Stability, Breadth and Guidance

Thomas Blanchard, Nadya Vasilyeva and Tania Lombrozo

Philosophical Studies, 175, 2263-2283 (2018)

This version of the article has been accepted for publication, after peer review, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: https://doi.org/10.1007/s11098-017-0958-6

Abstract

Much recent work on explanation in the interventionist tradition emphasizes the explanatory value of *stable* causal generalizations – i.e., causal generalizations that remain true in a wide range of background circumstances. We argue that two separate explanatory virtues are lumped together under the heading of `stability'. We call these two virtues *breadth* and *guidance* respectively. In our view, these two virtues are importantly distinct, but this fact is neglected or at least under-appreciated in the literature on stability. We argue that an adequate theory of explanatory goodness should recognize breadth and guidance as distinct virtues, as breadth and guidance track different ideals of explanation, satisfy different cognitive and pragmatic ends, and play different theoretical roles in (for example) helping us understand the explanatory value of mechanisms. Thus keeping track of the distinction between these two forms of stability yields a more accurate and perspicuous picture of the role that stability considerations play in explanation.

One of the main tasks for theorists of explanation is to account for the features that make some explanations better than others. In recent years, various authors have used the interventionist account of causal explanation (the dominant approach to explanation in philosophy of science) to identify and examine various explanatory virtues. In particular, much recent work in the interventionist tradition follows Woodward (2006, 2010) in emphasizing the explanatory value of invoking *stable* causal generalizations. While we are convinced that stability considerations play an important role in assessing the quality of explanations, we think that two quite different explanatory virtues are lumped together under the heading of `stability'. We call these two virtues *breadth* and *guidance* respectively. In our view, these two virtues are importantly distinct, but this fact is overlooked in the literature on stability. We argue that an adequate theory of explanatory goodness should explicitly recognize breadth and guidance as distinct virtues, and that keeping track of this distinction yields a more accurate and perspicuous picture of the role that stability considerations play in explanation.

Our discussion unfolds as follows. Section I briefly summarizes the main tenets of interventionism and introduces the notion of stability. Section II brings out the distinction between breadth and guidance as two forms of stability. In the remainder of the paper we argue that keeping track of the distinction is important for several reasons. Section III argues that making the distinction explicit helps us articulate more precisely the role that stability plays in various interventionist accounts of explanation, better situate these accounts with respect to one another and identify some of their limitations. In section IV we argue that the breadth/guidance distinction allows us to paint a more perspicuous picture of the cognitive and pragmatic value of stable generalizations. Finally, we argue in section V that guidance has an important but underappreciated role in helping us understand several aspects of our explanatory practices such as the importance of mechanisms. Section VI summarizes the main conclusions.

One remark before we begin. In our view, it is an open possibility that other accounts of causal explanation besides interventionism must also incorporate something like the breadth/guidance distinction. But in this paper we restrict our focus to interventionism, for two reasons. First, the notion of stability is most at home within the interventionist account of explanation. Examining whether the breadth/guidance distinction also matters for other accounts of explanation would require examining what form stability takes in these accounts, a task which is beyond the scope of this paper. Second, it is fair to say that interventionism has now become an extremely influential approach among philosophers interested in explanation, as witnessed by the explosion of work on explanation conducted from an interventionist perspective in the last decade.[1] Thus we think that the distinction we are drawing is important in its own right, independently of whether it applies beyond the confines of interventionism.

[1] Thus interventionism has been applied to explanation in biology (e.g. Woodward 2010), biomedicine (Malaterre 2011), neuroscience (e.g. Craver 2006), psychiatry (e.g. Campbell 2008), and sociology (e.g. Steel 2006). In addition, interventionism has been recruited to shed light on general questions about explanatory levels (e.g. Woodward 2008) and explanatory selection (e.g. Waters 2007). Interventionism is also an influential framework in the cognitive psychology of explanation (see e.g. Lombrozo 2010).

I. Interventionism and Stability

We start with a brief summary of the interventionist theory of causal explanation, focusing solely on explanations of singular events. According to interventionism, a minimally adequate causal explanation consists of three parts. The first part is a statement to the effect that some variable Y (the *explanandum* or effect variable) took value y in the circumstances under consideration. $Y=y$ represents the event to be explained. The second part is a statement to the effect that some variables $X_1 \ldots X_n$ (the *explanans* or cause variables) took values $x_1 \ldots x_n$ in the situation under consideration. These values represent particular events in terms of which the *explanandum* is to be accounted for. The third part is a true general causal claim G (the *explanatory generalization*) describing a relationship of causal dependence of Y on $X_1 \ldots X_n$. In the interventionist framework, causal dependence is understood as *counterfactual dependence under interventions*: roughly, Y causally depends on a variable or set of variables X when an intervention on X would change the value of Y. More precisely, G should be a true causal claim to the effect that if an intervention setting $X_1 \ldots X_n$ at certain non-actual values $x'_1 \ldots x'_n$ had occurred, Y would have taken some non-actual value y'. Heuristically, one can think of an intervention as a causal process that directly sets its target at some value while leaving the rest of the causal structure intact. Characterizing interventions more precisely is a delicate matter[2], but for the purposes of this paper this informal characterization will suffice.

To illustrate, suppose we want to explain why the car accelerated (*Acceleration*=*Yes*). According to interventionism, a causal explanation of this event may consist of (a) a true statement to the effect that the driver depressed the gas pedal (*Pedal*=*Yes*) and (b) a true general causal claim to the effect that depressing the gas pedal causes the car to accelerate, i.e. that setting *Pedal* at value *No* would change the value of *Acceleration* from *Yes* to *No*.

One crucial contention of interventionism is that for a singular causal explanation to be minimally adequate, the explanatory generalization G on which it relies need not hold in *all*

[2] See Woodward (2003, ch. 3).

possible circumstances. (In this way, interventionism recognizes that virtually all relationships of causal dependence break down under certain circumstances.) Instead, *G* need only hold in at least *some* background circumstances, including the ones that obtained in the actual situation. That is, it need only be the case that in at least some background circumstances (including the ones that actually obtained), the explanandum variable would have taken a non-actual value had an intervention on the *explanans* variables set those variables to some non-actual values. Thus the adequacy of the explanation of the car's acceleration isn't impugned by the fact that in some background circumstances (e.*g*., when the pedal isn't properly connected to the motor), intervening on the pedal isn't associated with any change in the value of *Acceleration*. All that matters for minimal explanatory adequacy is that in the actual circumstances, the claim that intervening on the pedal would have changed the value of *Acceleration* was indeed true.

Since the concept of a background circumstance will play a central role in what follows, it will be useful to have an explicit definition of it. Following Woodward (2006, 290), we'll say that a possible state of affairs *B* is a background circumstance relative to a causal claim `X →Y' just in case none of the values of *X* or *Y* convey any explicit information as to whether *B* obtains. For instance, the possible values of *Pedal* and *Acceleration* say nothing about whether the pedal is properly connected to the motor, so this state of affairs constitutes a background circumstance relative to the explanatory generalization `*Pedal→Acceleration*'.

This account of minimally adequate explanations provides an extremely fruitful point of departure for exploring the topic of explanatory virtues – i.e. the question of why, among all explanations that meet these minimal criteria of adequacy, some strike us as more powerful and illuminating than others. Here our focus is on a particularly important explanatory virtue that can be identified within the interventionist framework: the virtue of *insensitivity* or *stability*, discussed in detail in Woodward's (2006) and (2010).[3] As mentioned above, to figure in a successful explanation, an explanatory generalization must hold in at least *some* background circumstances (including the actual circumstances in which the event to be explained took

[3] As we will see below, Hitchcock and Woodward's (2003) – the first sustained attempt to articulate an interventionist theory of explanatory virtues – contains an early discussion of (a certain form of) stability.

place). The stability of an explanatory generalization Y= f(X) is tied to the *range* of background circumstances (i.e. circumstances mentioned in neither X nor Y) in which it holds. To the extent that the generalization holds in a large number of possible background circumstances – in particular, circumstances that we regard as `salient' and `important' – it is relatively stable; whereas if it holds only in a restricted and special set of circumstances, it is relatively unstable.[4] Woodward convincingly argues that the stability of an explanatory generalization importantly contributes to the quality of explanations in which it figures: *ceteris paribus*, an explanation is better insofar as it involves a more stable explanatory generalization.[5] To illustrate the notion of stability and motivate its status as an explanatory virtue, let's consider two examples put forward by Woodward. Each involves a comparison between two explanatory generalizations G and G', one of which is more stable than the other; in each case it is intuitively clear that the more stable generalization is more explanatory.

The first example we'll consider is one that Woodward borrows from Kendler (2005).[6] Suppose we discover that a certain gene *g* causes its bearers to engage in highly risky activities such as bungee-jumping, reckless driving and so on. Assume that Mary, who has gene *g*, bungee-jumped yesterday, and compare two possible ways in which one might explain this event in terms of Mary having *g*. First, one might explain Mary's behavior by appealing to the fact that she has gene *g* and the general causal claim

(1) Gene *g* causes bungee-jumping.

---

[4] Stability is also often called *invariance* (see e.g. Woodward 2010). It is worth noting, however, that in a number of places (e.g. Woodward 2003) Woodward uses the term `invariance' to designate a kind of robustness distinct from stability. Under this use of the term, the invariance of a generalization Y=f($X_1, \ldots, X_n$) depends on the extent to which it continues to hold under a wide range of possible interventions on the values of the independent variables $X_1, \ldots, X_n$. By contrast, stability has to do with the extent to which the relationship continues to hold under changes to factors other than $X_1, \ldots, X_n$. To avoid potential confusion we stick to the term `stability' in this paper.

[5] Comparing explanatory generalizations with respect to stability is a subtle affair. The easiest case is when the range of background circumstances in which a generalization G continues to hold is a proper subset of the range of circumstances in which some other generalization G' continues to hold. In this case, we can say that G is strictly less stable than G'. In cases where the relevant sets of background circumstances in which two generalizations hold are disjoint or only partially overlap, comparative judgments of stability are more difficult and may be impossible if we have no way of measuring the number and relative importance of the relevant circumstances. In this paper we leave aside this issue and concentrate on cases where stability comparisons are straightforward.

[6] See Woodward (2015, 22) and also (2010, 301-2).

Second, one might conceptualize Mary's behavior as an episode of risky activity, and account for this in terms of the fact that Mary has *g* and the general causal claim

(1*) Gene *g* causes risk-taking behavior.

Both explanations satisfy the minimal adequacy conditions put forward by the interventionist theory of explanation. In particular, both (1) and (1*) correctly describe patterns of counterfactual dependence under interventions that hold in at least some background circumstances (including, we may stipulate, the circumstances that actually obtained in Mary's case). But (1*) is a more stable explanatory generalization than (1). (1) holds only in a very special set of circumstances that may easily fail to obtain: namely, when the bearer of the gene can afford bungee-jumping, lives near bungee-jumping infrastructures, etc. By contrast, (2) presumably holds in a relatively wide set of circumstances – namely, any set of circumstances in which some risk-taking activity is available to the gene bearer. These include circumstances conducive to bungee-jumping, but also many others (e.*g*., circumstances in which *g* causes reckless driving). This difference in stability plausibly explains why the second explanation strikes us as much better than the first one. Specifically, the first explanation strikes us as defective in an important respect, and the reason seems to be that the explanatory generalization on which it relies holds only in a very restricted set of circumstances. By contrast, the second explanation seems perfectly fine, precisely because the explanatory generalization it recruits is one that holds in a relatively wide variety of background circumstances.[7]

---

[7] One might think that there is a simpler account of the superiority of (1*) over (1). After all, causes raise the probability of their effects, at least typically: and the more they do so, the stronger the causal relationship is. However we interpret the relevant notion of `probability', having gene *g* will presumably raise the probability of risk-taking behavior to a much larger extent than it raises the probability of bungee-jumping in particular. The superiority of (1*) might therefore be explained by the fact that it mentions a much stronger causal relationship than (1) does.

We think that there is something plausible to this account, and that there are important and under-explored connections between stability and probability. But there are important caveats. First, if `probability' simply means actual frequency, a causal relationship may be probabilistically strong and yet still be unstable in a way that reduces its explanatory power. For instance, it may be that coincidentally, all bearers of gene *g* are located in areas where bungee-jumping is the most easily accessible form of risk-taking behavior, in which case (1) and (1*) will be equally strong. Yet surely (1) would still be explanatorily defective in such an extraordinary circumstance. This means that any probabilistic account that suitably explains the superiority of (1) will presumably have to involve a

Here is a second example from Woodward (2015, 23-4), who draws on Spirtes and Scheines (2004). As is well-known, there are two kinds of cholesterol, high-density and low-density. High levels of low-density cholesterol (LDC) are a major risk factor for heart disease, whereas high levels of high-density cholesterol (HDC) are by and large harmless. Suppose that John has both a high level of LDC and heart disease, and compare the following two explanations. The first one explains John's heart disease in terms of his high level of cholesterol, together with the explanatory generalization

(2) High levels of cholesterol cause heart disease.

The second one explains John's heart disease by appealing to John's high level of LDC and the generalization

(2*) High levels of LDC cause heart disease.

It seems uncontroversial that the second explanation is more illuminating than the first, and here again the difference can be traced back to a difference in stability between the relevant explanatory generalizations. Specifically, the first explanation seems defective in an important respect because it fails to mention an important condition required for cholesterol to cause heart disease: viz. that the kind of cholesterol in question must be LDC. As a result, the generalization is fairly unstable: there is one salient kind of background circumstance (i.e. the cholesterol in

---

robustly modal notion of probability that is sensitive to the range of possible background circumstances in which the cause raises the probability of its effects. And such an account will amount to a probabilified version of the notion of stability. Second, all probabilistic measures of causal strength that we know of are functions of the *average* values of P(E/C) and P(E/not-C) in a population, and hence measure the average strength of the causal relationship in the population. Yet two relationships that are on average equally strong need not be equally stable: for instance, one may hold in all segments of this population while the second holds strongly in some segments of the population and not at all in others. There is evidence that in such circumstances, our explanatory practices still favor the more stable generalization (see Vasilyeva et al. 2016). Thus current probabilistic measures of causal strength cannot capture stability nor account for its role in our explanatory judgments. In addition, as pointed out to us by [name omitted for blind review] (p. c.), the notions of causal strength (understood probabilistically) and of stability are conceptually distinct: while the former requires a probability measure of all the circumstances relevant to the value of P(E/C), a causal generalization can be judged as more or less stable even in contexts where the relevant probabilities are unknown or undefined.

question being HDC) in which it fails to hold.[8] Because (2*) makes it explicit that only LDC causes heart disease, it is more stable, which plausibly accounts for our sense that an explanation of John's heart disease that appeals to this generalization is just fine.

II. Two Kinds of Stability

In our view, these examples (as well as the additional considerations marshalled by Woodward in his 2006 and 2010 papers) convincingly show that stability considerations play an important role in assessing the quality of an explanation. Yet we think that there are two different explanatory virtues that fall under the heading of `stability'. In this section, we introduce and explain these two virtues. In the remainder of the paper we will argue that these two virtues are importantly distinct, in a way that is obscured by lumping them together under the heading of `stability'.

The distinction we have in mind can be brought into focus by noting an important difference between the two examples discussed in the previous section: while (1*) and (2*) are more stable than (1) and (2) respectively, the increase in stability is not achieved in the same way in these two cases. To show this, let's say that a generalization *G* is *broader* than a generalization *G′* just in case every background circumstance in which *G*′ holds is a background circumstance in which *G* holds but not *vice versa*. Breadth is a form of stability: a generalization that is broader than another holds in a wider range of background circumstances and is thereby more stable. In our gene example, (1*) is more stable than (1) precisely because it is broader: every background circumstance in which gene *g* causes bungee-jumping is a circumstance in which it causes risk-taking, but not every background circumstance in which *g* causes risk-taking behavior is one where *g* causes bungee-jumping. As another example, consider the fact that pregnancy

[8] The claim that whether the kind of cholesterol under consideration is HDC or LDC is a background circumstance relative to (2) may appear to stretch the meaning of `background circumstance', but remember that we are using the term in a semi-technical way: a possible situation or state of affairs *B* is a background circumstance relative to a generalization X →Y just in case neither *X* nor *Y* encode any information about whether *B* holds. On this definition, whether the kind of cholesterol we're dealing with is HDC or LDC is a background circumstance relative to (2).

sometimes causes pulmonary embolisms by causing thrombosis, which is itself a major cause of pulmonary embolisms. Pregnancy is not the only cause of thrombosis, which can also be caused by (e.g.) birth control pills or smoking. As a consequence, the explanatory generalization

(3) Pregnancy causes pulmonary embolisms

is less broad and hence less stable than

(3*) Thrombosis causes pulmonary embolisms.

The reason is that pregnancy causes pulmonary embolisms solely by way of causing thrombosis, so that (3*) holds in every circumstance in which (3) holds. But since pregnancy isn't the only cause of thrombosis, there are circumstances in which thrombosis causes pulmonary embolisms but where (3) does not hold (e.*g*. when thrombosis was caused by smoking, not pregnancy). More generally, when we have a causal chain $A \rightarrow B \rightarrow C$ such that A causes C only by causing B and B can be caused by other factors besides A, the causal relationship $B \rightarrow C$ is typically broader and thereby more stable than the $A \rightarrow C$ relationship.

However, turning to the second example discussed in the previous section, note that while (2*) is more stable than (2), it isn't broader. After all, every circumstance in which a high level of LDC causes heart disease is a circumstance in which a high level of cholesterol causes heart disease – and the reverse holds as well, since low-density cholesterol is the only form of cholesterol that causes heart disease. What makes (2*) more stable than (2), then, isn't the fact that it holds in a broader range of background circumstances. Rather, (2*) enjoys a higher degree of stability because it makes explicit a background circumstance necessary for high cholesterol to cause heart-disease (viz. that it be low-density) and which (2) leaves unmentioned. As we might also put it, (2*) achieves a higher degree of stability than (2) not by being broader but by being more *guiding*: even though both generalizations apply in the same circumstances, (2*) explicitly conveys more information than (2) about what those circumstances are.

These examples thus show that there are two ways in which a generalization G can be more stable than G': by holding in a wider range of circumstances (breadth), or by making explicit certain background conditions required for *G′* to hold (guidance). Note that breadth is naturally glossed in terms of *generality*, whereas guidance is more accurately glossed in terms of *accuracy*. For example, because (3*) holds in a wider range of possible background circumstances than (3) does, it is appropriate to say that the former is more general or has wider scope than the latter. Not so for guidance: for instance, it would be strained to say that (2*) is more general than (2), and more appropriate to say that it provides a more *accurate* representation of the scope of the cholesterol – heart disease relationship. There is another illuminating way to unpack the distinction between breadth and guidance – one that also supports the idea that breadth and guidance are explanatorily valuable for distinct reasons. On the one hand, when a generalization *G* is broader than *G′*, *G* is more valuable than *G′* because it is more *inclusive*: it includes cases that should be included but that *G′* fails to include. For example, while (1) applies only to particular cases in which somebody has gene *g*, (1*) correctly describes what happens in a much wider range of cases in which somebody has gene *g*. Put differently, the superiority of (1*) over (1) has to do with the fact that the former but not the latter includes instances that should be included in an appropriately general description of the effects of gene *g*. Likewise, generalization (3) applies only in a very small and specific range of cases where the patient has pulmonary embolism, whereas (3*) holds true in a much wider range of cases of pulmonary embolism. Here the superiority of (3*) over (3) has to do with the fact that the former but not the latter includes instances that should be included in an appropriately general description of the causes of embolism.[9]

On the other hand, when *G* is more guiding than *G′*, *G* is more valuable not by being more *inclusive*, but by *excluding* cases that should be excluded and that *G′* fails to exclude. To make this concrete, return to the contrast between the two generalizations about heart disease (2) and (2*). As we pointed out, these generalizations hold in exactly the same actual instances: every instance in which LDC causes heart disease is an instance in which high cholesterol causes heart disease and *vice versa*. The difference between the two generalizations has to do with the fact that (2*) explicitly excludes instances in which the relevant causal relationship does not hold

[9] In this respect, breadth is tightly connected to the explanatory virtue of *proportionality* (Woodward 2010).

and which (2) fails to exclude, namely instances in which the relevant form of cholesterol is HDC rather than LDC.

Cast in this way, the breadth/guidance distinction can be seen as latching onto two different responses to the fact that virtually all causal generalizations we are in a position to discover and articulate involve exceptions.[10] When a generalization `X causes Y' involves exceptions, we can deploy two different strategies for arriving at a more reliable generalization. One strategy is to identify the range of circumstances in which the generalization doesn't hold and explicitly exclude those circumstances from the scope of the generalization, as when we move from (2) to (2*). In other words, we look for a more guiding generalization by incorporating what were previously background circumstances into our new generalization. (Note that these circumstances can be built into the characterization of the effect, rather than the cause, as in a case where we amend the claim "smoking causes cancer" to "smoking causes lung and throat cancer.") But another strategy is to jettison `X causes Y' in favor of a more wide-ranging and less exception-ridden description of the effects of X or the causes of Y. In other words, the strategy is to look for a broader generalization. This is what happens when we redescribe the effect of having gene *g* as risk-taking rather than bungee-jumping or when we come to understand the causal relationship between pregnancy and pulmonary embolism as mediated by a broader causal relationship between thrombosis and pulmonary embolism.

We take the foregoing considerations to make at least a strong *prima facie* case for the claim that breadth and guidance are two importantly distinct explanatory virtues. Yet the breadth/guidance distinction has not been recognized in the literature on stability.[11]

---

[10] When we consider exceptionless generalizations (such as, presumably, generalizations describing universal physical laws), the breadth/guidance distinction disappears. An exceptionless generalization is both maximally broad and guiding: if the generalization holds in all physically possible circumstances, it is by definition maximally broad and maximally guiding insofar as there are no background circumstances required for it to hold and *a fortiori* no such background circumstances that the generalization fails to makes explicit. It is only when we consider generalizations that fall short of holding in all possible circumstances that the distinction between our two kinds of stability can be drawn.

[11] However, we note that Potochnik (2015) offers a causal approach to explanation (although not a specifically interventionist one) that recognizes something like breadth and guidance as independent explanatory virtues. Indeed, Potochnik argues that some causal explanations are especially valuable because they clearly outline the

Woodward's recent discussion of stability covers both breadth and guidance, as witnessed by the fact that it contains examples illustrating both, yet he doesn't explicitly distinguish between them.[12] As we will see in the next section, some interventionist accounts of explanatory virtues in which `stability' figures prominently in fact cover only one of its two forms, yet do not make this explicit – presumably because the very fact that two distinct virtues are lumped together under the heading of `stability' has been so far unrecognized. In what follows, we will argue that keeping track of the distinction is important for several reasons.

III. Breadth, Guidance, and Interventionist Theories of Explanatory Virtues

The first reason why the breadth/guidance distinction is important and theoretically useful is that it helps us articulate a better picture of some prominent interventionist accounts of explanatory virtues and to situate them more clearly with respect to one another. Here we examine the accounts of Hitchcock and Woodward (2003) and of Ylikoski and Kuorikoski (2010). Both accounts of explanatory virtues are couched in interventionist terms and give a central place to considerations of stability. Upon closer examination, however, each account turns out to cover only one form of stability (guidance in the first case, breadth in the second) as a virtue. Making the distinction allows us explicitly thus allows us to bring out two important points that may be obscured by the single heading of `stability': the fact that despite their superficial similarities these accounts differ in important ways (and are in fact associated with different views of the ideal of explanation), and the fact that each of them arguably fails to capture the full extent to which stable generalizations are explanatorily valuable.

On Hitchcock and Woodward's (2003) account, an explanation is better (or, in their preferred terminology, `deeper') insofar as the explanatory generalization it involves can correctly answer

---

scope of the causal dependence pattern they pick out (which corresponds to what we call `guidance') but also recognizes (p. 1173) that picking out causal patterns of suitably broad scope matters for explanation.

[12] That being said, as we will see in section VI, in some of his discussions of the theoretical advantages of `stability' Woodward seems to have in mind breadth at the exclusion of guidance. Thus there is a tension in Woodward's discussion of the subject. One theoretical advantage of making the breadth/guidance distinction explicit is that it brings the tension to the foreground.

more `what-if-things-had-been-different' questions, or *what-if* questions for short. An explanatory generalization *G* correctly answers a *what-if* question of the form "what if *X* had taken value *x′*?' just in case *X* is one of the explanans variables figuring in *G* and the value assigned to the explanandum variable *Y* by *G* under the input `*X* = *x*" is the one that *Y* would indeed have taken had an intervention set *X* at value *x′*. While Hitchcock and Woodward do not use `stability' in the relevant sense of the word, their account is often presented as making what we call stability – robustness under background conditions – a central explanatory virtue.[13] Yet it is important to note that their account leaves room only for guidance – not breadth – as a virtue. Specifically, on their view, "perhaps the most fundamental way" (2003, 188) in which a generalization *G* can provide a better explanation than another generalization *G′* is by making explicit some factor causally relevant to the *explanandum* but was left unmentioned by *G′*. This is simply what we call *guidance*. The explanatory value of guidance is due to the fact that by making explicit the explanandum's dependence on *X*, *G* allows us to answer certain *what-if* questions on which *G* remains silent, viz. questions about whether and how the explanandum would vary under changes in *X*'s value. To borrow one of Hitchcock and Woodward's examples, suppose we want to explain why an object dropped from the top of the Pisa tower took time *t* to reach the ground. One way is to appeal to Galileo's law of free fall, and another is to appeal to Newton's second law. On their view, the second explanation is deeper because it highlights certain facts on which the explanandum depends and on which Galileo's law remains silent, e.g. that the Earth has certain mass *m*. This allows the second explanation to answer certain *what-if* questions on which the first remains silent, e.*g*. questions about how long the object would have taken to reach the ground had Earth's mass been different.

While Hitchcock and Woodward's account gives pride of place to what we call *guidance*, as far as we can see it leaves no space for *breadth* to count as an additional virtue, at least if we take the motivating idea behind their account literally. For instance, contrast an explanation of Jane's pulmonary embolism that mentions her pregnancy and generalization (3) with an explanation that instead mentions her thrombosis together with generalization (3*). Note that the sets of

[13] See for instance Weslake (2010, 278), who calls stability `portability'.

what-if questions that those explanations can truthfully answer are non-overlapping: the former answers questions about what would have happened had Jane not been pregnant, and the latter answers questions about what would have happened had Jane not been suffering from thrombosis. So there is no clear sense in which the latter explanation answers *more* w-questions than the former, and consequently no reason to count it as deeper on Hitchcock and Woodward's account.

Ylikoski and Kuorikoski (2010) offer another account of explanatory power couched in an interventionist framework, where once again stability (or `insensitivity' as they call it) plays a prime role – it is the first of the five explanatory virtues recognized by their account. Yet we think that on the most plausible way to interpret their view, what they mean by `stability' is what we have called *breadth*. Thus consider the following passage:

> Our first dimension of explanatory power is insensitivity of the explanatory relationship with respect to changes in background conditions… The more insensitive the explanatory relationship is with respect to background conditions, the more independent it is from these conditions. *This means that the same answer would be correct for a larger group of what-if questions*… (2010, 208-9; our emphasis).

The italicized sentence here is crucial. First, note that like Hitchcock and Woodward, Ylikoski and Kuorikoski appeal to sets of what-if questions. However, Ylikoski and Kuorikoski introduce these questions to impose a constraint on good explanatory generalizations that differs from Hitchcock and Woodward's. When Hitchcock and Woodward say that an explanatory generalization Y=f(X_1… X_n) answers what-if questions, they mean that it allows us to predict the value Y would take if X_1… X_n were to take certain values x_1… x_n, and an explanatory generalization is better to the extent that it can do so for more possible values of X_1… X_n. In contrast, Ylikoski and Kuorikoski introduce what-if questions to identify explanatory generalizations that give `the same answer' in a broad range of cases: an insensitive generalization ascribes to Y its actual value for a large range of counterfactual circumstances. Moreover, these counterfactual circumstances cannot be alternative value assignments to X_1…

X_n,[14] but must instead involve variations in background conditions. If this reading of the italicized sentence in the quote above is correct, then on their view, an explanatory generalization is stable insofar as for a large group of such background circumstances, Y would still have taken its actual value as long as $X_1 \ldots X_n$ had retained their actual values.

If this reading is correct, the kind of stability at the center of Ylikoski and Kuorikoski's account is breadth, not guidance. By way of example, suppose again that Mary bungee-jumped yesterday. One might attempt to explain this via the fact that Mary has gene *g* together with (1) or via the fact that Mary has gene *g* together with the broader generalization (1*), in which case the *explanandum* would be recast as an instance of risk-taking behavior rather than bungee-jumping specifically. There are many non-actual circumstances in which even if Mary still had gene *g* she wouldn't have bungee-jumped, e.*g*. if bungee-jumping had been too expensive for her. But in those non-actual circumstances, Mary would presumably have still engaged in some form of risk-taking behavior. Hence the answer given by the second explanation (that Mary engaged in risk-taking behavior because she had gene *g*) would still have been true in those circumstances, while the answer given by the first (that Mary bungee-jumped because she had gene *g*) wouldn't have. So the second explanation better satisfies Ylikoski and Kuorikoski's stability desideratum and is thus counted by their account as better than the first. And it does so because it involves a broader generalization, i.e. a generalization that involves a wider range of circumstances.

On the other hand, Ylikoski and Kuorikoski's definition of stability doesn't capture guidance. For instance, contrast again the two possible explanations of John's heart-disease in terms of (2),

---

[14] To see why Ylikoski and Kuorikoski must mean variation in background circumstances – and not in $X_1 \ldots X_n$, consider the following example. Suppose we want to explain why some water sample is frozen. One way to do so is to mention that the temperature in the room is -17.3 Celsius, together with a generalization that maps every possible fine-grained value of the temperature to the state of the water, and thus entails that in the actual circumstances the water must have been frozen. This generalization gives the same answer - viz. that the water is frozen – for many possible values of its explanans variables, namely every value below 0 Celsius. Yet this has nothing to do with the (in)stability of the generalization. Indeed, this feature of the generalization is a vice rather than a virtue in the present context, since it means that the corresponding explanation doesn't cite a cause that is `proportional' to the effect, by contrast to an explanation that merely mentions the fact that the temperature was below 0. See Woodward (2010) for a discussion of proportionality and its status as an explanatory virtue within the interventionist theory of explanation.

and (2*). It would be wrong to say that the answer provided by the second explanation would still be correct in a larger range of circumstances than the answer provided by the first explanation. After all, every instance in which (2) holds is an instance in which (2*) holds and *vice versa*. Rather, what makes the second explanation a better one is that it makes more explicit an important characteristic of all those instances (viz. that they all involve John having *low-density* rather than *high-density* cholesterol). Taken literally, Ylikoski and Kuorikoski's account of stability doesn't capture this dimension of explanatory power; in fact, as far as we can see, their complete list of explanatory virtues doesn't allow us to capture the fact that an explanation of John's heart disease that appeals to (2*) is better than one that appeals to (2).

This examination of the place of stability in current interventionist accounts of explanatory virtues holds an important lesson. We take it that insofar as the thesis that *both* breadth and guidance are explanatory virtues is *prima facie* plausible, the fact that the theories we just considered only recognize one form of stability as a virtue constitutes a limitation of these accounts. Yet this is obscured by the fact that both breadth and guidance are lumped together under the heading of `stability', which may give the impression that an account that captures one of these two kinds of stability *ipso facto* captures the other.

Once we have the distinction between breadth and guidance clearly in mind, it also becomes easier to appreciate why an account of explanatory goodness that recognizes one as an explanatory virtue may not recognize the other: both virtues contribute to the quality of an explanation in different ways. On the one hand, a broad generalization seems to contribute to the quality of an explanation insofar as it shows that the occurrence of the explanandum was to a certain extent *independent* of the background circumstances as they actually unfolded: had those circumstances been somewhat different, the explanandum would still have occurred. In showing that the explanandum's occurrence didn't depend on the minute details of the situation as it actually transpired, a broad explanatory generalization shows that the explanandum was in a certain respect *bound to happen*, whatever the actual circumstances turned out to be.[15] Thus an emphasis on breadth fits naturally with a view of explanation on which the ideal to which we aspire in explanation is to account for the explanandum in a way that *abstracts* as much as possible from the details of actual circumstances.[16] (We speculate that Ylikoski and Kuorikoski's account is largely driven by this idea, hence the privileged role of breadth on their account.)

Guidance, on the other hand, has little to do with showing that the occurrence of the explanandum was independent of the exact shape taken by the background circumstances in the relevant situation. In fact, it is precisely the reverse: a more guiding generalization is more explanatory insofar as it makes explicit certain circumstances on which the explanandum *depended* for its occurrence. (In this sense, guidance is a specific form of the more general virtue of accuracy.) For instance, insofar as (2*) is more explanatory than (2), it is because (2*) conveys more information about the factors on which the occurrence of heart disease depends. Thus an

---

[15] These remarks allow us to address an important concern raised by a reviewer. The concern is that we are using breadth in two quite different ways. To see why consider the following difference between the two examples used to illustrate breadth in section II. in the case of (1*) vs. (1), the generalization is broader because it accounts for a larger range of types of *explananda*. That is, it can be used to explain any episode that qualifies as risk-taking behavior, not just bungee-jumping. By contrast, (3*) is broader than (3) not because it accounts for more types of *explananda* (both generalizations apply to the same type of *explanandum*, namely pulmonary embolism), but because it accounts for this *explanandum* type in a wider range of circumstances. In light of this, one may suspect that two quite distinct phenomena are conflated under the heading of `breadth'.

We agree that there is an interesting difference between two senses of `broad generalization' here – one that at the end of the day may need to be incorporated into a full interventionist account of explanatory virtues. But we think that there are enough conceptual and theoretical similarities between these two forms of breadth to warrant common treatment for our current purposes. In particular, both kinds of breadth can be seen as contributing to the quality of an explanation in the same way, viz. by revealing that the explanandum was bound to happen, whatever the actual background circumstances turned out to be. Thus Jane's embolism is best explained by (3*) rather than (3) insofar as (3) picks out a cause of her embolism in light of which this outcome was to be expected, independently of how other aspects of the world turned out to be. And the exact same thing can be said when comparing an explanation of Mary's behavior in terms of (1*) and an explanation in terms of (1): the former makes it clear that the explanandum was more or less bound to happen, independently of what the actual background circumstances were. This is not to deny that there are interesting differences between these two cases. In particular, in the first case, the desired effect is achieved by selecting one causal factor rather than another (viz. thrombosis rather than pregnancy) as the explanans; in the second case the desired effect is achieved by describing the explanandum as an instance of risk-taking behavior rather than an instance of bungee-jumping. Nevertheless, there is a substantial and theoretically interesting sense in which both explanations display the same virtue.

[16] Strevens (2008) is a prominent advocate of this view of explanation.

emphasis on guidance fits naturally with a view of explanation on which the ideal to which we aspire in our explanatory practices is to convey as much *accurate* information as possible about the relationships of dependence linking the *explanandum* to other factors – which, it turns out, is precisely the view advocated by Hitchcock and Woodward (2003).

Breadth and guidance, then, fit naturally with different views of the ideal of explanation. It is therefore not surprising that some accounts of explanatory virtues may end up emphasizing one at the expense of the other (if only implicitly), depending on which of these ideals they favor. For our part, we think that breadth and guidance are both important, and that a proper theory of explanatory virtues should recognize the ideals with which each virtue is associated. That is, we are inclined to think that a proper theory of explanatory virtues should be inclusive and recognize both abstraction and accuracy as important ideals that we aspire to in our explanatory practices. Defending this view is beyond the scope of this paper. Here we simply want to point out that recognizing the breadth/guidance distinction can help further and sharpen inquiry into the nature of explanation.

IV. Breadth, Guidance, and Exportability

Another advantage of drawing a sharp distinction between breadth and guidance as distinct forms of stability is that doing so helps us paint a more perspicuous and accurate picture of the cognitive and practical value of explanations that invoke stable generalizations.

To articulate this point it will be useful to briefly discuss the *function* of causal explanations in our cognitive and practical lives. Within an interventionist framework, it is natural to think of the function of causal explanations along the lines of the "explanation for export" theory developed by cognitive psychologists Lombrozo and Carey (Lombrozo and Carey 2006, Lombrozo 2010, Lombrozo 2011). According to them, one of the central functions of causal explanations is to identify dependence patterns that are `exportable' insofar as they can be applied successfully in contexts besides the one that is the focus of the explanation, and which can therefore be recruited for future tasks of prediction and control. This view fits very naturally with the interventionist approach, which takes causal claims to encode information about robust patterns of association between variables that can be exploited for the purposes of manipulation and control. The exportability theory adds to this idea the further hypothesis that those causal claims that we regard as especially explanatory are those that are particularly useful for such purposes, as well as for prediction.[17]

From the perspective of the exportability theory, it is unsurprising that we give special value to stable generalizations, as stability makes for greater exportability. In this vein, Woodward writes that

> Stability… has to do with the extent to which a causal relationship is exportable from one set of circumstances or another… [M]ore stable relationships are more generalizable and provide more information relevant to manipulation and control. (2015, 22)

We think that while this is correct, distinguishing breadth and guidance also provides a more complete and perspicuous account of the connection between stability and exportability. While broad and guiding generalizations can both be said to provide information especially relevant to manipulation and control (as well as prediction), they do so in markedly different ways, each of them corresponding to a different sense of `exportability' or `generalizability' and tied to a specific pragmatic and cognitive interest.

On the one hand, *broad* generalizations provide information that is especially relevant to our practical aims insofar as such generalizations can be reliably applied in a wide variety of situations that we may encounter. For instance, `Thrombosis causes pulmonary embolisms' can reliably be applied for the purposes of predicting and preventing pulmonary embolisms whenever we encounter a patient with thrombosis, regardless of their idiosyncratic characteristics (age, gender, etc.), so that we are thereby spared the trouble of having to identify those characteristics. Correspondingly, what drives our interest for broad explanatory

---

[17] See Lombrozo (2011) for a review of the empirical evidence in favor of the exportability theory of explanation. The exportability theory can also be recruited to explain the function of judgments of singular causation (Lombrozo 2010, Hitchcock 2012). See also Phillips and Shaw (2015) and Murray and Lombrozo (2016), who show that people are less inclined to regard an agent as the cause of a bad outcome when a third-party intentionally controlled the agent. As these authors point out, this is plausibly due to the fact that the dependence of the outcome on the agent is very sensitive to the third-party's intentions and in that respect fairly unstable.

generalizations is a concern for picking out causal information that can be repeatedly and efficiently used in various cognitive and practical problem-solving tasks.

In contrast, a generalization that achieves stability in the form of *guidance* may not be repeatedly applicable in this way: an explanatory generalization may be applicable only in a very small set of situations, and nonetheless have a high degree of guidance. Instead, a very guiding generalization is especially `exportable' because it conveys a large amount of information about the conditions under which it may reliably be applied for the purposes of prediction, manipulation and control, and thereby allows us to effectively identify those circumstances. In that respect, guidance is valuable because it answers a different kind of pragmatic and cognitive interest: a concern for identifying the conditions under which a causal relationship may safely be relied upon.

In this way, breadth and guidance can be seen as two complementary but distinct responses to the problem of determining whether a causal relationship present in a situation can be generalized to another situation. Broad generalizations answer the problem by picking out patterns of dependence that hold in a wide variety of situations, while guiding generalizations answer it by explicitly conveying information that we can rely upon for the purposes of determining whether a causal relationship present in a certain case can be applied in another case.

In light of these remarks, it also appears that the value of breadth and guidance may depend on the context, and in particular on which cognitive and practical goals are driving the inquiry. Thus imagine a medical scientist concerned with identifying the causes of pulmonary embolisms in order to reduce the latter's incidence in the general population. In this specific practical and cognitive context, it is clear that what matters first and foremost is identifying broad generalizations involving pulmonary embolisms, so that (for example) a generalization such as (3*) is particularly explanatorily valuable: by picking a causal relationship that holds in many circumstances, this generalization also allows us to identify particularly effective policies for reducing the incidence of pulmonary embolisms in the general population. By contrast, consider a context in which the main cognitive goal is to design medical tests that allow us to precisely and reliably determine whether a pregnant woman is at a risk of having pulmonary embolisms. Here it is guidance that matters: ideally we would like a generalization that improves on (3) not by picking out a broader relationship involving embolisms, but by laying out explicitly the circumstances under which pregnancy causes embolisms.

Note also that breadth and guidance may sometimes trade off, and that which tradeoff is acceptable may once again depend on which cognitive and practical goals are driving the inquiry. To illustrate, suppose that whether a certain treatment *T* is effective against a disease *D* depends on the patient's genetic make-up. For individuals who have variant *A* of a certain gene, *T* is usually effective against *D*. For patients who have variant *B* of the gene, things are more complicated: while some of these individuals recover when they take the treatment, for some of them the treatment actually makes the disease worse. (Let's imagine that whether a patient with variant *B* recovers by taking *T* depends on certain extremely fine-grained aspects of the treatment-gene interaction, so that there is no cognitively tractable way of describing the conditions under which administering *T* to a patient with variant *B* will cure rather than exacerbate the disease.) Now consider these two descriptions of the causal relationship between *T* and recovery:

(4) *T* causes recovery

(4*) Among patients with gene variant *A*, *T* causes recovery

On the one hand, considerations of guidance may push us towards preferring (4*). After all, administering *T* to a patient with gene *B* may have negative effects, and the virtue of (4*) is that it is explicitly restricted to those cases in which *T* can reliably and safely be administered. But this comes with a cost in breadth: (4*) does not apply to those cases in which a patient with variant *B* recovers from the disease through *T*. On the other hand, (4) does capture these cases and thus scores high on breadth, but this comes with a cost in guidance: using (4) as a guide to treatment administration may in certain circumstances lead to bad consequences. Thus whether (4) or (4*) is a more useful generalization will once again depend on which cognitive and practical interests (maximizing potential benefits of the treatment vs. minimizing potential harms) prevail in the context.

To summarize, drawing attention to the distinction between breadth and guidance draws attention to the fact that there are two different cognitive and practical interests that drive our preference for more stable generalizations, and that which form of stability is privileged may depend on which of these interests are driving inquiry. Conversely, the fact that these two kinds of stability derive their value from different pragmatic and cognitive concerns further supports the claim that breadth and guidance are importantly distinct explanatory virtues. (We hasten to add that these concerns, while distinct, are not incompatible: in many contexts our explanatory concerns are driven by an interest in both breadth and guidance.)

V. Weighting the Relative Importance of Breadth and Guidance

We now turn to the final point that we want to make in this paper. Although the distinction between breadth and guidance hasn't been explicitly drawn in the literature, we think that most discussions of the explanatory virtue of stability (including Woodward's seminal 2006 and 2010 contributions) tend to emphasize the explanatory value of what we call `breadth'. We think that in the process, the explanatory value of guidance tends to be neglected, or at least underestimated. We want to point out that guidance also has an important theoretical role to play (including a role sometimes ascribed to what we call breadth), and that recognizing this role allows us to paint a more accurate and balanced picture of the importance of stability considerations in explanation.

To illustrate, let's return to the example of the defective generalization

(1) Gene *g* causes bungee-jumping

which Woodward uses in a number of places (2010, 2015) to illustrate the explanatory importance of what we call breadth. We agree that one way to `fix' the defective generalization (1) is indeed to replace it with a related generalization holding in a wider range of circumstances, such as

(1*) Gene *g* causes risk-taking behavior

Yet it should appear clear by now that this is not the only way to improve on (1). Another option is to replace (1) with a more guiding generalization of the form

(1**) In circumstances *B*, gene *g* causes bungee-jumping

where *B* describes the conditions required for the relationship to hold, such as e.g. the availability of bungee-jumping technology. As suggested above, which of these explanatory strategies is more appropriate may depend on the context, and specifically on the particular cognitive and practical goals and values driving the explanatory inquiry. To illustrate it is best to turn to a realistic example, which we borrow from Potochnik (2015). It has been shown that a certain variant on the BDNF genetic locus is associated with a propensity to smoke (Amos et al. 2010). Presumably the generalization

(5) The BDNF genetic locus causes smoking

is highly unstable, insofar as the relationship holds only in very specific social, cultural, and psychological background circumstances. In a context where the goal is to reliably predict the consequences of having the BDNF variant in question in the general population, it may be better to appeal to a broader generalization than (5), e.g.

(5*) The BDNF genetic locus causes stress

But in a context where we are specifically interested in designing effective interventions on smoking behavior, turning to a generalization like (5) may not be the most useful strategy: instead, it may be more appropriate to fix the defectiveness of (5) by turning to a more guiding generalization which makes explicit the circumstances required for the causal relationship between BDNF and smoking to hold. (One can read Amos et al. as relying on such a generalization, since they allude to those circumstances at the outset of the paper.[18]) This example drives home the point that guidance can play certain theoretical roles that should not

[18] `Nicotine dependence results from an interplay of neurobiological, environmental and genetic factors. Patterns of smoking initiation reflect individual differences in sensitivity to nicotine, the availability of tobacco and social norms.' (Amos et al. 2010: 366).

be attributed to breadth: in certain contexts where specific theoretical and practical aims are operative, it is more appropriate to solve the instability of a generalization by increasing guidance than by increasing breadth.

There is another point worth making concerning the relative importance of breadth and guidance. One of the main tasks for which stability has been recruited in the literature is to account for the explanatory value of information about *mechanisms*. We think this is another issue where guidance has an important theoretical role to play, in a way that qualifies the importance of breadth. As has been emphasized by the `new mechanists' (Machamer et al. 2000), mechanisms play a central role in scientific and ordinary explanation. In particular, in many areas of science, uncovering the mechanism by which a cause produces a certain effect is regarded as perhaps the most important step in theory-construction and explanation-building. A good theory of explanation should explain this feature of our explanatory practices. As Woodward (2011: 423) has pointed out, the interventionist account of explanation can nicely explain why mechanistic information is explanatory by appealing to stability (in the form of breadth). In general, discovering the mechanism underlying a causal relationship from X to Y provides information about the intermediate steps in the causal chain leading from X to Y. As we noted in section I, when we have a causal chain of the form $X \rightarrow Z \rightarrow Y$, the relationship $Z \rightarrow Y$ is in general broader than the relationship $X \rightarrow Y$, insofar as it holds even in circumstances where the causal relationship from X to Z is not operative. Thus, by discovering the mechanism linking X to Y, we can identify broader causal relationships involving Y as an effect. Woodward concludes that `a concern with identifying stable [i.e. broad] relationships thus leads directly to an interest in identifying intermediate links in underlying mechanisms' (2011: 423). By way of example, suppose that all we knew about pulmonary embolisms is that they are sometimes caused by pregnancy. Discovering the mechanism by which this happens *ipso facto* reveals a broader generalization involving pulmonary embolisms (namely that they are caused by thrombosis, including in circumstances that do not involve pregnancy). Insofar as we have good pragmatic reasons to look for broad generalizations (see the previous section), it is no wonder that we regard this sort of information about mechanisms as explanatory.

We agree with Woodward that concerns for breadth play an important role in our explanatory interest in mechanisms. Nevertheless, we think it cannot be the whole story: an adequate interventionist account of the explanatory value of mechanisms must additionally recognize the important role of guidance. The reason is that there are many cases in which mechanistic information is explanatorily valuable, but fails to reveal interesting broader generalizations about the *explanandum*. Instead, the explanatory value of mechanistic information in those cases can plausibly be traced back to an interest in guiding generalizations. Consider for instance the mechanism by which neurotransmitters are released from neurons to other neurons via synapses, a paradigmatic example of a mechanism which has received a lot of attention in the `new mechanistic' literature.[19] Simplifying for the sake of exposition, this mechanism involves the release of neurotransmitters from the synaptic vesicle through a voltage-gated channel, after which the neurotransmitters travel through the `synaptic cleft' (the gap between the neurons) and bind to receptors situated on the membrane of the neuron on the other side of the synapse. While information about this mechanism is obviously valuable, we do not think that it can be accounted for in terms of breadth. In particular, it seems strained to claim that the individual causal links in this chain are broader than the overall causal chain: after all, there are few if any realistic circumstances in which neurotransmitters get to travel through the synaptic cleft without having been previously released from the synaptic vesicle, and few if any realistic circumstances in which neurotransmitters get bound to receptors without having previously traveled through the synaptic cleft. At any rate, it seems implausible to us to attribute the explanatory value of this mechanistic information to the fact that it reveals that the postsynaptic neuron would still receive neurotransmitters in such unlikely circumstances.

Instead, within an interventionist framework, it is far more natural to trace the explanatory value of this mechanism back to a concern for guidance. By learning the mechanism, we also learn more about the conditions that must be operative for the generalization `neurons release transmitters to other neurons via synapses' to hold. Here we are taking our cue from Strevens (2007), who argues that information about mechanisms is explanatorily useful in part because it

[19] See e.*g*. Craver 2006.

allows us to identify conditions under which a generalization may break down – and (conversely), to zoom in on those circumstances that are required for a causal generalization to hold. For instance, information about the relevant mechanism reveals that this generalization will not hold if the voltage-gated channel is malfunctioning, or if other molecules are already bound to the receptors on the postsynaptic membrane (which may be the case if the subject has ingested certain drugs). More generally, when we learn the mechanism underlying a causal relationship between two variables *X* and *Y*, we thereby acquire information about which circumstances may disrupt this causal relationship by preventing one of the intermediate causal links in the mechanism from operating, and hence we learn more about the conditions that must be present for the causal relationship to hold. The result is that the bare explanatory generalization `*X* causes *Y*' is replaced by a more guiding generalization of the form `In circumstances *B*, *X* causes *Y*'. Since we have a strong pragmatic interest in learning such generalizations, it is no wonder that we find mechanistic information explanatory even when such information fails to reveal broader generalizations involving the explanandum, as in the example we just considered.

VI. Conclusion

In sum, we have argued for an important distinction between two ways in which an explanatory generalization can be stable: by offering breadth or by offering guidance. Moreover, we have argued that it's important to keep this distinction in mind, as breadth and guidance track different ideals of explanation, satisfy different cognitive and pragmatic ends, and play different theoretical roles in (for example) explaining the explanatory value of mechanisms.

Despite the striking differences between breadth and guidance, we speculate that there are a few reasons why they have not been adequately distinguished in the literature to date. Most obvious is the common label: breadth and guidance have been lumped together under "stability" or "insensitivity" or "robustness". However, this common label is more likely to be a symptom of deeper issues than a cause. First, breadth and guidance both offer ways to diagnose defective explanatory generalizations: those that either leave something out or include cases they should rightly exclude. Second, breadth and guidance can both be articulated as requirements on answers to *what-if* questions, albeit (as we have seen in section III) these requirements are importantly distinct. Third, breadth and guidance both offer ways in which an explanatory generalization can be exportable in the sense that it supports generalizations to other cases: by either applying widely, or by building in its conditions of application. These considerations help explain how breadth and guidance occupy related conceptual space, and why they might rightly be considered as instances of stability. But the lesson of our paper is that we would do well not to lose sight of their differences.

References

Amos, C. I., Spitz, M. R., and Cinciripini, P. (2010). Chipping Away at the Genetics of Smoking Behavior. *Nature Genetics* 42: 366–368.

Craver, C. (2006). When Mechanistic Models Explain. *Synthese* 153: 355-376.

Hitchcock, C. (2012). Portable Causal Dependence: A Tale of Consilience. *Philosophy of Science* 79: 942-951.

Hitchcock, C. and Woodward, J. (2003). Explanatory Generalizations, Part II: Plumbing Explanatory Depth. *Noûs* 37: 181-199.

Kendler, K. (2005). A Gene For…: The Nature of Gene Action in Psychiatric Disorders. *American Journal of Psychiatry* 162, 1243-1252.

Lombrozo, T. (2010). Causal-Explanatory Pluralism: How Intentions, Functions, and Mechanisms Influence Causal Ascriptions. *Cognitive Psychology* 61, 303-332.

Lombrozo, T. (2011). The Instrumental Value of Explanations. *Philosophy Compass* 6: 539-551.

Lombrozo, T. and Carey, S. (2006). Functional Explanation and the Function of Explanation. *Cognition* 99: 167-204.

Machamer, P., Darden, L. and Craver, C. (2000). Thinking about Mechanisms. *Philosophy of Science* 67: 1-25.

Malaterre, C. (2011). Making Sense of Downward Causation in Manipulationism: Illustrations from Cancer Research. *Studies in the History and Philosophy of the Life Sciences*, 33: 537-62.

Murray, D. and Lombrozo, T. (2016). Effects of manipulation on Attributions of Causation, Free Will and Moral Responsibility. *Cognitive Science*. Advance online publication. DOI: 10.1111/cogs.12338

Phillips, J. and Shaw, A. (2015). Manipulating Morality: Third-Party Intentions Alter Moral Judgments by Changing Causal Reasoning. *Cognitive Science* 39: 1320-1347.

Potochnik, A. (2015). Causal Patterns and Adequate Explanations. *Philosophical Studies* 172: 1163-1182.

Spirtes, P., & Scheines, R. (2004). Causal Inference of Ambiguous Manipulations. *Philosophy of Science* 71, 833–845.

Steel, D. (2006). Methodological Individualism, Explanation, and Invariance. *Philosophy of the Social Sciences* 36: 440-463.

Strevens, M. (2007). Why Represent Causal Relations? In Gopnik, A. & Schulz, L. (eds.), *Causal Learning: Psychology, Philosophy and Computation*. Oxford: Oxford University Press, pp. 345-60.

Strevens, M. (2008). *Depth*. Cambridge, MA: Harvard University Press.

Vasilyeva, N., Blanchard, T. and Lombrozo, T. (2016). Stable Causal Relationships are Better Causal Relationships. Unpublished manuscript.

Weslake, B. (2010). Explanatory Depth. *Philosophy of Science* 77: 273-94.

Waters, K. (2007). Causes that Make a Difference. *Journal of Philosophy* 104: 551-579.

Woodward, J. (2006). Sensitive and Insensitive Causation. *Philosophical Review* 115: 1-50.

Woodward, J. (2010). Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation. *Biology & Philosophy* 25: 287-318.

Woodward, J. (2011). Mechanisms Revisited. *Synthese* 183: 409-427.

Woodward, J. (2015). The Problem of Variable Choice. *Synthese*, forthcoming. Online version available at http://link.springer.com/article/10.1007%2Fs11229-015-0810-5.

Ylikoski, P. and Kuorikoski, J. (2010). Dissecting Explanatory Power. *Philosophical Studies* 148: 201-219.