., most members of B_2 are already in B_1 etc.) the overall $X_1 \rightarrow X_n$ relationship will be less stable than any of the individual links $X_i \rightarrow X_{i+1}$. Thus to the extent that we value finding stable causal relationships, we will often be able to accomplish this goal by looking

[6] Relatedly, it is no part of my argument that relatively stable gene → gross phenotypical traits relationships are common. Arguably (e.g., Greenspan 2001) they are not, but if so, we still require the notions of stability/instability to express this fact.

[7] This second condition is not redundant; even if each individual link in the chain satisfies **M**, there may be no overall counterfactual dependence between X_n and X_1. See Woodward (2003, pp. 57ff).

for more proximate causal relationships that mediate distal relationships.[8] It follows that a concern with stability can sometimes (but need not always)[9] drive us in a "reductive" direction, toward the identification of more fine-grained, "micro" relationships. Note, though, that this does *not* mean that stability is just another name for how proximate a causal relationship is. For one thing, it is perfectly possible for a distal relationship to be relatively stable (and even no less stable than its individual proximate links) given the right relationship between B_1, B_2, B_n). More generally, how proximate a causal relationship is seems to be relative to the coarseness of grain in variable description one employs.[10] By contrast, stability is not representation-dependent in this particular way.

One reason why the stability of a causal relationship matters biologically is that this may bear on the question of how readily the relationship can be altered, whether by processes such as natural selection or by human intervention (the latter consideration mattering for biomedicine and social policy) and on the extent to which this alteration can occur independently of changes in other processes. For example, in eukaryotes the causal relationship between a particular DNA sequence and the pre-mRNA for which it codes is more proximal and also more stable than the DNA sequence → mature mRNA sequence relationship since the latter is mediated by the activity of various splicing enzymes. The DNA → mRNA relationship is in turn more stable than the relationship between DNA sequence and yet more distal phenotypical features. (Arguably it is also true the DNA → mRNA relationship in prokaryotes is more stable than this relationship in eurkaryotes since the former is not affected by splicing agents.) Of course it is true that even the DNA → pre-mRNA relationship is not completely stable—it depends on factors like the presence of RNA polymerase and various other cellular features. But the relation between the DNA sequence and its more distal effects in eurkaryotes is even less stable—it depends both on these factors and on more besides such as the activities of various splicing enzymes.

One consequence is that in eukaryotes it may be easier (in the sense that there are more possible changes that will produce this outcome) for natural selection or mutation to alter the relationship between DNA sequence and mRNA than for these to alter the relationship between DNA sequence and pre-mRNA. The former

[8] As Kendler has pointed out to me, this is essentially the logic behind looking for so-called endophenotypes in psychiatric genetics, when these are construed as common pathway variables that are causally intermediate between genotype and phenotype—see, e.g., Gottesman and Gould (2003). Ideally, relationships between endophenotype and phenotype will be more stable than genotype—phenotype relationships and also perhaps more causally specific in the 1–1 sense described in Sect. 5.

[9] Some macro-level relationships may be highly stable (under, say, some range of changes in features of their components) and may better satisfy other conditions like proportionality described below. Relationships among thermodynamic variables provide examples. Whether stable relationships are to be found at more micro or more macro levels is thus always an empirical question.

[10] With respect to a set of variables like {wish for victim's death, firing of gun, victim's death}, the relationship between the second and third variables will be "direct" or "proximal". With respect to an expanded more fine grained set of variables {wish for victim's death, firing of gun, penetration of victim's heart by bullet, loss of blood supply to brain, victim's death} the relationship between firing and death is mediated or distal. But the overall stability of the firing → death relationship does not depend on whether we employ a representation with these intermediate variables.

alteration might occur, for example, via changes in the genetic regulation of the activities of splicing enzymes which leave the DNA → pre-mRNA relationship unchanged. Similarly, changes in regulation of expression of structural genes can have profound phenotypic effects even though the relationship between the structural genes themselves and the proteins for which they code remains stable.[11]

Before leaving the notion of stability an additional remark may be helpful in placing this notion in a more general perspective. The issue of whether biology contains "laws" (and if so, which biological generalizations count as laws) has been the subject of a great deal of discussion among philosophers of biology. I won't try to settle this question here, but two points seem uncontroversial. First, there is an obvious connection between lawfulness and stability: paradigmatic laws drawn from physics and chemistry are very stable generalizations—they hold over a wide range of background conditions. Second, many biological generalizations, including many we think of as describing causal relationships, have somewhat more restricted ranges of stability than fundamental physical and chemical laws—for many such generalizations there are not just nomologically possible but actually occurring, biologically relevant conditions under which they break down or have exceptions. It is an important point that we may ask about the conditions under which such generalizations are (or are not) stable and, as illustrated above, make assessments of their relative stability without trying to settle the difficult question of whether the generalizations are properly regarded as "laws". In other words, at least some of the concerns that motivate discussions of the role of laws in biology can be addressed by focusing directly on the notion of stability, rather than the notion of law.

Proportionality and the choice of an appropriate level of explanation

It is a common idea that some choices of level of explanation or causal description are more appropriate of perspicuous than others, although there is little consensus about what exactly this means. The version of this idea that I want to defend involves the claim that, depending on the details of the case, causal description/explanation can be *either* inappropriately broad or general, including irrelevant

[11] Suppose one has a network of interacting causal structures or units, with, e.g., C_1 causing C_2, C_2 in turn influencing both C_3 and C_4 and so on. I have elsewhere (Hausman and Woodward 1999; Woodward 1999, 2003) characterized such a structure as *modular* to the extent that various of these causal relationships can be changed or disrupted while leaving others intact—that is, a relatively modular structure is one in which, e.g., it is possible to change the causal relationship between C_1 and C_2 while leaving the causal relationship between C_2 and C_3 intact. When modularity is so understood, it is one kind or aspect of stability—it involves stability of one causal relationship under changes in other causal relationships (which we can think of as one kind of background condition). Like stability, modularity comes in degrees and relative modularity is a feature of some sets of causal relationships, not all. (As recognized in Woodward 1999). Hausman and Woodward (1999) contains some mistaken assertions to the contrary, appropriately criticized in Mitchell (2009). Notions of modularity figure importantly in recent discussions of genetic regulatory networks and other structures involved in development and in evolutionary change—see, e.g., Davidson (2001). Obviously, it is an empirical question to what extent any particular example of such a structure is modular (see Mitchell 2009 for additional discussion.) My claim is simply that modularity (and its absence), like stability more generally, is a feature of causal relationships and their representation that is of considerable biological interest.

detail, or overly narrow, failing to include relevant detail. Thus, which level (or levels) is (are) most appropriate will be in large part an empirical, rather than a priori matter—empirical in the sense that it will depend on the causal structure of the situation under investigation. This contrasts with the common philosophical tendency to think there is a single, universal level of causal description that is most appropriate—either a level of maximal specificity/detail (perhaps given to us by physics or biochemistry) or, alternatively, a preferred level of maximal generality or abstractness (as perhaps implied by some versions of unificationist accounts of explanation).

Although my focus in what follows will be mainly on how the choice of level is (and should be) influenced by empirical considerations, I should emphasize that it is fully consistent with this that the investigator's purposes, and in particular what it is that the investigator wishes to explain or understand should also influence the choice of level. Roughly speaking, the investigator's purposes or interests influence what she wants to explain (the choice of *explanandum*), and once this is fixed, empirical considerations play a large role in influencing the "level" at which an explanation for this *explanandum* is most appropriately sought. As an illustration, developed in more detail below, suppose an investigator wishes to understand how individual neurons generate spike trains with particular temporal features (described by the functional forms governing firing rates over time.) Then the details of the anatomy and molecular processes at work within the individual neurons likely will be relevant to this *explanandum*. If instead the investigator wishes to understand how and why assemblies of neurons produce (in response to certain inputs) certain outputs that in turn influence behavior, then some of this anatomical and molecular detail *may* no longer be relevant—no longer relevant because it may turn out to be the case, as an empirical matter, that the features of the neuronal output that influence the behavior in question do not depend upon these temporal features of individual spike trains or the factors that influence these, but are instead just sensitive to certain aggregate population level statistics of the incoming neural signals, such as average firing rates across these populations.

To explore the idea of an appropriate choice of level of explanation, consider a proposal due to Stephen Yablo (1992). Yablo suggests that causes should "fit with" or be "proportional" to their effects—proportional in the sense that they should be just "enough" for their effects, neither omitting too much relevant detail nor containing too much irrelevant detail. Yablo spells out this idea by appealing to "essentialist" metaphysical ideas but I want to focus on his underlying intuition, which is quite plausible. One of the illustrations Yablo uses to motivate his proposal is this: A pigeon is trained to peck at a target when and only when presented with a stimulus of any shade of red. Suppose, on some particular occasion or series of occasions, the pigeon is presented with a particular shade of scarlet and pecks at the target. Consider the following two causal claims/causal explanations:

(3.1) The presentation of a scarlet target caused the pigeon to peck.
(3.2) The presentation of a red target caused the pigeon to peck.

Yablo holds (and I agree) that (3.2) seems to provide a description of the causal structure of this situation that is in some way more perspicuous or appropriate (or

furnishes a better explanation) than the representation provided by (3.1). In Yablo's language, the cause cited in (3.2) fits better with or better satisfies the requirement of proportionality with respect to its effect than the cause cited in (3.1). Intuitively, this is because the cause cited in (3.1) contains, on at least one natural interpretation, irrelevant detail or fails to convey relevant detail: any shade of red would have caused the pigeon to peck but (3.1) fails to convey this information. Indeed, it arguable that (3.1) is most naturally understood as (wrongly) suggesting instead that that the specifically scarlet color of the target is causally relevant to pecking. By contrast (3.2) correctly identifies the conditions changes in which (from a red to non-red target) will lead to a change from pecking to non-pecking behavior.

For our purposes, it does not matter exactly how we go onto characterize this limitation of (3.1). We could say that (3.l) is false on its most natural interpretation (that is, when interpreted as claiming that a change from scarlet to any non-scarlet color will change whether there is pecking) and hence that **M** fails to provide a sufficient condition for causation.[12] Alternatively, we could say that (3.1) is literally true but pragmatically misleading, and/or furnishes a less adequate causal explanation than (3.2). However described, it is this limitation of (3.1) that we have in mind when we say it exhibits a failure of proportionality.

Given a specification of an effect variable or *explanandum* (which will include a specification of a range of possible values this variable can take) I propose that a cause will be proportional to its effect (or will come closer to satisfying the constraint of proportionality) when (or to the extent that) the following condition is met:

> (**P**) There is a pattern of systematic counterfactual dependence (with the dependence understood along interventionst lines) between different possible states of the cause and the different possible states of the effect, where this pattern of dependence at least approximates to the following ideal: the dependence (and the associated characterization of the cause) should be such that (a) it explicitly or implicitly conveys accurate information about the conditions under which alternative states of the effect will be realized *and* (b) it conveys *only* such information—that is, the cause is not characterized in such a way that alternative states of it *fail* to be associated with changes in the effect.

A cause that fails to convey the information described under (a) will fail to contain causally relevant detail and a cause that fails condition (b) will convey irrelevant detail.[13] Both conditions are not (fully) satisfied in the case of (3.1) since (3.1) both (a) fails to convey full and accurate information about the conditions under which non-pecking will occur and (b) suggests (at least on the "natural" interpretation described above) that changes from scarlet to non-scarlet are

[12] That is, there is a change in the condition cited in (3.1) (from scarlet to non-red) which is associated with a change in pecking, so that **M** judges that (3.1) is true; hence requires revision if (3.1) is false.

[13] Another way of understanding proportionality is in terms of employing variables that allow for the parsimonious maximization of predictive accuracy. When **P** fails there will either be a characterization of the cause such that variation in it could be exploited for predictive purposes but is not so used or else "superfluous" variation in the cause which does not add to the predictability of the effect.

associated with changes in whether or not pecking occurs. By contrast, (3.2) does not exhibit these defects. I suggest that this is what underlies the judgment that the cause cited in (3.1) fails to be proportional to or fit with its effect and that it introduces irrelevant detail.

The pigeon example may seem somewhat artificial but in fact there are many examples with a similar structure that arise naturally in biological contexts. A common view among neurobiologists is the neural code is primarily or entirely a "rate" code—that is, information is carried by firing rates of ensembles of neurons, so that what matters is simply the average number of times firing occurs within some temporal interval and not, e.g., the precise temporal location of the spikes within this interval or the detailed temporal features of firing patterns of individual neurons. An alternative possibility (defended in detail in Rieke et al. 1997) is that other features of the spike trains of individual neurons, such as their detailed temporal pattern (as when, e.g., changes in the probability of firing follow a sinusoidal pattern over time) also carry important information. Suppose we wish to provide a true causal claim about (or explanation of) the response of some neural structure to incoming stimuli. On any given occasion, these stimuli will exhibit a detailed, particular temporal pattern. But if the rate code hypothesis is correct, most of this detail will be causally irrelevant to the neural response—only the firing rates reflected in the incoming signals (and presumably only some aggregate of these) will matter. (Empirically, this would mean that the neural response remains the same across variations the temporal course of the incoming signals in as long as the incoming firing rates remain the same.) A claim about the causes of the neural response that adverted to this irrelevant detail would violate constraint **P** (and would be like (3.1) in the example above)—features of the temporal pattern of firing such that changes in it left the overall firing rate unchanged would not be associated with changes in the neuronal response in violation of condition (b) above.

Suppose next, by way of contrast, that the rate hypothesis is wrong and that other features of the temporal pattern of firing matter for neuronal response. Then a causal claim that attributes the neuronal response just to the overall firing rate of the incoming signals would also violate the proportionality constraint—now the claimed cause would fail to incorporate relevant detail in contravention of condition (a) above. Thus, depending on the empirical details of the case, considerations of proportionality may lead either toward the incorporation of more fine-grained detail in the specification of causes (in a "reductive" direction, if you like) or toward specifications that abstract away from such detail.

Note that considerations of proportionality represent constraints that are at least partly independent of the considerations having to do with stability discussed in the previous section. The relationship between presentation of a scarlet target and pecking may be just as stable as the relationship between presentation of a red target and pecking; nonetheless it may be more appropriate to describe the cause in this case as the target's being red rather than its being scarlet if this description better satisfies the proportionality constraint. More generally, a causal claim may do a good job of satisfying the proportionality constraint but involve generalizations that are relatively unstable; alternatively the claim may involve generalizations that are relatively stable but do a poor job of satisfying the proportionality constraint.

So far, we have focused on cases in which, intuitively, *causes* were at the wrong level for their effects, either containing too much or too little detail. Interestingly, a failure of proportionality (or a mismatch of levels) between cause and effect can also occur on the *effect* side: the candidate effect either may also involve detail that is irrelevant to the cause.

Illustrations of this possibility are provided by some thought experiments, again taken from Kendler (2005), concerning the appropriate choice of "level of explanation" (his terminology) in connection with claims of genetic causation. According to Kendler, another criterion (in addition to non-contingency of association) for when it is correct to talk of a "gene for" some phenotypic trait is whether the level of explanation connecting gene and trait is "appropriate". Kendler writes:

> To illustrate how this issue—the appropriateness of level of explanation—may apply to our evaluation of the concept of "a gene for…" consider these two "thought experiments":
>
> Defects in gene *X* produce such profound mental retardation that affected individuals never develop speech. Is *X* is a gene for language?
>
> A research group has localized a gene that controls development of perfect pitch. Assuming that individuals with perfect pitch tend to particularly appreciate the music of Mozart, should they declare that they have found a gene for liking Mozart? (2005, pp. 398–399)
>
> According to Kendler:
>
> For the first scenario, the answer to the query is clearly "No." Although gene *X* is associated with an absence of language development, its phenotypic effects are best understood at the level of mental retardation, with muteness as a nonspecific consequence. *X* might be a "gene for" mental retardation but not language.
>
> Although the second scenario is subtler, if the causal pathway is truly gene variant → pitch perception → liking Mozart, then it is better science to conclude that this is a gene that influences pitch perception, one of the many effects of which might be to alter the pleasure of listening to Mozart. It is better science because it is more parsimonious (this gene is likely to have other effects such as influencing the pleasure of listening to Haydn, Beethoven, and Brahms) and because it has greater explanatory power. (2005, pp. 398–399)
>
> Kendler adds "a final scenario":
>
> Scientist A studied the behavioral correlates of a particular variant at gene *X* and concluded "This is a very interesting gene that increases the rates of sky diving, speeding, mountain climbing, bungee jumping, and unprotected casual sex." Scientist B studied the same variant and concluded "This is a very interesting gene and effects levels of sensation-seeking."
>
> He then asks:
>
> Who has done the better science? Since sensation seeking (and its close cousin novelty-seeking) are well studied traits, scientist B has provided results that

> are more parsimonious and potentially provide greater explanatory power. For example, only scientist B could predict that this gene ought to be related to other behaviors, like drug taking, that are known to be correlated with sensation-seeking. (2005, p. 399)

I agree with Kendler's judgments and think they fall naturally out of proportionality constraint **P** and the interventionist account of causation and explanation sketched above. In the first scenario, changes in whether gene *X* is defective or not (that is, changes that replace the normal form of the gene with a defective form and vice versa) are associated not just with changes in the ability to speak but with many other changes as well—in various general cognitive abilities and so on. In describing the effect of variations in gene *X*, we prefer a characterization that captures the fact that such variations are associated with all these other changes as well, and which presents such information in a parsimonious way, revealing what all these particular consequences have in common. Thinking of the defects in the gene as causing mental retardation accomplishes this—it provides more information regarding the answers to what-if-things-had-been-different questions than if we merely think of the defect in the gene as causing muteness. Extending a bit the characterization of proportionality under **P** above, if we describe the gene as causing muteness, then there will be changes that result from alterations in whether or not one possesses this gene that are not captured by this causal claim—namely changes in other features (in addition to muteness) associated with mental retardation.

A similar analysis holds for the gene that influences pitch perception—indeed, in this case the fit between the gene and the "effect" of liking Mozart is even more imperfect than in the previous case. Again, as Kendler notes, changes in whether one possesses this gene will likely be associated with many other changes besides whether one likes Mozart—for example, changes in whether one likes other musicians, changes in one's musical abilities, and so on. A characterization of the effect of the gene in terms of perfect pitch does a better job of capturing these additional patterns of dependency than the characterization in terms of liking Mozart. In addition, even assuming those with the gene are more likely to like Mozart, whether they do so will depend on much else besides possession of the gene (for example, on exposure to Mozart's music)—in this sense the gene → liking Mozart relationship is like Dawkins' gene → reading relationship in being comparatively unstable or non-invariant. Moreover, the relationship between *lacking* the gene and *not* liking Mozart is also unstable in that many people who lack the gene will still like Mozart. A similar analysis applies to the third example.

Causal specificity

I now turn to yet another causal notion—*causal specificity*—that is also important (indeed ubiquitous) in the biological sciences.[14] My discussion will connect the

[14] A point recognized by many writers. Greenspan (2001) writes, "specificity has been the shibboleth of modern biology" (383) and Sarkar (2005) that "specificity was one of the major themes of twentieth century biology" (263).

notion of specificity to two interrelated concepts, one of which is a variant on David Lewis' (2000) notion of influence (the *fine-grained influence* conception of specificity) and the other of which embodies the prima-facie implausible idea that each cause should have a single (relevant) effect and/or each effect should have a single relevant cause (the *one to one* conception). I will then explore the relationship between these two conceptions.

An interesting invocation of "specificity" is provided by Davidson (2001). Davidson appeals to this notion as a reason for focusing on what he calls the "control circuitry embodied in the DNA" as opposed to "other cellular machinery" in answering the question of where do the "causal differences responsible for morphological diversity reside and how exactly do they function?".

Davidson's answer is as follows:

> A large part of the answer lies in the gene control encoded in the DNA, its structure, and its functional organization. …. In physical terms the control circuitry encoded in the DNA is comprised of *cis*-regulatory elements, i.e., the regions in the vicinity of each gene which contain the specific sequence motifs at which those regulatory proteins which affect its expression bind; plus the set of genes which encode these specific regulatory proteins (i.e., transcription factors). …. Of course the *trans*-regulatory apparatus can be considered much more broadly. If one relinquishes the constraint of considering only those *trans*-regulatory molecules which directly interact with DNA, by recognizing and binding at *cis*-regulatory target site sequences, then large components of both nuclear and cytoplasmic cellular biochemistry might also be included. Among these would be all those signaling pathways, adaptor proteins, cofactors, and other entities that affect the activity of transcription factors. But it seems clear that most of this cellular machinery is in general ubiquitous or in any case relatively nonspecific; that it is always utilized for so many diverse regulatory tasks in each organism; and that by far the most important genomic determinants of animal diversity are the regulatory elements which encode the genetic program for development. (2001, pp. 1–2, underlining reflects my emphasis)

Davidson's claim is thus that although both the DNA sequence and the other "cellular machinery" involved in transcription play a causal role in development, they play a different or asymmetric role; the former is more causally specific than the latter.

A very similar claim is defended by the philosopher Ken Waters in a recent paper:

> DNA is a specific difference maker in the sense that different changes in the sequence of nucleotides in DNA would change the linear sequence in RNA molecules in many different and very specific ways. RNA polymerase does not have this specificity. Intervening on RNA polymerase might slow down or stop synthesis of a broad class of RNA molecules, but it is not the case that many different kinds of interventions on RNA polymerase would change the linear sequence in RNA molecules in many different and very specific ways.

> This shows that DNA is a causally specific potential difference maker. The fact that many such differences in DNA do actually exist and these differences actually explain the specific differences among RNA molecules indicates that DNA is the causally specific actual difference maker with respect to the population of RNA molecules first synthesized in eukaryotic cells. (2007, pp. 574–575)[15]

Both Davidson and Waters claim that the DNA sequence is causally specific (with respect to RNA, proteins synthesized and in Davidson's case, also "morphological diversity") and that at least some other parts of the machinery involved in RNA production, protein synthesis and so on are not similarly specific, thus introducing a causal asymmetry between the role played by DNA and this other machinery. I want to put aside the question of whether these claims are correct and instead to ask the prior question of what the notion (or notions) of causal specificity invoked in the above passages amounts to. What do these writers *mean* when they claim that DNA is causally specific? One consideration to which both Davidson and Waters draw attention is the fact that there is a great deal of actually occurring variation in the DNA sequence in different genes and in the expression of these genes at different times and places in the organism. In contrast, many of the elements of the other "cellular machinery" involved in production of proteins are, at least when the cell is functioning properly, to a first approximation common and ubiquitous to all cases of transcription and protein synthesis. This suggests that if you want to understand why *different* proteins (or different mRNA molecules) are synthesized at different times and places in the organism, the answer is to be found in what varies—what Waters calls the actual difference makers—rather than in what is common, unchanging or constant, such as the presence of the appropriate cellular machinery for synthesis.

But while this consideration seems relevant to Davidson's and Waters' judgments that the DNA sequence is causally specific, it doesn't seem to be the only consideration. For one thing, it is certainly possible for aspects of this cellular machinery to be changed or disrupted—as Waters notes, an experimenter or nature might alter the amount of RNA polymerase present in a particular prokaryotic cellular environment, thereby altering the rate of RNA synthesis but (he claims) this would not turn RNA polymerase into a specific cause of the RNA sequence produced.

In the quotation above, Waters suggests that the reason (or an additional reason) why we regard RNA polymerase as non-specific is that it is *not* the case that many

[15] Waters speaks in this passage of DNA as "the" causally specific actual difference maker for RNA molecules "first synthesized" in eukaryotic cells (i.e., presumably pre-mRNA) but he goes onto note that in eukaryotes different varieties of RNA polymerase and different splicing agents are involved in the synthesis of mature RNA, with different splicing agents also acting as causally specific actual difference makers for this mature RNA. Thus, according to Waters, while DNA is causally specific actual different maker for mature RNA in eukaryotes it is not the only such causally specific agent. As previously emphasized, this will not affect my discussion below, which focuses on what it might mean to say that DNA is causally specific with respect to RNA and not on whether other causes are also present that act in a causally specific way. Also the DNA that acts as a causally specific actual difference maker is of course activated DNA.

different kinds of interventions on RNA polymerase would change the linear sequence in RNA molecules in many different and very specific ways. By contrast, we regard the DNA sequence as a specific cause at least in part because interventions that change this sequence in many different and specific ways will also change the linear sequence of RNA molecules in different and specific ways. I agree that this captures one aspect or element in the notion of causal specificity. In what follows, I want to flesh this idea out, and connect it to some recent philosophical discussion.

As Waters observes elsewhere in his paper (following a suggestion of mine) it seems natural to connect the aspect of causal specificity under discussion to Lewis' (2000) notion of "influence". Lewis characterizes this notion as follows:

> Where C and E are distinct actual events, let us say that C influences E if there is a substantial range C_1, C_2,… of not too distant alterations of C (including the actual alteration of C) and there is a range E_1, E_2… of alterations of E, at least some of which differ, such that if C_1 had occurred, E_1 would have occurred, and if C_2 had occurred, E_2 would have occurred, and so on. Thus, we have a pattern of counterfactual dependence of whether, when and how on whether, when, and how. (2000, p. 190)

Although my use of this notion will be broadly similar, I will modify Lewis' treatment in several respects. First, Lewis proposes to use the notion of influence to define or characterize causation *simpliciter* (that is, to provide necessary and sufficient conditions for C to cause E.) In particular, he suggests that "C causes E if there is a chain of stepwise influence from C to E". My view is that this is not very promising, basically because there seem to be many examples of causal relationships not involving influence (or at least much influence) in Lewis' sense.[16] However, I do think that (as with stability and proportionality) we may use the notion of influence (along with specificity) to distinguish in a useful way *among* causal relationships, rather than treating it as a "criterion" of causation. That is, some causal relationships involve considerable influence (in Lewis' sense) and some do not, this difference is related to whether those causal relationships seem "specific" or not, and whether a causal relationship is specific (or involves influence) matters in biological contexts.

Second, Lewis thinks of causation as a relationship between events (i.e., relata that correspond to binary variables). One consequence is that Lewis finds it natural to focus on variations in the time and place of occurrence of C and E, and on whether there is systematic dependence between such variations in characterizing influence, since these are the obvious dimensions along which events can vary. Without denying the relevance of such temporal and spatial dependence, I prefer (as explained above) to think of causal relata as variables that can be in any one of a number of different states (or can take a number of different values), rather than just two. For such variables, influence (and one notion of specificity) will have to do with whether there are systematic dependencies between a range of different possible states of the cause and different possible states of the effect, as well as

[16] See Kvart (2001) for examples.

dependencies of the time and place of occurrence of E on the time and place of C. My proposal is that, other things being equal, we are inclined to think of C as having more rather than less influence on E (and as a more rather than less specific cause of E) to the extent that it is true that:

> (**INF**) There are a number of different possible states of C ($c_1 \ldots c_n$), a number of different possible states of E ($e_1 \ldots e_m$) and a mapping F from C to E such that for many states of C each such state has a unique image under F in E (that is, F is a function or close to it, so that the same state of C is not associated with different states of E, either on the same or different occasions), not too many different states of C are mapped onto the same state of E and most states of E are the image under F of some state of C. This mapping F should describe patterns of counterfactual dependency between states of C and states of E that support interventionist counterfactuals. Variations in the time and place of occurrence of the various states of E should similarly depend on variations in the time and place of occurrence of states of C.

In other words, C will influence E to the extent that by varying the state of C and its time and place of occurrence, we can modulate the state of E in a fine-grained way. One might think of the ideal case of influence (and the notion of specificity to which it is related) as one in which the mapping F is a function that is both 1–1 (injective) and onto (surjective)—that is, F is bijective.[17] In real-life cases this condition will rarely be met, but we have more influence/specificity the closer we get to it. From the point of view of assessing influence, usually it will matter more that the mapping F is a function and that many states of E are the image under F of some state of C, then that F be 1–1, since the notion we are trying to capture is that the state of C exerts a fine-grained kind of control over which state of E is realized. It will be less damaging to the achievement of such control if several different states of C lead to the same state of E (redundancy), than if the same state of C leads to a

[17] A mapping F from X to Y is a function iff $F(x_1) = y_1$ and $F(x_1) = y_2$ implies $y_1 = y_2$. A function F is 1–1 iff $F(x_1) = F(x_2)$ implies $x_1 = x_2$. F is onto iff for every y in Y, there exists an x in X such that $F(x) = y$. This characterization may be compared with the characterizations in and Weber (2006) and in Sarkar (2005), which I discovered only after formulating the ideas above. I believe that Sarkar's intent is to capture notions that are very similar to mine, but have some difficulty in understanding how the mechanics of his definitions work. In particular his use of "equivalence classes" seems to make his condition on "differential specificity" redundant; satisfaction of this condition is insured just by the assumption that different elements in the domain of the mapping, a and a', belong to different equivalence classes. In other respects there is close parallelism: Sarkar's condition (ii) that B be "exhausted" is (I assume) just the assumption that F is onto and the intent of his "reverse differential specificity" condition seems to be captured by the assumption that F is 1–1.

Weber (2006) suggests that "causal specificity is nothing but the obtaining of a Woodward-invariance for two sets of discrete variables". Weber's paper is highly illuminating about the role of specificity in Crick's central dogma, but his characterization of specificity is very different than mine: a functional relationship might be invariant and involve discrete variables but not be 1–1 or onto, might relate only two-valued variables (in violation of the "many different states" requirement in **INF**) and might violate the one cause one effect condition described below. Weber's condition seems to me to have more to do with stability than specificity.

.

number of different states of E or if there are many states of E that cannot be produced at all by realizing states of C.[18]

Applied to the passages from Davidson and Waters above, the idea is that there are many possible states of the DNA sequence and many (although not all) variations in this sequence are systematically associated with different possible corresponding states of the linear sequences of the mRNA molecules and of the proteins synthesized. (In some cases there will be a unique corresponding state of mature RNA or protein; in others in which alternative splicing is possible, there will still be a rather limited set of such possible corresponding states.) It is also true of course that because of the redundancy in the genetic code, several different DNA sequences may lead to the same protein, but, as noted above, this is less damaging to control than if the same sequence produced (even without the mediation of other causal factors such as slicing agents) different RNA sequences or different proteins on different occasions. This dependency also involves time and place, in the sense that variations in the time and place at which a particular DNA sequence is activated will systematically lead to variations in the time and place in which particular RNA sequences and proteins are produced.

To the extent such dependency is present, varying the DNA sequence provides for a kind of fine-grained and specific *control* over which RNA molecules or proteins are synthesized. According to Davidson and Waters, this contrasts with the relationship between the presence of RNA polymerase and some other features of the cellular machinery and the specific RNA sequence or proteins produced. Focusing for the moment on prokaryotes in which just one form of RNA polymerase is present, RNA polymerase and other aspects of the cellular machinery are certainly causally relevant to protein synthesis (in the sense captured by **M**)—as Waters says, by reducing the amount of RNA polymerase in the cell, one may interfere with the synthesis of RNA molecules or modify the rate of synthesis. However, one cannot modulate or influence which linear sequences of RNA are synthesized in a fine—grained way by altering the amount of RNA polymerase that is present. In other words, the functional relationship between DNA and RNA sequences is such that one can alter the latter in very specific ways by altering the former; but this is not so for the relationship between RNA polymerase and RNA sequence. The role of RNA polymerase in RNA and protein synthesis instead seems more *switch-like*.

A pure case of switch-like causation would be a case in which a causal factor S affects an outcome E in the following way: E can be in a number of different possible states, $e_1,\ldots,e_n$. S can be in just two possible states ("on" and "off"). S is causally relevant to E in the sense that under some conditions (having to do with, e.g., the state of some third variable C), changing the state of S from onto off or vice versa will change the state of E from one alternative e_1 to another e_2. However, most changes from one state of E to another are not achievable just by changing S—one can't affect whether e_i rather than e_j is realized for most values of i and j, just by

[18] This way of formulating matters makes it clear that Proportionality and specificity in the sense of **INF** are related notions. To the extent that, e.g., there are states of E that cannot be reached by realizing states of C, there will be a failure of proportionality.

varying *S*. In this case, *S* has little influence in the sense of **INF** over *E*—one can't use *S* to modulate or fine-tune the state of *E*. Instead the influence of *S* on *E* is relatively coarse-grained. If, on the other hand, for one of the states of *S*, variations in the state of *C* are available which will affect which of each of the possible states of *E* is realized, *C* will have a considerable amount of fine-grained influence over *E*. In contrast to *S*, *C* will be a relatively specific cause of *E*.

A simple example is provided by a radio with (a) an on/off switch and (b) a rotary dial, the position of which controls which of a number of possible stations is received and hence the content of what is heard. Here (a) is a switch in the sense characterized above. The position of the dial (b) influences, in Lewis sense, the station and content—there are many possible positions of the dial, many possible stations, and a systematic relationship between these such that the position of the dial gives one relatively fine grained control over which station is received, assuming that the switch is on. In this sense, the relationship between the position of the dial and the station received is relatively causally specific. By contrast, while the state of the on/off switch is causally relevant in the sense of **M** to whether any station at all is received, the switch has little influence on which station is received—one can't modulate or fine-tune which station is by varying the state of the switch. In this sense, the switch is not causally specific with respect to which program is received. For similar reasons the writing on the paper placed in a copying machine is a more specific cause of what comes out of the machine than the state of the on/off switch for the machine or whether it is plugged in.

Whether RNA polymerase or various other aspects of the cellular machinery involved in protein synthesis approximate to a switch-like role is an empirical question. The point of my discussion is not to try to settle this question but rather simply to observe that to the extent that one variable C_1 plays a switch-like role with respect to an outcome *E* and a second variable C_2 has influence in the sense of **INF** on *E*, then even though both are causally relevant to *E* in the sense of **M**, there will be a causal asymmetry in the way they affect *E*. This asymmetry arises because fixing C_1 to some specific value (on), many different states of *E* will be associated with different states of C_2, while fixing C_2 to some specific value, only two possible states of *E* will be associated with different states of C_1. My suggestion is that this asymmetry is one thing one might have in mind in claiming that the DNA is a causally specific factor in protein synthesis in a way that the presence of RNA polymerase is not. Moreover, this asymmetry (when it exists) is commonly thought of as having biological significance, as the passages from Davidson and Waters illustrate.

There are many other examples drawn from biological and psychological contexts involving causal factors that exhibit, to varying degrees, either fine-grained influence or more switch-like, coarse-grained behavior. Massive damage to the dorsal lateral prefrontal cortex is causally relevant to performance on tests of IQ and short term memory, but one cannot change a subject's test scores by small amounts in a controllable way by imposing such damage. In this sense, brain damage is a non-specific cause of test performance. Examples of highly specific causes (again in the sense of conforming to **INF**) of test performance are arguably harder to find, but one might imagine, for example, that length of a list of items to be memorized or the

amount of time allowed for memorization might influence test performance in a more specific way than measures that interfere with general health.

In my discussion so far, I have suggested that the notion of specificity (or at least an important aspect of it) might be understood in terms of fine-grained influence. However, one also finds in the biological literature a second understanding of specificity that seems prima-facie rather different from the notion captured by **INF**. Put very roughly, this second idea is that a causal relationship is specific to the extent that a single (type of) cause produces only a single (type of) effect and to the extent that each single type of effect is produced only by a (type of) single cause. A non-specific causal relationship in this sense is one in which tokens of several different types of causes produce (are sufficient in the circumstances for) the same effect (e.g., both smoking and asbestos exposure cause lung cancer) or a single cause (smoking) produces a number of different effects (lung cancer, heart disease).

In epidemiology, a classic source for this idea is Hill (1965). One of Hill's examples is the increased incidence of two different kinds of cancer (cancer of the lung and cancer of the nose) among nickel refiners exposed to certain chemical processes in South Wales in the early twentieth century. According to Hill, two relevant facts are that the workers show only an increased incidence of these two kinds of cancer (rather than an increased incidence of cancer more generally) and that the increased incidence occurs only among workers at certain specific sites during the period 1900–1923, after which the chemical processes employed in the refining at those sites were changed. Hill describes this as illustrating "specificity of association". He writes:

> specificity of the association [is the] the third characteristic [pertaining to whether a relationship is causal] which invariably we must consider. If, as here, the association is limited to specific workers and to particular sites and types of disease and there is no association between the work and other modes of dying, then clearly that is a strong argument in favor of causation. (1965, p. 297)

The epidemiologist Susser characterizes this notion of specificity in the following way:

> By the term *specificity of association*, then, we describe the precision with which the occurrence of one variable will predict the occurrence of another. The ideal, a one to one relationship, encompasses the element of strength of association as well as of precision, and might be better reduced to the statement that one thing and only one thing causes one effect… (1977, p. 13)

Kendler, in the paper cited above, appeals to a similar notion of causal specificity in the context of his discussion of genetic causation:

> The second criterion to evaluate the appropriateness of the concept of "*X* is a gene for *Y*" is the degree of specificity in the relationship between *X* and *Y*. …does *X* influence risk for any other disorders in addition to *Y*? Or are there other genes that contribute to *Y* in addition to *X*? (2005, pp. 395–396)

Kendler characterizes Mendelian genes as having

> quite specific phenotypic effects. That is, one gene influenced pea color but not shape or height while another influenced shape but not height or color. (2005, p. 396)

He then asks:

> How specific are individual genes in their impact on risk for psychiatric disorders? Do most genes influence risk for one and only one psychiatric disorder? (2005, p. 396)

His answer is that

> genetic risk factors for psychiatric disorders are often nonspecific in their effect. A large-scale twin study of seven psychiatric and substance use disorders found one common genetic risk factor predisposing to drug abuse, alcohol dependence, antisocial personality disorder, and conduct disorder and a second common genetic factor influencing risk for major depression, generalized anxiety disorder, and phobia. (2005, p. 396)

Here non-specificity is associated with pleiotropy or the extent to which the same gene has many "different" sorts of effects. Kendler also considers, however, the extent to which single psychiatric disorders are influenced by many different genes, noting that there is evidence that susceptibility to many common disorders (bipolar disorder, schizophrenia) is affected by multiple loci. Kendler's general conclusion is that to the extent that the relationship between gene and disorder is causally non-specific (in either the sense that the gene causally influences many disorders or in the sense that the disorder is influenced by many different genes) the characterization of the gene as a gene for the disorder is in some way misleading or non-perspicuous.

One also finds a similar notion of specificity in other biological contexts. For example, enzymes are commonly described as "very specific" with respect to the substrates on which they act and the reactions they catalyze. This is usually understood to mean that a particular enzyme will often act only a particular substrate or a small set of these (rather than a large class of different substrates) and that it will catalyze just one kind of reaction with this substrate. In other words, the smaller the number of different substrates an enzyme can bind, the greater its "specificity".

The familiar "lock and key" account of enzyme action, originally due Emil Fischer, represents one possible model of this sort of specificity—a model now thought to inadequate but which nonetheless illustrates the idea under discussion. Fischer's idea was that the enzyme and its substrate possess complementary geometric shapes that fit together in the specific and precise way a lock and key do. Suppose that we have a variety of different keys and locks such that each particular make or shape of key will fit and open one and only one variety of lock and, conversely, each different variety of lock can be opened by only one kind of key. Then the relationship between the keys and locks is specific in the sense that we are presently interested in. A master key or (even better, a crowbar) which could be used to open all locks would be the counterpart of a non-specific cause.

A similar notion of specificity is also commonly invoked in characterizations of the mammalian immune system. In particular, the immune system is often described as having a high degree of specificity in the sense that different antibodies are formed in response to different antigens and these antibodies interact preferentially with those antigens and not others. If instead the immune system produced only a small number of general purpose agents which responded to large numbers of different antigens and which were capable of attacking a range of different sources of infection, then it would be less "specific" in its operation. Similarly, a general antiseptic agent such as hydrogen peroxide is non-specific in its effect on pathogens since it will kill many different kinds of pathogens.[19]

Let us call this notion of specificity the *one cause–one effect* notion. This notion raises several questions. One concerns its relationship to the notion described by **INF**. However, a second and prior question is whether the one cause–one effect notion has any plausible application to biological systems. As many commentators have noted, many-many causal relationships (that is, causal relationships in which effects result from the operation of many causes and in which causal agents have many effects) seem ubiquitous in biology (and, for that matter, in many other contexts). In epidemiology, this is often advanced as a reason for rejecting specificity (in its one cause–one effect interpretation) as a "criterion" of causality: it is obviously misguided to conclude that smoking is not a cause of lung cancer on the grounds that it causes many other diseases as well. Thus, after introducing the notion of specificity in the passage quoted above, Susser mentions the many diseases caused by smoking and goes onto say:

> Arguments that demand specificity are fallacious, if not absurd. There can be no logical reason why any identifiable factor, and especially an unrefined one, should not have multiple effects. (1977, p. 13)

Indeed, Hill himself, in a passage that immediately follows the one quoted above, goes onto say, regarding specificity:

> We must not, however, over-emphasize the importance of the characteristic. Even in my present example there is a cause and effect relationship with two different sites of cancer—the lung and the nose.

He adds:

> We must also keep in mind that diseases may have more than one cause. It has always been possible to acquire a cancer of the scrotum without sweeping

[19] This one-cause-one-effect notion of specificity is also closely intertwined with the notion of an intervention, as discussed in Woodward (2003). One wants the relationship between an intervention *I* and the variable *C* intervened onto be "targeted" or surgical in the sense that *I* affects *C* but does not indiscriminately affect other variables—in particular, those that may affect the candidate effect *E* via a route that does not go through *C*. A manipulation lacking this feature is not properly regarded as an intervention on *C* with respect to *E*. Thus, to use an example from Campbell's (2006), derived originally from Locke, pounding an almond into paste is not a good candidate for an intervention on its color because this operation alters so many other properties of the almond. Often, as this example illustrates, the most causally significant variables in a system will be those we can manipulate specifically. Moreover, in many cases, these will be "mechanical" variables like position, density etc.

> chimneys or taking to mule spinning in Lancashire. One-to-one relationships are not frequent. Indeed I believe that multi-causation is generally more likely than single causation though possibly if we knew all the answers we might get back to a single factor. (1965, p. 297)

I agree that these observations about many-many causal relationships are a good reason for rejecting the contention that specificity understood in terms of one cause-one effect is a "criterion" of causation in the sense that it is necessary condition for a relationship to qualify as causal at all. However, in the spirit of the remarks at the beginning of this essay, one may then go onto explore a different question: does specificity (either in the one cause–one effect or any other sense) mark some important or interesting distinction *among* those relationships that are causal? That is, even if it would be wrong to conclude that a candidate causal relationship that is non-specific in the one cause-one effect sense is not genuinely causal, does this notion of specificity (or some plausible reconstruction) capture an interesting or important feature of some causal relationships?

Let me begin with a friendly amendment to the one-cause one-effect idea. As already noted, if one thinks of candidate causes and effects in a completely unrestricted way, it seems uncontroversial that most causes will have very many different effects and many effects can be produced by different causes. For example, any individual molecule will, in addition to the possibly biologically interesting interactions it undergoes with other molecules, exert a small gravitational force on surrounding molecules.

So let us begin by restricting the one cause–one effect idea in the following way: given a candidate cause, we consider only possible kinds of effects within some limited set or range of alternatives, rather than all possible effects to which the cause may contribute. For example, in connection with enzymes we consider only effects that involve chemical interactions with substrates and not other sorts of effects. In connection with the immune system, we consider various antibodies that may be produced in reaction to a given antigen, but not other possible effects that may be caused by the presence of the antigen. Similarly for the causes of particular kinds of effects: we consider only whether there are alternative possible causes that fall within some pre-specified class all of which can produce the effect of interest. Obviously, in applying this idea, a great deal will turn on how this restricted range is specified. I have no general theory to offer about this, but claim, as the above examples illustrate, that it will often be intuitive enough what sort of range is reasonable and non-artificial in particular cases.

Assuming this restricted range idea, the issue of whether a causal relationship is specific in the sense of satisfying the one cause-one effect condition becomes something like this: is it the case that within the specified range of kinds of effects, a particular kind of cause produces only one kind of effect from that range and is it the case that for a given effect, it is (capable of being) caused only by a single kind of cause within some pre-specified set of alternatives? This can be generalized to make specificity a matter of degree—*C* will be a more (rather than less) specific cause (in the one to one sense) to the extent that it causes only a few different kinds of effects within a pre-specified range.

When restricted in this way, the one cause-one effect idea arguably begins to look more reasonable as a description of some (although obviously not all) causal relationships. Even it has other effects; a particular kind of enzyme often will interact only with single sort of substrate rather than with many different kinds of substrate or at least will interact much more strongly or preferentially with a single substrate rather than alternatives. Similarly, an antigen–antibody relationship may be one–one with respect to alternative antigens and alternative antibodies even if it is not one–one in the sense that each antigen has only one effect in general and each antibody only one cause in general (in the sense that nothing else but the antigen is involved in its production). At the very least it is not trivially false that the immune system is specific (or relatively specific) in this sense.

It remains the case that even assuming this restricted range idea many other relationships of biological interest such as the relationship between common diseases and disease causing agents are, as Hill notes, not specific in the one–one sense. But given that specificity (in any of its senses) is being advanced as a feature that is present in some causal relationships and not others and not as a necessary condition for a relationship to be causal at all, this does not seem an unreasonable result.

Given this understanding, whether a relationship is specific in the one–one sense is obviously going to depend on how we "carve up" or restrict the range of alternative causes and alternative effects employed in making this assessment. For example, the relationship between possession of a particular gene and various gross phenotypic traits is often non-specific in the sense that the gene may figure in the production of many different phenotypic traits. However, if we instead ask about the relationship between the gene and its more proximate effects—the protein or proteins for which the gene codes, the mRNA sequence associated with the gene, or the pre-mRNA sequence, the gene comes closer to the ideal of producing just one effect (or at least just a few effects) among the candidates in some restricted set of alternative effects—e.g., the set which consists of just the possible proteins produced by the gene or the set which consists of different mRNA sequences. In this sense, the relationship between the gene and its more proximate effects is "more specific" than the relationship between the gene and its more distal effects.

More generally, by "splitting" a single variable into several or by lumping what was previously regarded as several variables into a single variable, we may sometimes achieve a representation of a casual relationship according to which it is more specific than previously supposed. As an illustration, return to Kendler's example of a gene *G* that causes liking Mozart. If we employ this level of description of the gene's effects, then, as Kendler notes, *G* will be a gene that has many other effects as well—e.g., liking Haydn, Beethoven, and Brahms. The relationship between *G* and these effects will thus be relatively unspecific in the one–one sense. Moreover, as already noted, the relationship between *G* and any one of these particular effects is likely to be relatively unstable, since whether some one who possess *G* exhibits any of these particular preferences for musicians is likely to depend on the contingencies of exposure, musical training etc. On the hand, if we describe *G* as a gene for (that causes) the single more unified effect of perfect pitch, this relationship is not only likely to be more stable but also more specific in the one

to one sense.[20] While I do not claim it is always preferable to replace causal descriptions that represent a cause as having many different effects with more unified descriptions that describes the cause as having a single effect, I do claim, that depending on the empirical details of the situation under investigation, considerations having to do with stability and proportionality can often push us in the direction of representations of causal relationships that are also more specific in the one–one sense.

With this as background, let us return to the question of the relationship between the two notions of specificity (**INF** and the one–one notion). We should now be able to see that the two notions are, even if not quite the same notion, at least interconnected.

First, paradigm cases in which there is specificity in the sense of **INF** will also be cases which can be represented as approximating to specificity in the one–one sense, assuming an appropriately restricted range of alternative effects. The paradigm of specificity in the sense of **INF** is a relationship that may be represented by a function from a range of possible states of a cause-variable (capable of taking many values) to possible states of a (many-valued) effect-variable that is both 1–1 and onto. Thus, one may think of each *state* of the cause-variable as causing one and only one *state* of the effect-variable, so that (in this sense) the one–one requirement is satisfied with respect to states. For example, to the extent that it is true that each different coding region in the DNA sequence in a prokaryote is associated with a single distinct protein and each such protein is associated with a distinct coding region, one will have both satisfaction of **INF** and a situation in which each cause (among a set of possible causes consisting of different DNA sequences) is associated with one and only possible effect (in the set of proteins synthesized).[21]

What about the converse claim? Does specificity in the one to one sense always give us specificity in the sense of **INF**? It appears not—when the relationship between cause and effect is switch-like, there is little influence in the sense of **INF**, but it might be also natural to think of the cause as having only that particular effect (within some class of possible effects of interest). Perhaps some real-life switches are like this—the only interesting and relevant effect (within some pre-specified range of alternatives) of the light switch may be to determine whether the light is on.

[20] Referring back to Kendler's discussion, recall he describes muteness as a "nonspecific consequence" of the hypothetical gene *X* (which causes mental retardation) in the first of his scenarios. Prima-facie, this may seem puzzling. After all muteness seems, if anything, more specific in the sense of being less abstract and a "narrower" category than mental retardation. The sense in which muteness, in comparison with mental retardation, a non-specific consequence of *X* seems to be that muteness is one of many effects of *X*, in contravention of the one cause-one effect ideal of specificity.

[21] Compare Crick's sequence hypothesis: "the specificity of a piece of nucleic acid is expressed solely by the sequence of its bases, and […] this sequence is a (simple) code for the amino acid sequence of a particular protein" and his association, in his statement of the Central Dogma, of both specificity and "information" with the *precise* determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein" (Crick 1958, 152, 153). The ideas of causal specificity and information are obviously closely linked; as this example illustrates, biologists tend to think of structures as carrying information when they are involved in causally specific relationships. I regret that I lack the space to explore this connection in more detail.

A second connection between the two notions of specificity is this: Suppose that C_1 is a specific cause of some effect E and C_2 a non-specific cause, both in the sense of **INF**—e.g., C_1 is DNA sequence, C_2 is the state of the cellular machinery or the presence of RNA polymerase. Suppose that C_2 is manipulated or changed in some way. The effect of this manipulation on E will depend on which of its many different possible states C_1 occupies. In the common case in which there is actual variation in the state of C_1, C_2 will seem to have many different effects, depending on the state of C_1. Thus, C_2 will be non-specific in the second sense of specificity, violating the one cause–one effect ideal. For example, manipulating the level of RNA polymerase in a cell (a non-specific cause in the sense of **INF**) will typically have many different effects in the sense that it will impact the transcription of many different RNA molecules and the synthesis of many different proteins. Davidson alludes to this in the passage quoted above when he says that the "cellular machinery" is in involved in "many diverse regulatory tasks". So if C_1 is specific cause of E (in the sense of **INF**) and there is actual variation in C_1 and C_2 is a non-specific cause of E (again in the sense of **INF** sense) then C_2 is likely to be non-specific in the one–one sense. (Think of a light that is controlled by a switch and a dimmer—if the state of the dimmer varies continuously, controlling the illumination, the on position of the switch will look non-specific in the sense that it is capable of having many different effects, depending on the position of the dimmer.)

One sees a similar pattern in connection with the other examples of non-specific causes (in the sense of **INF**) described above. Interfering with the operation of someone's heart is non-specific with respect to test performance in the sense that it does not give one fine-grained control over that performance, but it is likely also to affect much else beside test performance. That is, such interference, in addition to not conforming to **INF**, is also likely to be non-specific in the sense of violating the one cause-one effect ideal.

The significance of specificity (and of stability and proportionality)

I have suggested several ways to make sense of the notion of a causal relationship being more or less specific. But why does it matter (why should we care) whether a causal relationship is specific or not? As my discussion above has attempted to illustrate, part of the interest of this notion has to do with the way it connects up with other notions, such as stability and proportionality that we also care about. However, there is also a more general point to be made. One of the guiding ideas of an interventionist approach to causation is that causal relationships are relationships that are potentially exploitable for purposes of manipulation and control; our concern with identifying causal relationships and constructing causal explanations is in part motivated by and structured by our interests in controlling the world. This interest helps to explain why we distinguish between causal and merely correlational relationships but it also influences or structures the way in which we think about causal relationships (and the distinctions we make among causal relationships) in other ways as well. Any relationship that is minimally causal in the sense of satisfying **M** is potentially exploitable for some limited sorts of control, in

the sense that there will be some changes in C in some circumstances such that if we are able to bring them about, this will change E. However, a relationship that is minimally causal in this sense may be much less useful for many control-related purposes than we would like. For example, it is consistent with the satisfaction of **M** that the relationship between C and E is highly unstable, holding only in very special background circumstances B. In this case (particularly if we are only rarely in B) the $C \rightarrow E$ relationship may not be very helpful for control purposes. Similarly if there are many possible states of C and E, and alterations in only a very few states of C are associated with alterations in E, so that we cannot use C to control which of most states of E occur.

Indeed, all of the various causal notions investigated above share the common feature that they have to do with possibilities for more fine-grained, extensive and targeted control than is afforded by satisfaction of **M** alone. For example, other things being equal, causal relationships that are more stable are likely to be more useful for many purposes associated with manipulation and control than less stable relationships. Similarly, for causal relationships that are specific in the sense of **INF**: these often offer in principle opportunities for finer grained modulation of effects than less specific relationships. And similarly for relationships that are specific in the one–one sense. For one thing, when C produces some effect E that we want to manipulate but also produces many other different effects (either at the same time or on different occasions); it will often be the case in biological or biomedical contexts that these additional effects are unwanted or deleterious. At the very least, when we employ a non-specific causal relationship we need to monitor and adjust for the presence of these additional effects. By contrast a causal agent C that, so to speak, specifically targets E, and produces no other relevant effects allows us to avoid such complications. For example, it is a defect in most currently available chemotherapy drugs that they adversely affect not just the cancerous cells that one wishes to eliminate but much else as well, including many healthy cells. Current chemotherapy is highly non-specific in the one–one sense. More specific chemotherapy drugs that target only cancerous cells and leave other cells unaffected would be highly desirable and would provide much finer grained control over cancer.

Control is important in part because of its implications for what human agents can or cannot do, but it is also important in biological contexts independently of this. In many biological systems, the successful operation of control structures involves the use of causal relationships having features like stability and specificity. That is, it is often essential to the effective operation of biological control structures that they not have coarse-grained and indiscriminate, unstable effects on many different systems but that instead that they have precise and specific effects on a limited number of target systems, that they affect target systems in stable ways and so on.[22] This is true of many of the structures involved in gene regulation (as Davidson's talk of the "control circuitry embodied in the DNA" suggests), in the control of immune response, in the control of many biochemical reactions that occur

[22] Here, though, we should keep in mind the caveat in footnote 1: it may be that specific stable control is achieved through the interaction of a number of different agents which taken individually have a much less stable and specific effect on the outcome of interest.

within the cell and so on. In other words, one reason why it matters, in biological contexts, whether causal relationships are stable, specific and so on is that these features are relevant to understanding how some biological structures exercise fine-grained control over others.[23]

Consequences

I conclude, briefly, with several other consequences of the ideas described above. First, consider claims about "causal parity". Suppose that several different factors—e.g., C_1 and C_2—are causally relevant to E in the sense of **M**. It is a natural thought that it is invidious or unprincipled to distinguish (on other than pragmatic grounds) between the causal roles played by C_1 and C_2. As noted in Sect. 1, this argument is often made in biological contexts—for example, both Griffiths and Gray (1994) and Oyama (2000) suggest that once it is recognized that both DNA sequence and "other cellular machinery" are casually relevant to protein synthesis, it is misguided to single out or "privilege" the causal role of DNA. It is urged instead that all causally relevant factors be treated the same, in a spirit of "causal democracy".

This argument overlooks the possibility that even if C_1 and C_2 are both causally relevant to E, the causal relationship between C_1 and E may nonetheless differ from the causal relationship between C_2 and E in virtue of *other* features such as those discussed in this essay. For example, even if C_1 and C_2 are both causally relevant to E, the relationship between C_1 and E may be more specific, more stable or better satisfy requirements of proportionality than the relationship between C_2 and E, thus introducing an asymmetry between the two factors. I take this to be one way of interpreting Davidson's claim about the role of DNA in comparison with "other causal machinery". Of course, even if this claim is correct, it is a further issue whether biologists should focus largely or exclusively on the role of DNA and neglect or downplay the role of other causal factors. (A discussion of this issue is

[23] As a pre-cautionary move, let me try to head off some possible misunderstandings of this argument. When the issue is control by a human agent, whether a relationship is useful or not for that agent of course depends on (among other considerations) the agent's purposes and values. In some cases, potential manipulators may not care that some cause has non-specific effects on many other variables (because they regard those effects as neutral) or may even think of this as making the cause a particular good target for intervention, as when these various non-specific effects are all regarded as undesirable and the cause provides a handle for affecting all of them. For example, smoking and childhood sexual abuse have many non-specific effects, virtually all of which are bad and this provides strong reason for trying to intervene to reduce the incidence of both causes. My discussion above is not intended to deny this obvious point. Rather my claim is simply that causal relationships that are stable, specific etc. have control-related features that distinguish them from relationships that are unstable, non-specific etc. Second, and relatedly, I emphasize that my aim has been the modest one of suggesting some reasons why the distinctions between stable and unstable relationships, specific and non-specific relationships and so on is biologically significant. Obviously nature contains (or at least our representations represent nature as containing) stable, specific etc. and unstable, non-specific relationships. I do *not* claim that the former are always more "important", fundamental, valuable, or more worthwhile targets of research than the latter. One can coherently claim that the distinctions I have described are real and have biological significance without endorsing such contentions about importance. Thanks to Ken Kendler for helpful discussion of this point.

well beyond the scope of this paper.) Nonetheless, to the extent that there are real differences in stability and so on in the role played by one group of causal factors in comparison with others, this opens up the possibility that these differences might justifiably serve as a reason for differential treatments of these factors.

A related point concerns the role that stability etc. may play in capturing other distinctions among different causal factors, all of which may be relevant to an outcome. Consider the notion of an "enabling factor", as in Thompson's (2003) example of building *B* a bridge that enables *P* to perform an action *A* that *P* would not have otherwise performed, such as robbing a bank that would have been inaccessible in the absence of *B*. Here it seems more natural to describe *B* as an enabling (or perhaps background) condition for *A* rather than as cause of *A*, despite the counterfactual dependence between *B* and *A*. A natural conjecture is that, other things being equal, when a factor *X* satisfies **M** with respect to *Y*, but *X* is a non-specific cause of *Y* and/or the *X*–*Y* relationship is unstable, we are more likely to regard *X* as a mere enabling (or background) condition for *Y*.[24] As another illustration, the (elusive) contrast between "permissive" and "instructive" causes sometimes employed by biologists might be similarly understood in terms of the idea that permissive causes are generally non-specific in comparison with instructive causes.

A final consequence has to do with the role (or goal) of "reduction" in biological theorizing. Philosophers with reductionist sympathies tend to emphasize the connection of reduction with very general, abstract goals; such as showing that biological phenomena are "nothing but" physical/chemical phenomena (hence satisfying some ideal of ontological economy). My view is that in many biological investigations the motivation for seeking reductive accounts instead has to do with discovering particular causal relationships that have the features (stability etc.) discussed above. That is, it often (although by no means always) turns out that causal generalizations framed at a relatively fine grained level (in terms of physical/chemical concepts) are more stable, and/or specific, and/or better satisfy requirements of proportionality than causal generalizations that are framed at "higher" macro levels. For example, as noted above, the intermediate links in an overall relationship connecting some genetic regulatory network to a gross phenotypical trait may be more stable than the overall relationship itself and these intermediate links may be best specifiable in physical/chemical terms. This yields a motivation for reduction that is piece-meal rather than global, that is guided by specific empirical considerations bearing on where stable, specific etc. causal relationships are to be found, and that such suggests that the motivation for reduction will be stronger in some empirical circumstances than others.

Acknowledgments Versions of this paper were given as talks at a Boston Studies in Philosophy of Science Colloquium on causation in biology and physics, October, 2006, a University of Maryland conference on causation and mechanisms in April, 2007, at the University of Pittsburgh, October, 2007 and at meetings of the SPSP and the Behavioral Genetics Association in June, 2009. Particular thanks to

[24] I don't claim that these are the only considerations relevant to the classification of a factor as an enabler.

James Bogen, Lindley Darden, Peter Machamer, Sandra Mitchell, Ken Schaffner, Ken Waters, Marcel Weber, and especially Ken Kendler for helpful discussion.

References

Campbell J (2006) Manipulating color: pounding an almond. In: Gendler T, Hawthorne J (eds) Perceptual experience. Oxford University Press, Oxford, pp 31–48
Crick F (1958) On protein synthesis. Symp Soc Exp Biol 12:138–163
Davidson E (2001) Genomic regulatory systems: development and evolution. Academic Press, San Diego
Dawkins R (1982) The extended phenotype: the long reach of the gene. Oxford University Press, Oxford
Gottesman I, Gould T (2003) The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry 160:636–645
Greenspan R (2001) The flexible genome. Nat Rev Genet 2:383–387
Griffiths P, Gray R (1994) Developmental systems and evolutionary explanation. J Phil 91:277–304
Hausman D, Woodward J (1999) Independence, invariance and the causal Markov condition. The Br J Philos Sci 50:521–583
Hill A (1965) The environment and disease: association or causation? Proc R Soc Med 58:295–300
Kendler K (2005) A gene for…: the nature of gene action in psychiatric disorders. Am J Psychiatry 162:1243–1252
Kvart I (2001) Lewis' 'causation as influence'. Australas J Philos 79:409–421
Lewis D (1986) Postscript c to 'causation': (insensitive causation). In: Philosophical papers, vol 2. Oxford University Press, Oxford, pp 184–188
Lewis D (2000) Causation as influence. J Phil 97:182–197
Mitchell S (2000) Dimensions of scientific law. Phil Sci 67:242–265
Mitchell S (2009) Unsimple truths: science, complexity, and policy. University of Chicago Press, Chicago
Oyama S (2000) Causal contributions and causal democracy in developmental systems theory. Phil Sci 67:S332–S347
Rieke F, Warland D, van Steveninck R, Bialek W (1997) Spikes: exploring the nature of the neural code. MIT Press, Cambridge
Sarkar S (2005) How genes encode information for phenotypic traits. In: Sarkar S (ed) Molecular models of life. MIT Press, Cambridge
Susser M (1977) Judgment and causal inference: criteria in epidemiologic studies. Am J Epidemiol 105:1–15
Thompson J (2003) Causation: omissions. Phil Phenomenol Res 66:81–103
Waters K (2007) Causes that make a difference. J Phil CIV:551–579
Weber M (2006) The central dogma as a thesis of causal specificity. Hist Philos Life Sci 28:595–609
Woodward J (1999) Causal interpretation in systems of equations. Synthese 121:199–257
Woodward J (2003) Making things happen: a theory of causal explanation. Oxford University Press, New York
Woodward J (2006) Sensitive and insensitive causation. Phil Rev 115:1–50
Yablo S (1992) Mental causation. Phil Rev 101:245–280