Penultimate draft. Please cite published version: https://doi.org/10.1016/j.shpsa.2023.08.008

Explaining Individual Differences

Zina B. Ward
Florida State University

Published 2023 in *Studies in History and Philosophy of Science*

Abstract: Most psychological research aims to uncover generalizations about the mind that hold across subjects. Philosophical discussions of scientific explanation have focused on such generalizations, but in doing so, have often overlooked an important phenomenon: variation. Variation is ubiquitous in psychology and many other domains, and an important target of explanation in its own right. Here I characterize explananda that concern individual differences and formulate an account of what it takes to explain them. I argue that the notion of actual difference making, the only causal concept in the literature that explicitly addresses variation, cannot be used to ground such an account. Instead, I propose a view on which explaining individual differences involves identifying causes that could be intervened on to reduce the variability in the population. This account provides criteria of success for explaining variation and deepens our understanding of causal explanation.

1. Variation, Explanation, and The Dress

In 2015, a photograph of a striped dress went viral.[1] Some people were certain that the dress had a blue body with black lace stripes, while others were just as sure it had a white body with gold stripes. The public was fascinated that people could have such different perceptions of the same object (Holderness 2015). Remarkably, researchers working on color perception were "almost equally surprised" by the phenomenon, which sparked a lively discussion on the Color and Vision Network mailing list (Olkkonen and Ekroll 2016, 176). The same year the photograph appeared on Tumblr, there was a trio of commentaries on it in *Current Biology*, and in 2017 the *Journal of Vision* dedicated a special issue to it. To date, the photograph has been the subject of

[1] The photograph can be seen at http://en.wikipedia.org/wiki/The_dress

dozens of research papers (Martín-Moro et al. 2018). Scientists and laypeople alike want to know: why do people see such different things when they look at #TheDress?

The explanandum picked out by this question is different from many other scientific explananda in that it concerns variation rather than regularity. Psychologists often want to know things like what role working memory plays in language, why we are susceptible to the McGurk effect, and how people learn new perceptual categories. These explananda are about regularities across minds rather than differences between them. In this paper, I'll spell out what is different about explananda that concern variation and provide a philosophical account of what it takes to explain them. In doing so, I'll draw examples from research on #TheDress.

Individual differences are of interest not only in psychology, but across the sciences, from medicine to education to economics. Especially in applied disciplines, variation can have practical consequences. It is important to know, for example, not just whether certain biomedical treatments or pedagogical interventions are effective on average, but also whether they have different effects on different individuals. Ignoring such variation can lead to one-size-fits-all approaches that cause harm to patients and students who fall outside the norm. Variation between other units of analysis matters too: we also want to be able to explain why different solar systems, muscle cells, bees, labor markets, and hurricanes are different from one another. Establishing regularities that hold across all hurricanes, but not understanding why different hurricanes follow different paths and development trajectories, would leave us with an impoverished meteorology.

Despite its importance, variation remains philosophically undertheorized. It has received the most attention from philosophers of biology, in part because variation is a precondition for evolution by natural selection (Lewontin 1970). But variation need not play a role in an evolutionary process to be of scientific interest. By taking psychology as my entry point, I bracket the distinctive concerns of evolutionary biology, constructing an account of the explanation of variation that has wide applicability. This account deepens our general understanding of scientific explanation and sheds light on an underexplored dimension of science.

My plan is as follows. In Section 2, I delineate the subclass of explananda that I call "variation explananda." To formulate an account of what it takes to explain a variation explanandum, I adopt an interventionist framework, introduced in Section 3. I then articulate two interventionist theories of the explanation of individual differences, rejecting the first and endorsing the second. Section 4 argues against an account that appeals to Waters' (2007) notion of actual difference making. On my alternative view, presented in Section 5, explaining a variation explanandum involves exhibiting causes of variation, which are variables that could be intervened on to reduce population variance. My account provides criteria of success for the explanation of individual differences and reveals a new way in which the causal background must be kept in view when making claims about populations.

2. Variation Explananda

A primary aim of psychology is to understand how the mind works. Understanding the mind requires the formulation of generalizations about its operation that hold across subjects. Much psychological research tries to identify regularities of this kind. There is another dimension of psychological understanding, however, that is equally important: we want to know what *differences* exist between the minds of individuals and explain why they occur. Interest in psychological variation has long animated fields such as psychopathology, cultural psychology, and differential psychology, but variation is now gaining increasing attention outside disciplines explicitly concerned with difference. Psychologists interested in color, for example, want to understand both the general mechanisms underlying color perception and why the photograph of #TheDress produces radically different perceptions in different people.[2]

There are different questions scientists can ask about individual differences. One can ask why variation is structured in a certain way in the population, as in, "Why are individuals' scores on this task bimodally distributed?" (Ward 2022). Or one can ask why there isn't *more* variation than there is, as in, "Why does the dress elicit *just two* different responses?" The most basic question to ask about variation, though, is why it exists at all: why there is (any) variability in the variable of interest? Questions that take this form include, "Why are some people faster than others at the Stroop task?" and "Why is language processing lateralized to the right hemisphere in some people and the left in others?" Such questions are the starting point for individual differences research and will be my focus here.

Questions about the existence of variation pick out what I will call "variation explananda." Let variable Y represent a trait or behavior that ranges over individuals in a population p. A variation explanandum asserts that Y takes different values for (at least some of the) different individuals in p. In other words, it asserts that there are two or more values of Y represented in the population, rather than one.[3] In psychology, the individuals in question are typically individual people, but other domains focus on other units of analysis. In astronomy, for instance, a variation explanandum might assert that there is variability in the densities of individual planets. I will

---

[2] I'll talk loosely about people "seeing," "perceiving," or "interpreting" the dress differently. I intend to remain neutral about whether the phenomenon is fully perceptual.

[3] One might wonder why variation explananda are here characterized by appeal to the first-order variable Y rather than a binary variable (say, Y') representing the presence or absence of variation in Y. The reasons for this choice will become apparent in the interventionist analysis to follow. Introducing variables that depend non-causally on other variables in one's model greatly complicates analysis of the causal relations, as discussed in the literature on mental causation (e.g., Woodward 2015). Moreover, given interventionist assumptions, use of a binary variable would imply that only variables which can be intervened on to completely eliminate variation help to explain it (see Section 5). This is far too demanding a requirement.

motivate my account using psychology in what follows, but its scope is not limited to that field (see Section 6).

The relevant population *p* in a variation explanandum is occasionally obvious, but more often must be inferred. When we ask why there is variation in how people interpret the colors of the dress, *p* is arguably comprised of adults who encounter the dress photograph online. Even when *p* is left implicit, it is an important component of a variation explanandum. There are no individual differences without a population of individuals. In what follows I will focus on variation explananda that concern potentially recurring individual differences, not those that are about a specific instance of variation.[4]

Variation explananda incorporate an implied contrast: they state that there is variation *rather than uniformity* in *Y*. This requirement is needed because claims about the presence of variation can be associated with different contrasts, not all of which make variation the explanatory focus.[5] Consider the statement, "People's category representations differ from one another." In asking for an explanation of this, one might be asking, "Why are people's category representations, *but not their perceptual representations*, different?" or "Why do people have different *rather than the same* category representations?" Only the latter targets a variation explanandum. The former is about differences between category and perceptual representations rather than differences between individuals. The use of explanatory contrasts also allows us to distinguish questions about why there is variation at all from other questions one can ask about variation, such as those mentioned above. For instance, "Why does the photograph of #TheDress elicit two different

[4] Hence my account will target generic or type-level rather than singular or token-level explanations of variation. Future work might consider how it could be modified to accommodate the explanation of particular instances of variation.

[5] Understanding variation explananda as incorporating a contrast is in keeping with the common view that explanation is essentially contrastive (Lipton 1990, Woodward 2003). What is distinctive about variation explananda is the particular sort of contrast they involve.

responses?" can be understood as asking, "Why does the photograph elicit two different responses *rather than one*?" or "Why does the photograph elicit *just two* responses *rather than many?*" The latter question also concerns individual differences, but it does not target a variation explanandum.[6]

The formulation of variation explananda can be laborious. Consider the statement, "People see the dress as either blue and black or white and gold." Color-matching studies were required to establish that people really do have different perceptual experiences when they look at the dress, and that they are not simply using color names differently (Brainard and Hurlbert 2015, Chetverikov and Ivanchei 2016). The explanandum also remains controversial because some researchers believe the distribution of color perceptions produced by the dress is continuous rather than discretely bimodal (Gegenfurtner et al. 2015, Hugrass et al. 2017, Aston and Hurlbert 2017). Others think there are not two perceptions but three, since a small number of people report that the dress is blue and brown (Lafer-Sousa and Conway 2017, Jonauskaite et al. 2018).

To recap: variation explananda are a subclass of explananda that state that there is variation rather than uniformity in variable Y in population p. In addition to studying regularities across minds, psychologists and other scientists are interested in explaining variation explananda. My aim in what follows is to propose a philosophical account of what this requires. In doing so, I'll use the dress as an extended case study. Let Y be a binary variable representing the perceived colors of the dress: Y=0 if an individual sees the dress as blue and black, and Y=1 if they see it as white and gold. Our focal variation explanandum asserts that Y takes different values (rather than a single value) in the population p of individuals who come across the dress photograph online.

[6] Future work might extend the account proposed below to these more complex explananda. For now I define my target narrowly, tackling variation explananda because they often constitute the starting point for individual differences research.

Although individual differences in perceptions of the dress are especially striking, they have the same features as more mundane patterns of variation of interest to psychologists: they are measurable, predictable differences between people with high intra-individual reliability (Drissi-Daoudi et al. 2020). While I have taken variation in perceptions of the dress to be discrete, a great deal of variation in psychology and elsewhere is arguably continuous. However, this difference does not matter to the accounts of explanation that will be considered below. One could just as well illustrate the account by appeal to individual differences in speech planning (Swets 2015), risky decision making (Glöckner and Pachur 2012), working memory (Unsworth et al. 2004), visual imagery (Reeder et al. 2017), or any other variable psychological or non-psychological phenomenon. Moreover, as we will see, potential explanations of variation in interpretations of the dress are not at all atypical, appealing to long-established psychological principles (e.g., de Lange et al. 2018). We can therefore safely generalize from an in-depth examination of the investigation and explanation of variation in #TheDress to variation more broadly (see Section 6).

3. Interventionism

In my view, the best current theory of causal explanation is interventionism (Woodward 2003, Woodward and Hitchcock 2003). Since interventionism has been defended elsewhere, I will not discuss its merits here. I will instead ask what the interventionist ought to say about what it takes to explain a variation explanandum, focusing on Woodward's (2003) account as the most thoroughly developed version of the view.

Interventionism conceives of causation as a relation between variables that take different values. Roughly speaking, X is a cause of Y if and only if Y can be changed by intervening on X while holding other factors fixed. The notion of an intervention here is critical and somewhat technical. Intervening is like conducting an ideal experiment to determine whether *X* causes *Y*: one sets the value of *X*, severing its connections with factors that usually influence it, in order to see whether the value of *Y* changes as a result. More precisely, an intervention on *X* with respect to *Y* is the taking of a specific value by an intervention variable *I* that acts as a switch for *X*. That is, when *I* takes on (a) particular value(s), the value of *X* is a function of *I* alone. *X* ceases to depend on any other variables. Moreover, *I* only influences *Y* through *X*; it does not exert an independent influence on *Y*. Interventions need not be practically feasible or even physically possible. They need only be logically or conceptually possible and well-defined (Woodward 2003, 127-132).

To generate causal claims from facts about interventions, one needs to decide where to hold fixed the aforementioned "other factors." Specifically, when considering a potential intervention on *X* with respect to *Y*, values must be assigned to the variables not on a causal path from *X* to *Y*. (If off-path variables weren't held fixed, then changes to *Y* could be the result of changes to those variables rather than the intervention on *X*.) Woodward argues that the proper handling of off-path variables depends on whether we are interested in type or token causation. For type-causal claims, off-path variables can be fixed at any possible value.[7] As long as there is an intervention on *X* with respect to *Y* that changes the value of *Y* under *some possible assignment* of values to off-path variables, then *X* is a type cause of *Y*. Second, Woodward suggests that we can capture most of our token causal judgments by setting off-path variables at their actual values.[8] As long as there is an intervention on *X* with respect to *Y* that changes the value of *Y* when off-

[7] Here and elsewhere, when I discuss causes or causal claims, I mean to invoke Woodward's notion of a contributing cause. Other causal concepts (e.g., total cause, direct cause) require different treatment.
[8] Woodward concedes that this view doesn't generate the correct judgments in token cases of symmetric overdetermination. To accommodate such cases, he settles on a view that appeals to the notion of the "redundancy range" of a variable (Woodward 2003, 74-86; Hitchcock 2001). Since his initial account is far simpler and captures almost all of our judgments, and since cases of symmetric overdetermination are arguably rare, I will stick with the initial proposal that off-path variables be held at their actual values.

path variables are set to their *actual values*, then *X* is a token cause of *Y*. Part of my aim in what follows will be to show that neither of these proposals is adequate for causal claims about variation.

Colloquially, the interventionist view of causation is often presented in the language of "wiggling": *X* is a cause of *Y* iff there's some way of wiggling *X* that changes *Y*, with off-path variables held at the appropriate values. Rules for the assignment of values to off-path variables delineate the background conditions against which wiggling *X* must change *Y* for *X* to count as a cause of the relevant kind. (In this respect, limits on the assignment of values to off-path variables within an interventionist framework serve some of the same functions as Mackie's [1974] notion of a "causal field.") Explanations involve the provision of causal information, understood in interventionist terms. This means they show how the explanandum depends counterfactually on the explanans, where the counterfactuals in question describe the outcomes of interventions. Such counterfactuals are sometimes called "interventionist counterfactuals." Explanations can be used to answer what Woodward dubs "what-if-things-had-been-different" questions because they show us what difference it would have made to the explanandum "if the factors cited in the explanans had been different in various possible ways" (Woodward 2003, 11).

The rest of the paper will develop an interventionist analysis of the causal explanation of individual differences. Given the interventionist's commitments, what should she say about what it takes to explain a variation explanandum? The next section will consider and reject a proposal that appeals to the notion of actual difference making. Informed by the shortcomings of this account, I'll then lay out my own view.

4. The ACTUAL Account

One potential interventionist account of how to explain a variation explanandum draws on a concept developed by Waters (2007). Waters claims that there is an ontologically special subset of causes called actual difference makers (ADMs) that are responsible for "actual differences" in a population. One might argue, then, that the variables that are explanatorily relevant to variation are ADMs. In this section, I'll argue against this Waters-inspired account.

4.1 Introducing ACTUAL

According to Waters (2007), *X* is *an actual difference maker* with respect to *Y* in population *p* if and only if:

(i) *X* causes *Y*.
(ii) The value of *Y* actually varies among individuals in *p*.
(iii) The relationship *X* causes *Y* is invariant over at least parts of the space(s) of values that other variables actually take in *p*.
(iv) Actual variation in the value of *X* partially accounts for the actual variation of *Y* values in population *p* (via the relationship *X* causes *Y*).

Loosely speaking, a variable *X* is an ADM for *Y* in *p* when *X* and *Y* both vary and *X* causally accounts for some of the variation in *Y.* Waters endorses interventionism and suggests that the causal concepts in his definition be explicated using Woodward's (2003) framework. Actual difference making is meant to capture the intuitive distinction between the causes of a trait and the causes of "actual differences" in the trait. For example, consider the actual differences in sun-related skin damage in the human population. The causes of these differences include differences in lifetime sun exposure, skin melanin concentration, and sunscreen usage. Such factors are all ADMs because they vary and are partly responsible for existing variation in skin damage among humans. One factor that is a cause of skin damage but not an ADM is the amount of ultraviolet radiation emitted by the sun. The sun's total UV radiation does not account for any of the variation among humans since it is the same for everyone.

Waters' paper is controversial because of how he uses the concept of actual difference making to reject causal parity theses, which claim that different causes in biology (such as genes and physiological background conditions) are ontologically on a par. Waters argues that biologists are most interested in ADMs and that genes are the most causally specific ADMs in biology. His aim is to defend gene-centric biology against the critiques of developmental systems theorists and others. Waters' argument has been criticized for purportedly implying that biologists don't care about potential difference makers (Stegmann 2012, Currie 2018, Baxter 2019), for ignoring ADMs that aren't genes (Griffiths and Stotz 2013), for characterizing the distinction between ADMs and all other causes as ontological (Northcott 2009), and for oversimplifying the views of those who endorse causal parity theses (Griffiths and Stotz 2013). These critiques have challenged Waters' conclusions about biology but not the definition of an ADM itself.

Actual difference making may seem to be just the tool we need to formulate an interventionist account of the explanation of variation explananda. The causes that Waters singles out as ADMs are those variables that account for individual differences in a population. One might then formulate an account on which ADMs are what explain variation:

ACTUAL The variables that explain a variation explanandum are actual difference maker(s) X with respect to Y in population p.

ACTUAL is inspired by Waters but not intended to capture his view.[9] The account has prima facie plausibility. We do frequently explain variation by tracing differences in behavior back to differences in some other property of individuals. It is intuitive to think that difference must be explained by appeal to difference. This idea is often implicit in the interpretation of statistical

[9] I am not attributing ACTUAL to Waters (2007). Waters' focus is on making sense of experimental practice in biology. He is directly concerned with causal selection: understanding why biologists single out particular causes for special attention. ACTUAL is one possible *extension* of these views in the domain of explanation. On my reading, Waters too could reject ACTUAL by granting that merely potential difference makers are still explanatorily relevant to variation, just not as explanatorily privileged as ADMs.

analyses of variance (ANOVAs). When one variable "accounts for" variance in another, scientists often say that the former explains variation in the latter. But only ADMs can statistically account for variance in an ANOVA. Hence statistical practices seem to sanction ACTUAL's identification of ADMs with explanatorily relevant variables (cf. Lewontin 1974).

Further support for ACTUAL comes from the observation that explanations appealing to ADMs abound in psychology. #TheDress provides a ready illustration. One factor that strongly influences the perceived colors of the dress is how subjects interpret the illumination in the photograph (Chetverikov and Ivanchei 2016, Wallisch 2017, Witzel et al. 2017b, Toscani et al. 2017).[10] People who interpret the image as showing the dress in shadow (or under blueish illumination) tend to see it as white and gold, while those who take it to be directly lit (by yellowish illumination) typically see it as blue and black. Wallisch (2017) explains why this is so: "Shadows overrepresent short wavelengths. In other words, shadows appear bluish. If someone assumes that the dress was in a shadow, color constancy mechanisms could be expected to discount the effect of the shadow, rendering the conscious appearance of the dress more yellowish" (5). Different subjects make different assumptions about illumination, leading color constancy mechanisms to "correct" the perception in different ways. Brainard and Hurlbert (2015) call this the "color constancy explanation" of individual differences in the perceived color of the dress. It is a relatively proximal and partial explanation, to be sure, since it does not explain why people make different assumptions about illumination, but it does provide some explanatory information.[11]

[10] Although I'll talk about people "interpreting" or "making assumptions about" illumination, the process is usually subpersonal. It would be more accurate to speak of the illumination assumptions made "by the visual system."
[11] Witzel et al. (2017b) and Metzger and Drewing (2019) argue that the dress provides an example of the cognitive penetration of perception: people's high-level beliefs about illumination shape their perception of color. This is another reason the dress is philosophically interesting, though I am skeptical of their claim.

ACTUAL seems to capture the explanatory import of the color constancy hypothesis for our focal variation explanandum. Individuals' assumptions about illumination (*X*) are an ADM with respect to the perceived colors of the dress (*Y*) in the population (*p*) of individuals who encounter the dress online. Let *X*=0 if a subject assumes that the dress is in shadow, and *X*=1 if they take it to be directly illuminated. Recall that binary variable *Y* represents how an individual perceives the colors of the dress. Different people see the colors of the dress differently (i.e., *Y* takes different values in *p*) satisfying Waters' condition (ii) on actual difference making. There is also strong experimental evidence that people's varying assumptions about the illumination in the photograph (*X)* causally influence their perceptions of the dress (*Y)*, satisfying (i) and (iii). Witzel et al. (2017b), for example, manipulate subjects' interpretations of the illumination by substituting realistic scenes which are either sunny or in shadow to serve as the background behind the dress. They find that when subjects are exposed to the image of the dress against a sunny background, they tend to see the dress as blue and black; when the background is in shadow, they tend to see it as white and gold. Together with survey results (Wallisch 2017), such experiments suggest that variation in assumptions about illumination partially account for variation in perceived dress color, satisfying Waters' condition (iv). Hence, people's illumination assumptions, which partly explain individual differences in perceived dress color, appear to fulfill Waters' criteria for being an ADM, lending credence to ACTUAL.

4.2 The Explanatory Role of Uniform Variables

ACTUAL is ultimately unsatisfactory, however, because it overlooks an important class of explanatory variables. To be an ADM, a variable has to take different values for different individuals in the population. But there are variables that take the same value for all individuals that can nonetheless help explain the presence of variation in the population. So, ACTUAL is wrong to equate the explanation of individual differences with the exhibiting of ADMs.

For a variable to be an ADM, it must take different values for different individuals in population *p*. This is because of Waters' condition (iv), which states that variation in *X* partially accounts for variation in *Y* in *p*. Waters fleshes out the condition as follows: "*X* partially accounts for the actual variation of *Y* values in population *p* (via the relationship *X* causes *Y*) if and only if conditions (i)–(iii) [in the definition of an ADM] are satisfied and an intervention on *X* with respect to *Y* that changed the *X* values in one or more individuals in *p* to the *X* value that one of the individuals had sans intervention would change *Y* values in *p*" (Waters 2007, 571). Let's call this kind of intervention a "swapping intervention," since it involves swapping one individual's *X* value for the *X* value of a different individual in *p*.[12] Condition (iv) is satisfied iff there is a swapping intervention on *X* that would change the value of *Y* for at least one individual in *p*. There can only be such an intervention if at least some of the individuals in *p* have different *X* values. Waters embraces this feature of his account, emphasizing that no variable that "exhibit[s] uniform values in the actual population" can be an ADM for that population (*ibid.,* 570).

ACTUAL claims that all of the variables that explain variation are ADMs, and hence take different values in the population *p*. The problem is that there seem to be variables that are uniform in *p* but that can still help explain variation. Consider the following example. Many researchers studying #TheDress have noticed that the pixel chromaticities in the photograph fall primarily along the blue-yellow axis (Brainard and Hurlbert 2015). This means that the chromaticities closely mirror variation in natural daylight, tracking the so-called "daylight locus." Research has shown that this chromaticity profile is essential to the photograph's capacity to generate different

[12] This term could be misleading, since it might suggest an *exchange* of two individuals' *X* values. On the reading of "swapping intervention" I intend, only one individual's *X* value must be modified.

perceptions in different people. Gegenfurtner et al. (2015) shift the RGB values in the photograph around the hue circle away from the daylight locus, making blueish pixels become pinkish and yellowish pixels become greenish. This adjustment makes the individual differences effectively disappear, with subjects reporting that the dress body is pink or red, and the lace green. It is not yet known why the photograph's blue-yellow chromaticity profile produces such stark individual differences. Building on work showing that people's color discrimination is poorest along the daylight locus (Pearce et al. 2014), Brainard and Hurlbert (2015) speculate that "there might be something especially ambiguous about images whose RGB values vary primarily in this way…[S]uch images might evoke a response shaped more than usual by implicit prior expectations about illumination spectra, which in turn might vary across individuals" (R553). Follow-up work investigating this hypothesis has produced mixed results (Wallisch 2017, Witzel et al. 2017a, Witzel et al. 2017b, Aston and Hurlbert 2017, Wallisch and Karlovich 2019).

The chromaticity profile of the original photograph is not an ADM for ordinary people's perceptions of the colors of the dress, but it does help explain variation in those perceptions. In our focal explanandum, Y represents the perceived colors of the dress. Let X be a continuous variable representing the distribution of chromaticities in the photograph of the dress. X is not an ADM for Y in the population p of individuals who come across the dress photograph online because X takes the same value for each individual. Even so, I contend that chromaticity profile does help explain why people see the dress differently. Citing the photograph's unusual distribution of chromaticities along the daylight locus allows us to partly understand why there is variation in how we interpret its colors. For one thing, chromaticity mediates the influence of other variables on people's perceptions of the dress. Subjects' prior expectations about illumination would be causally irrelevant to the perceived color of the dress, for instance, if the photograph's chromaticity profile were different. Gegenfurtner et al.'s (2015) experiment also shows that there is an intervention on chromaticity profile that eliminates variation in the perceived color of the dress. Of course, to fully understand why the variation exists, we would need to determine how exactly chromaticity profile interacts with other factors to produce different interpretations of the dress. But we shouldn't get hung up on what is required for a complete explanation, since science rarely traffics in complete explanations. The fact remains that the chromaticity profile of the dress does genuine explanatory work.[13]

A defender of ACTUAL might try to argue that chromaticity profile is an ADM after all. What separates the photograph of #TheDress from photographs of other dresses – what makes its color unusually ambiguous – is partly its chromaticity profile. So chromaticity profile is an ADM for perceived color when one considers the broader category of dress photographs. It is also an ADM for the population of participants in Gegenfurtner et al.'s (2015) study. For those participants, *X* took different values on different trials, and the varying chromaticity values made a difference to *Y,* their perceptions of the dress. While it is true that chromaticity profile is an ADM in such experimental contexts, it is not an ADM with respect to our focal variation explanandum. In the explanandum we are considering, population *p* is comprised of individuals who come across the original photograph of the dress in an ordinary way, not subjects in a chromaticity-altering study. Moreover, the population is a group of humans, not photographs. The question is why different *people* see this particular photograph differently, not why some photographs are subject

[13] One might object that chromaticity profile does explain how variation in perceptions of the dress is *possible*, but doesn't explain why variation is *present*. The proposed distinction between variables that make a phenomenon possible and variables that bring it into existence is not part of the standard interventionist framework. Wysocki [in preparation] proposes a modified interventionist account that applies to causal claims about possibility. His expanded framework also treats causes that make an event possible and causes that bring about the event on a par. On his account, then, if chromaticity profile causes variation to be *possible,* it also causes variation to be *present.*

to variable interpretation while others are not.[14] Chromaticity profile is hence not an ADM for the population that matters to our focal variation explanandum, even though it does help explain it.

This example shows that variables whose values are uniform in a population can still play an important role in explaining differences between individuals. A Waters-inspired account of how to explain variation is missing something important. Contra ACTUAL, ADMs are not the only variables that contribute to the explanation of variation explananda.

5. SHRINKing Individual Differences

Waters' account of actual difference making is the primary extension of interventionism in the literature that directly concerns variation. The failure of ACTUAL, then, forces us back to the drawing board. In this section, I'll present my own account of how to explain a variation explanandum and lay out its virtues.

5.1 Introducing the SHRINK Account

The idea behind SHRINK is that explaining a variation explanandum requires explaining why the population variance of *Y* is nonzero. This is accomplished by describing (an) intervention(s) on some variable *X* with respect to *Y* that would reduce (or "shrink") Var(*Y*). I call *X* a cause of variation in *Y*. The full account, which I'll explain below, is as follows:

[14] Recall the setup of our variation explanandum: *p* is the population of individuals who come across the original dress photograph online, *X* represents chromaticity profile, and *Y* represents how a subject sees the colors of the dress. Here *X* is not an ADM for *Y* because it is uniform in *p*. The objector is entertaining a different setup in which *p* is the population of (dress) photographs, *X* represents chromaticity profile, and *Y* represents whether a particular image is subject to variable interpretation or not (i.e., whether there are individual differences in how people perceive its colors). In this latter scenario, *X* is indeed an ADM with respect to *Y* in *p*. But that is beside the point: my claim is not that chromaticity is *never* an ADM, but rather that it is not an ADM with respect to the variation explanandum I characterized in Section 2.

SHRINK A variable *X* is a cause of variation in *Y* in population *p* iff there is a possible intervention on *X* with respect to *Y* that would reduce Var(*Y*) in *p* and in which the actual distribution of values of variables not on a path from *X* to *Y* in *p* is held fixed.

To explain a variation explanandum, one must exhibit a true causal generalization *G* relating change(s) in a cause of variation *X* with change(s) in *Y* that are associated with a reduction in Var(*Y*).[15]

SHRINK captures the explanatory import of uniform variables, avoiding the problem that plagued ACTUAL. Causes of variation, as defined by SHRINK, can be the same or different across individuals in *p*. For example, the chromaticity profile of the dress photograph qualifies as a cause of variation under SHRINK because there is an intervention on chromaticity profile with respect to perception of the dress that reduces the variance in the latter, when the distribution of off-path variables is held fixed (Gegenfurtner et al. 2015). Off-path variables in this case include things like viewing conditions (lighting, screen type, etc.) and people's prior expectations about illumination spectra. SHRINK therefore correctly implies that the photograph's chromaticity profile helps to explain individual differences in how the dress is perceived.

Several features of the view require further elaboration. First, SHRINK is an account of the explanation of variation that applies in deterministic contexts (though a probabilistic extension, invoking the probability distribution over Var(*Y*), would be easy to formulate). The intervention on *X* with respect to *Y* that it requires must be an intervention for at least one of the individuals in *p*. The account also assumes that there is causal homogeneity in the target population, meaning that the individuals in *p* conform to the same causal model. This may seem like an unwarranted assumption, but it is not as strong as it appears. First, it is a commonplace assumption in modeling and inference. Many statistical techniques require assumed homogeneity of some kind among the

[15] Population variance, written Var(*Y*) or σ^2, captures how much the values of *Y* differ from their mean $\bar{y}$ in a population with *n* members. It is standardly defined as follows:

$$\sigma^2 = {\sum(y_i - \bar{y})^2}/{n}$$

units. Second, individuals can differ from one another and still conform to the same causal model. They can have different values for any of the variables that range over individuals. Third, even if the individuals in *p* fall into (say) two qualitatively distinct groups, each associated with a different constellation of causal relationships, they can all be understood as conforming to the same model provided the model contains a variable representing group membership.

There is also a question about what justifies the asymmetry at the heart of SHRINK. Why do we only count as causes of variation those variables that can be intervened on to *reduce* population variance? The characterization of variation explananda offered in Section 2 provides the answer. Recall that a variation explanandum comes with a contrast: it states that there is variation *rather than uniformity* in *Y*. The interventionist counterfactuals exhibited to explain any explanandum must be responsive to its contrast, in the sense that they must describe counterfactual outcomes that approach the contrast. This means an explanation of a variation explanandum has to show how changes in the explanans variable *X* would have produced greater uniformity in *Y*. Variation explananda are not explained by interventionist counterfactuals in which Var(*Y*) increases because they do not contrast the presence of variation with uniformity. Note that this is an additional point of contrast with ACTUAL. Waters' condition (iv) on ADMs is satisfied even if all possible swapping interventions on *X* with respect to *Y* would *increase* the variance in *Y* (Waters 2007, 571-2). But it seems wrong to say that *X* partially accounts for (and thus helps explain) the variation in *Y* if *X*'s taking the values it does actually *minimizes* the variance in *Y*. Because it lacks SHRINK's asymmetry, ACTUAL does not capture the contrast with uniformity that characterizes a variation explanandum.

SHRINK's emphasis on the reduction of variance does not imply that the only evidence for causal claims about variation comes from experiments that reduce Var(*Y*). One must distinguish between the interventionist counterfactuals that make a causal claim true and the counterfactuals about experimental outcomes that support it. The latter may not mirror the former for practical reasons. For *X* to be a cause of variation under SHRINK, there must be a hypothetical intervention on *X* with respect to *Y* that would reduce Var(*Y*). However, under certain circumstances, such as when intervening on *X* to reduce Var(*Y*) is not feasible, an experimental manipulation of *X* that brings about an increase in Var(*Y*) may be taken as evidence that *X* is a cause of variation in *Y*. Imagine that psychologists found that interfering with subjects' attention increased variance in perceptions of #TheDress. Under certain assumptions about attention, this might be taken to support the idea that attention is a cause of variation. The (sometime) epistemic significance of interventions that increase Var(*Y*) is not a threat to the asymmetry in SHRINK, given that SHRINK is an account of the truthmakers of causal claims about variation.

5.2 Off-Path Variables and the Causes of Variation

Another feature of SHRINK that requires explanation is its handling of off-path variables. Recall that, when considering an intervention on *X* with respect to *Y*, one must decide where to hold fixed the variables not on a path from *X* to *Y*. Woodward (2003) proposes two strategies, appropriate for different kinds of causal claims: (i) set off-path variables at any possible combination of values, or (ii) set off-path variables at their actual values. Although I have adopted the interventionist framework, my account proposes to (iii) set off-path variables at values that maintain the *actual distribution* of values in the population of interest. Unlike (ii), which holds fixed the actual properties of individuals, (iii) holds fixed actual population-level features. SHRINK claims that a variable *X* is a cause of variation in *p* iff there is an intervention on *X* with respect to *Y* that would reduce Var(*Y*) in which the actual distribution of values of variables not on a path from *X* to *Y* in *p* is held fixed. This means that there must be some way of wiggling *X* that reduces the variance in *Y*, where that wiggling is done against a causal backdrop in which off-path variables remain unchanged at the population level.

The intuitive motivation for this approach is that it guarantees that causal claims about individual differences in *p* reflect the characteristics of *p*. As was noted in Section 2, all claims about variation are restricted to a specific population. This population-relativity can be captured by assigning values to off-path variables in a way that preserves the actual distribution of values in the target population. *X* is reasonably called a cause of variation in *Y* in *p* iff wiggling *X* brings about the requisite changes in *Y* when other background conditions in *p* are kept constant.

Although it may be appropriate to adopt strategies (i) and (ii) for the handling of off-path variables when evaluating typical type- and token-causal claims, neither strategy is suitable for type-causal claims about variation. First, consider what would happen if we adopted (i), that is, if a cause of variation were any variable that could be wiggled to reduce Var(*Y*) against *any possible causal background* (i.e., with off-path variables set to any of their possible values). This proposal makes specification of the population largely inert, thereby failing to capture many intuitive judgments about the causes of variation in a population. Imagine that screen brightness (*B*) and screen type (*T*) are potential causes of subjects' color perceptions of the dress (*Y*). *B* is a continuous variable that takes values between 0 and 100, with higher values corresponding to greater brightness. *T* is binary: *T*=0 for conventional screens, and *T*=1 for brand new super screens. These super screens, we may imagine, eliminate color ambiguity when turned up to their maximum brightness (*B*=100). When an individual is using a conventional screen, however, manipulating brightness has no effect on their perception of the dress. Since super screens are not in commercial circulation, each person who comes across the dress online views it on a conventional screen. But if people were using super screens, it would be possible to increase screen brightness and in so doing cause everyone to see the dress in the same way.

I suggest that in this scenario, brightness is intuitively not a cause of variation in perceptions of the dress among current internet users (our focal explanandum), since every individual in *p* is using a conventional screen. SHRINK delivers this verdict. To assess whether *B* causes variation in *Y*, we hold fixed the actual distribution of *T* values in *p* (*T*=0 for all individuals) and find that there is no way to wiggle *B* that reduces Var(*Y*). So *B* is not a cause of variation in *Y* in this population. However, had we adopted strategy (i) for handling off-path variables, we would have gotten the counterintuitive verdict that screen brightness *is* a cause of variation in perceptions of the dress. Under (i), we can set off-path variables to any possible values. If we set *T*=1 for all individuals in *p*, intervening to make *B*=100 reduces Var(*Y*) to zero. So, *B* would count as a cause of variation in *Y*. This example shows that adopting strategy (i) renders the population (and the values of variables which characterize it) largely irrelevant to the formulation of causal claims about variation. This is unacceptable: the causes of variation in a trait or behavior *Y* should reflect the characteristics of *p*.[16]

Defining causes of variation in the manner described by (i) therefore does not do justice to the population-restricted nature of causal claims about variation. What about strategy (ii), in which off-path variables are set to their actual values? Had we adopted (ii) instead of (iii), we would

[16] Note the important difference between the case of chromaticity, discussed in Section 4.2, and this toy example: the former concerned whether chromaticity, a uniform factor in the target population, can *itself* help explain variation in perceptions of the dress. (I argued that it can.) In this case, we're asking whether screen brightness can explain the variation given that the *background conditions* involve a variable, screen type, that is uniform in *p*. (I suggest that it can't.) In other words, the analogue of chromaticity in this toy example is screen type, not screen brightness. My claim that screen brightness is not explanatory is compatible with the claim that screen type *is* explanatory. There is therefore no inconsistency between my earlier treatment of chromaticity and my discussion of this example. Still, the possibility that screen type might be explanatory raises legitimate questions about *which* uniform variables contribute to the explanation of individual differences; I return to this issue in Section 5.4.

count X as a cause of variation in Y iff wiggling X reduces Var(Y) while keeping all off-path variables at their actual values. While this generates the correct verdict in the toy example just discussed, it is needlessly restrictive. It requires that the values of off-path variables for every individual in p be fixed at their actual values. But the presence of individual differences is a population-level, not individual-level, phenomenon. To ask for an explanation of variation is to seek information about a group rather than any of its individual members. Causes of variation are factors that have a certain kind of influence on the population; the impact they have on specific individuals does not matter in itself. What matters in the assignment of values to off-path variables, then, is that the causal background conditions in the population are held constant, not that every single individual remains the same.

For these reasons, extant strategies for handling off-path variables are not suitable for the definition of a (type-)cause of variation. SHRINK incorporates an alternative strategy, one that allows some flexibility in the assignment of values to off-path variables but ensures that the assigned values are representative of the target population. It is worth reflecting briefly on what this modification means for interventionism generally. Interventionists have shown that strategies (i) and (ii) successfully capture most of our causal judgments about paradigm type- and token-causal claims like "smoking causes cancer" or "the striking of the match caused it to light." However, I have argued that causal claims about variation require (iii), the holding fixed of off-path variables' distribution of values in the population. There is little reason to think that claims about variation are unique in requiring a more nuanced approach to the assignment of values to off-path variables. For example, I suspect that (iii) is apt for other causal claims that are restricted to a population, such as, "For American schoolteachers, smoking reduces stress." This claim seems to be true only if intervening on a variable M representing smoking changes the value of a variable

S representing stress when the actual distribution of off-path variables characterizing the population of American teachers is held fixed. Let's imagine that wiggling *M* does indeed change the value of *S* under these conditions. It is irrelevant that this might not be the case if, contrary to fact, American teachers were not underpaid and overworked. We must hold fixed the distribution of values of variables that are not on a causal path between *M* and *S* to see whether smoking has an impact given the actual background conditions that hold among American teachers.

Moreover, the fixing of a population-level distribution of values is not the only alternative way of handling off-path variables that is needed. Consider the type-causal claim, "Smoking causes cancer even when cigarettes do not contain tar." This claim would seem to be true only if there is an intervention on variable *M* representing smoking with respect to variable *N* representing the probability of getting cancer that changes the value of *N* when an off-path variable *R* representing the presence of tar in cigarettes is held fixed at the value for "no tar" (say, $R=0$). If the only interventions on *M* that change *N* occur against a causal background in which tar *is* present ($R>0$), then the above claim is false. This suggests that the distinction between type and token causation cannot fully guide the assignment of values to off-path variables. The appropriate assignments depend on the causal claim under consideration. Strategies (i) and (ii) may be adequate for the paradigmatic causal claims that have animated philosophical discussions about causation, but analyzing more complicated causal claims requires a subtler approach.

Investigating the explanation of variation has therefore drawn our attention to a more general issue: within interventionism, the way one assigns values to off-path variables must be sensitive to the fine-grained content of the causal claim under consideration.

5.3 SHRINK and Scientific Practice

Let's now examine what can be said in favor of SHRINK, beyond the fact that it captures intuitive judgments about toy cases. As mentioned above, SHRINK, unlike ACTUAL, correctly implies that uniform variables like the chromaticity profile of the dress can help to explain variation explananda. It also captures key aspects of experimental design. Much scientific work on individual differences involves the same basic procedure: variation is identified, a hypothesis about its source is formulated, and an experiment is performed in which the hypothesized cause is manipulated with the aim of reducing the variation. If the manipulation succeeds, the causal hypothesis is favorably assessed. Research on the color constancy explanation conducted by Hesslinger and Carbon (2016) follows this procedure. Their Experiment 1 involves scrambling the image of the dress (i.e., dividing the image up into squares and randomly mixing them up) to give subjects less information about the illumination in the scene. They reason as follows: "If the color constancy explanation…is right, that is if interindividual differences in the use of illumination information significantly contribute to the differences in the perception of the dress, these differences will decrease when the illumination information of the image is reduced…by means of image scrambling" (*ibid.*, 3). They find that, indeed, the more scrambled the image, the less variation in subjects' color reports about the dress. They take this to support the hypothesis that divergent assumptions about illumination cause individual differences in perceptions of the (original) photograph.

SHRINK nicely captures the motivation for Hesslinger and Carbon's experiment. To assess whether assumptions about illumination are a cause of the variation in subjects' color judgments, they try to find an intervention on illumination assumptions (X) with respect to color perception (Y) that reduces Var(Y), while keeping fixed the actual distribution of values of other variables in p. The reason the authors used image scrambling is that it disrupts illumination cues while "preserv[ing] the local color information of the image" (*ibid.,* 4). Image scrambling is an appropriate intervention, they claim, because it influences the variance in color judgments only through its effect on assumptions about illumination.

Experimental design is often subject to criticism, and SHRINK captures the reasoning behind these challenges as well. Jonauskaite et al. (2018) cast doubt upon the evidential value of Hesslinger and Carbon's experiment by drawing attention to another potential cause that is affected by image scrambling: apparent texture. Since image scrambling reduces information about texture as well as illumination, the experiment does not provide strong support for the color constancy explanation relative to a texture-based explanation. Filtered through SHRINK, Jonauskaite and colleagues are arguing that Hesslinger and Carbon fail to establish that illumination assumptions are a cause of variation because image scrambling is not an intervention on illumination assumptions with respect to color perception. A genuine intervention on *X* with respect to *Y* must not influence causes of *Y* that are not on a path from *X* to *Y.* Image scrambling changes textural information, an off-path cause of color perception, so it is not an intervention of the requisite kind.

Finally, SHRINK also helps make sense of the close connection between research on variation and regularity. The explanation of variation in a population and the explanation of regularity within a *sub*population are often two sides of the same coin. People's differing assumptions about illumination help explain variation in perceptions of the dress among all the individuals who came across the photograph online. But we might also want to explain regularities within subpopulations, such as when we ask why people who see the dress as white and gold see it that way. Here too illumination assumptions have an explanatory role to play. We might explain that people in this subpopulation tend to assume that the dress is in shadow, triggering color constancy mechanisms which discount the short wavelengths of the illuminant, making the dress appear yellowish (see quote from Wallisch [2017] above). Illumination assumptions are thus an important variable for explaining *both* variation in the wider population *and* regularities in each subpopulation. From the perspective of SHRINK, this is to be expected: what it counts as causes of variation will often turn out to be causes of regularity for subpopulations of individuals as well. The variables that can be intervened on to reduce variance in the population as a whole can likely be intervened on to change outcomes in a subpopulation too.[17] SHRINK therefore captures the continuity between the explanation of variation and the explanation of regularity.

5.4 An Objection

One might object to SHRINK on the grounds that it categorizes too many variables as causes of variation. Variables that exert an extremely minimal influence on Var(Y) count as explanatory under SHRINK, as do counterfactuals like, "If there had been a meteor strike that killed all life on earth, there would be no variation in perceptions of the dress." For a more realistic example, consider Hugrass et al.'s (2017) work showing that subjects' perceptions of the dress can be changed by inducing illusory shifts in the brightness of the photograph. Hugrass and colleagues produce illusory dimming and brightening using an established technique in which rotating sawtooth gratings are superimposed on an image. When the photograph of the dress seems to be dimming, subjects are more likely to see it as blue and black; when it seems to be brightening, responses tend toward white and gold. Consider a continuous variable S that represents illusory brightness shift in the image of the dress ($S > 0$ for brightening, $S < 0$ for dimming). Assuming

[17] I am here assuming that a standard interventionist story about how to pick out regular causes, with the caveat that, if my suggestions toward the end of Section 5.2 are on the right track, causal claims about subpopulations require a nonstandard approach to the handling of off-path variables. (See the example above about American schoolteachers, which involves a claim about regularity within a subpopulation.)

there is an illusory brightness shift that would reduce the variance in perceptions of the dress, *S* counts as a cause of variation under SHRINK. This may seem strange. It doesn't seem explanatory to say that people perceive the colors of the dress differently in ordinary contexts because, as a matter of fact, the apparent brightness of the photograph is *not* changing (S=0). Although some variables that are uniform in the population (like chromaticity profile) help explain individual differences, other uniform variables (like brightness shift) arguably do not. But SHRINK does not distinguish between them.

There are resources in the literature to counter this objection, which is a species of the broader worry that interventionism counts too many things as causes. First, interventionists are not committed to all explanations being equally informative. Deeper explanations "give answers to a finer-grained, more detailed, and wider-ranging set of what-if-things-had-been-different questions" (Woodward 2003, 222). The brightness shift explanation may not answer a wide range of w-questions, which is partly why its explanatory power is limited.

Second, we can invoke the notion of a "serious possibility" to rule out outlandish explanations. Woodward argues that, "if a change in a purported explanans is associated with some corresponding change in an explanandum, but the change in the explanans is not a serious possibility, then the information that the explanandum will change under this change in the explanans is typically not regarded as explanatory, or at least the purported explanation is not seen as satisfying or relevant" (*ibid.,* 227).[18] The absence of a meteor strike does not explain the presence of variation in perceptions of the dress because it is not a serious possibility. Similarly, the possibility that sawtooth gratings could be applied to the dress photograph is not one that we take seriously outside Hugrass and colleagues' laboratory. By contrast, we readily imagine that the

[18] Against the charge that the notion of a "serious possibility" renders interventionism overly subjective, Woodward argues that alternative accounts of causation require an analogous concept (*ibid.,* 89).

chromaticity profile might have been different – for instance, if the photograph had been taken in slightly different lighting conditions, or if the dress itself had faded in the wash.

Third, SHRINK can be defended by appealing to the philosophical literature on causal selection (Hesslow 1988). Even though SHRINK does indeed count many variables as causes of variation, there are strategies for selecting the most important or explanatory causes from among the full set. Criteria that have been proposed for causal selection include causal strength, invariance, proportionality, and specificity (Hitchcock and Woodward 2003, Woodward 2010, Ross forthcoming).

For these three reasons, it is not a problem that SHRINK's notion of a cause of variation is quite liberal, nor that SHRINK does not distinguish uniform variables that have significant explanatory power from those that do not. Interventionist (and even some non-interventionist) accounts of causation and causal explanation cast a wide net. It is generally recognized that they must be supplemented with additional (often pragmatic) constraints that pick out the variables most relevant to the explanation at hand.

5.5 SHRINK on Noise

One attractive upshot of SHRINK is that it suggests a principled way of distinguishing between noise and variation. Researchers frequently distinguish individual differences that are unstructured or uninteresting from differences that are systematic or theoretically important. The former are called "noise" and the latter, "variation" (e.g., Finn et al. 2017, Seghier and Price 2018). Different definitions of noise have been offered in the scientific (Faisal et al. 2008), statistical (Upton and Cook 2008), and philosophical (McAllister 1997) literatures. One common theme is that noise, unlike variation, is fundamentally inimical to explanation. While the sources of variation can be catalogued and understood, noise is by its very nature resistant to explanation.

SHRINK can be used to make this idea precise. Noise, on a SHRINK-inspired view, is the heterogeneity between units of analysis that cannot be explained given the variables one is willing to countenance. When there are no causes of variation (as defined by SHRINK) within one's potential variable set, one is dealing with noise. No variable in the set can be intervened on in the appropriate way to reduce the population variance. Genuine variation, meanwhile, is heterogeneity that is amenable to explanation in the manner laid out in SHRINK. The distinction between noise and variation, then, boils down to a distinction between heterogeneity that could be reduced by hypothetical intervention, and heterogeneity that could not be so reduced.

This view of noise leans heavily on the notion of a variable that one is willing to countenance. Heterogeneity that counts as noise for one researcher (e.g., a cognitive psychologist interested in cognitive variables) may count as variation for a different researcher (e.g., an electrophysiologist who deals with variables representing properties of neurons). There may be no variable that the cognitive psychologist would entertain that could (in principle) be intervened on to reduce the population variance, even though such a variable does exist for the electrophysiologist. The identification of noise is thus relative to a variable set. While some might object to this aspect of the view, there is widespread recognition that scientific context does partly determine whether heterogeneity qualifies as noise or variation (Cronbach 1957). Moreover, although the relevant variable set depends on one's interests, once that set is determined, it is an objective matter whether the differences in question constitute noise or variation. Either there is a variable in the set that meets the conditions in SHRINK or there is not.

Defending this view of noise would require a thorough comparison with alternative accounts. The suggestion itself, however, illustrates the applicability of SHRINK to topics beyond causation and explanation.

6. Conclusion

I have argued here that explananda about variation constitute a distinct and undertheorized class of scientific explananda. Drawing from research on #TheDress, I explored what the interventionist ought to say about how to explain a variation explanandum. One plausible proposal, ACTUAL, was rejected. Explaining individual differences does not exclusively involve actual difference makers because explanatorily relevant variables may be uniform in the target population. Instead, I argued, explaining a variation explanandum requires exhibiting causes of variation, which are variables that can be intervened on to reduce population variance. This account, SHRINK, captures the explanatory relevance of uniform variables, reflects experimental practice, and suggests an appealing way of drawing the distinction between noise and variation.[19]

Although my focus has been on the explanation of individual differences in psychology, SHRINK applies to other domains as well. Causally explaining variation in a variable *Y* that ranges over any unit of analysis involves identifying factors that could be intervened on to reduce population variance in the way described by SHRINK. In biology, the units of analysis might be individual organisms; in physiology, individual muscle cells; in meteorology, individual hurricanes. SHRINK implies that explaining variation in hurricane trajectories requires identifying

[19] Since I have limited my discussion to causal explanation, there may be more to say about the non-causal explanation of variation explananda. Perhaps there are statistical, mathematical, or constitutive explanations of individual differences (Lange 2013a, 2013b; Yilkoski 2013; Lyon 2014). I suspect that non-causal explanations in science are less common than is usually supposed but cannot argue for this here.

variables which could hypothetically be intervened on to reduce the variance in trajectories, while holding fixed the distribution of other features of hurricanes in the population.[20]

With its broad applicability, SHRINK enriches our philosophical understanding of explanation – not just in psychology, but across the sciences – by characterizing what is distinctive about explaining variation rather than regularity. It also shows that there is a need for subtlety in the application of interventionism: we must attend to background conditions when analyzing causal claims about populations. Consideration of variation therefore promises not only to shed light on an underexplored area of psychological research, but also to contribute to the refinement of general philosophical accounts of science.

Acknowledgments: I started working on this paper while visiting Australian National University in 2019. I'm grateful to Colin Klein for inviting me to ANU and to everybody there for a fantastic visit. My sincere thanks to those who have offered feedback on the paper: Colin Allen, Mazviita Chirimuuta, David Danks, Dmitri Gallow, Gabby Johnson, Edouard Machery, Shivam Patel, and Jim Woodward. I've also had helpful conversations with Janella Baxter, John Campbell, Maria Olkkonen, Joel Velasco, Tom Wysocki, and others who attended presentations of the paper. This work was supported by a Mellon/ACLS Dissertation Completion Fellowship.

[20] Woodward (1995) argues that linear regression equations "typically explain actual variation in some specific population," assigning larger coefficients to factors that are of greater causal significance to variability in the dependent variable (46). Woodward's interpretation of linear regression models is consistent with (though not entailed by) SHRINK. An independent variable assigned a nonzero coefficient in a regression can be manipulated to reduce the variance in the dependent variable, so long as certain assumptions (e.g., about the direction of causation, the variable's manipulability) hold. If Woodward is right that regression equations pick out causes of variation in a population, then the scope of my account is very wide indeed, given the ubiquity of linear regression in science.

References

Aston, Stacey, and Anya Hurlbert. 2017. "What #theDress Reveals about the Role of Illumination Priors in Color Perception and Color Constancy." *Journal of Vision* 17 (9): 4–4.

Baxter, Janella. 2019. "How Biological Technology Should Inform the Causal Selection Debate." *Philosophy, Theory, and Practice in Biology* 11 (2).

Brainard, David H., and Anya C. Hurlbert. 2015. "Colour Vision: Understanding #TheDress." *Current Biology: CB* 25 (13): R551-554.

Capizzi, Mariagrazia, Ettore Ambrosini, and Antonino Vallesi. 2017. "Individual Differences in Verbal and Spatial Stroop Tasks: Interactive Role of Handedness and Domain." *Frontiers in Human Neuroscience* 11 (November).

Chetverikov, Andrey, and Ivan Ivanchei. 2016. "Seeing 'the Dress' in the Right Light: Perceived Colors and Inferred Light Sources." *Perception* 45 (8): 910–30.

Cronbach, Lee J. 1957. "The Two Disciplines of Scientific Psychology." *American Psychologist* 12 (11): 671–84.

Currie, Adrian. 2018. *Rock, Bone, and Ruin: An Optimist's Guide to the Historical Sciences*. Cambridge, MA: MIT Press.

de Lange, Floris P., Micha Heilbron, and Peter Kok. 2018. "How Do Expectations Shape Perception?" *Trends in Cognitive Sciences* 22 (9): 764–79.

Drissi-Daoudi, Leila, Adrien Doerig, Khatuna Parkosadze, Marina Kunchulia, and Michael H. Herzog. 2020. "How Stable Is Perception in #TheDress and #TheShoe?" *Vision Research* 169 (April): 1–5.

Faisal, A. Aldo, Luc P. J. Selen, and Daniel M. Wolpert. 2008. "Noise in the Nervous System." *Nature Reviews. Neuroscience* 9 (4): 292–303.

Finn, Emily S., Dustin Scheinost, Daniel M. Finn, Xilin Shen, Xenophon Papademetris, and R. Todd Constable. 2017. "Can Brain State Be Manipulated to Emphasize Individual Differences in Functional Connectivity?" *NeuroImage* 160: 140–51.

Gegenfurtner, Karl R., Marina Bloj, and Matteo Toscani. 2015. "The Many Colours of 'the Dress.'" *Current Biology* 25 (13): R543–44.

Glöckner, Andreas, and Thorsten Pachur. 2012. "Cognitive Models of Risky Choice: Parameter Stability and Predictive Accuracy of Prospect Theory." *Cognition* 123 (1): 21–32.

Griffiths, Paul, and Karola Stotz. 2013. *Genetics and Philosophy: An Introduction*. New York: Cambridge University Press.

Hesslinger, Vera M., and Claus-Christian Carbon. 2016. "#TheDress: The Role of Illumination Information and Individual Differences in the Psychophysics of Perceiving White–Blue Ambiguities." *I-Perception* 7 (2): 2041669516645592.

Hitchcock, Christopher. 2001. "The Intransitivity of Causation Revealed in Equations and Graphs." *The Journal of Philosophy* 98 (6): 273–99.

Hitchcock, Christopher, and James Woodward. 2003. "Explanatory Generalizations, Part II: Plumbing Explanatory Depth." *Noûs* 37 (2): 181–99.

Holderness, Cates. 2015. "What Colors Are This Dress?" BuzzFeed. February 26, 2015. https://www.buzzfeed.com/catesish/help-am-i-going-insane-its-definitely-blue.

Hugrass, Laila, Jana Slavikova, Melissa Horvat, Alaa Al Musawi, and David Crewther. 2017. "Temporal Brightness Illusion Changes Color Perception of 'the Dress.'" *Journal of Vision* 17 (5): 6–6.

Jonauskaite, Domicele, Nele Dael, C. Alejandro Parraga, Laetitia Chèvre, Alejandro García Sánchez, and Christine Mohr. 2018. "Stripping #The Dress: The Importance of Contextual Information on Inter-Individual Differences in Colour Perception." *Psychological Research*, September.

Knecht, S., M. Deppe, B. Dräger, L. Bobe, H. Lohmann, E. Ringelstein, and H. Henningsen. 2000. "Language Lateralization in Healthy Right-Handers." *Brain: A Journal of Neurology* 123 (January): 74–81.

Lafer-Sousa, Rosa, and Bevil R. Conway. 2017. "#TheDress: Categorical Perception of an Ambiguous Color Image." *Journal of Vision* 17 (12): 25.

Lange, Marc. 2013a. "Really Statistical Explanations and Genetic Drift." *Philosophy of Science* 80 (2): 169–188.

———. 2013b. "What Makes a Scientific Explanation Distinctively Mathematical?" *British Journal for the Philosophy of Science* 64 (3): 485–511.

Lewontin, R. C. 1970. "The Units of Selection." *Annual Review of Ecology and Systematics* 1: 1–18.

———. 1974. "The Analysis of Variance and the Analysis of Causes." *American Journal of Human Genetics* 26: 400–411.

Lipton, Peter. 1990. "Contrastive Explanation." *Royal Institute of Philosophy Supplement* 27: 247-266.

Lyon, Aidan. 2014. "Why Are Normal Distributions Normal?" *British Journal for the Philosophy of Science* 65 (3): 621–649.

Mackie, J. L. 1974. *Cement Of The Universe*. Oxford: Oxford University Press.

Martín-Moro, J.G., F.P. Garrido, F.G. Sanz, I.F. Vega, M.C. Rebollo, and P.M. Martín. 2018. "Which Are the Colors of the Dress? Review of an Atypical Optic Illusion." *Archivos De La Sociedad Espanola De Oftalmologia* 93 (4): 186–92.

McAllister, James W. 1997. "Phenomena and Patterns in Data Sets." *Erkenntnis* 47 (2): 217–28.

Northcott, Robert. 2009. "Is Actual Difference Making Actually Different?" *Journal of Philosophy* 106 (11): 629–633.

Olkkonen, Maria, and Vebjørn Ekroll. 2016. "Color Constancy and Contextual Effects on Color Appearance." In *Human Color Vision*, 159–88. Springer Series in Vision Research. New York, NY: Springer.

Pearce, Bradley, Stuart Crichton, Michal Mackiewicz, Graham D. Finlayson, and Anya Hurlbert. 2014. "Chromatic Illumination Discrimination Ability Reveals That Human Colour Constancy Is Optimised for Blue Daylight Illuminations." *PLOS ONE* 9 (2): e87989.

Reeder, Reshanne R. 2017. "Individual Differences Shape the Content of Visual Representations." *Vision Research* 141: 266–81.

Ross, Lauren N. forthcoming. "Causal Control: A Rationale for Causal Selection." In *Causal Reasoning in Biology*, by B. Hanley, C.K. Waters, and J.F. Woodward. Minnesota Studies in the Philosophy of Science. University of Minnesota Press.

Seghier, Mohamed L., and Cathy J. Price. 2018. "Interpreting and Utilising Intersubject Variability in Brain Function." *Trends in Cognitive Sciences* 22 (6): 517–30.

Stegmann, Ulrich E. 2012. "Varieties of Parity." *Biology and Philosophy* 27 (6): 903–918.

Swets, Benjamin. 2015. "Psycholinguistics and Planning: A Focus on Individual Differences." In *Individual Differences in Speech Production and Perception*, edited by Susanne Fuchs, Daniel Pape, Caterina Petrone, and Pascal Perrier, 89–122. New York, NY: Peter Lang.

Tabery, James. 2009. "Difference Mechanisms: Explaining Variation with Mechanisms." *Biology and Philosophy* 24 (5): 645–664.

Toscani, Matteo, Karl R. Gegenfurtner, and Katja Doerschner. 2017. "Differences in Illumination Estimation in #thedress." *Journal of Vision* 17 (1): 22–22.

Unsworth, Nash, Josef C. Schrock, and Randall W. Engle. 2004. "Working Memory Capacity and the Antisaccade Task: Individual Differences in Voluntary Saccade Control." *Journal of Experimental Psychology. Learning, Memory, and Cognition* 30 (6): 1302–21.

Upton, Graham, and Ian Cook. 2008. *A Dictionary of Statistics*. 2nd ed. Oxford: Oxford University Press.

Wallisch, Pascal. 2017. "Illumination Assumptions Account for Individual Differences in the Perceptual Interpretation of a Profoundly Ambiguous Stimulus in the Color Domain: 'The Dress.'" *Journal of Vision* 17 (4): 5–5.

Wallisch, Pascal, and Michael Karlovich. 2019. "Disagreeing about Crocs and Socks: Creating Profoundly Ambiguous Color Displays." https://arxiv.org/abs/1908.05736v1.

Ward, Zina B. 2022. "Cognitive Variation: The Philosophical Landscape." *Philosophy Compass* 17 (10): e12882.

Waters, C. Kenneth. 2007. "Causes That Make a Difference." *Journal of Philosophy* 104 (11): 551–579.

Witzel, Christoph, J. Kevin O'Regan, and Sabrina Hansmann-Roth. 2017a. "The Dress and Individual Differences in the Perception of Surface Properties." *Vision Research* 141: 76–94.

Witzel, Christoph, Chris Racey, and J. Kevin O'Regan. 2017b. "The Most Reasonable Explanation of 'the Dress': Implicit Assumptions about Illumination." *Journal of Vision* 17 (2): 1–1.

Witzel, Christoph, and Matteo Toscani. 2020. "How to Make a #theDress." *JOSA A* 37 (4): A202–11.

Woodward, James. 1995. "Causation and Explanation in Econometrics." In *On the Reliability of Economic Models: Essays in the Philosophy of Economics*, edited by Daniel Little, 9–61. Recent Economic Thought Series. Dordrecht: Springer Netherlands.

———. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford, UK: Oxford University Press.

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

———. 2015. "Interventionism and Causal Exclusion." *Philosophy and Phenomenological Research* 91 (2): 303–47.

Woodward, James, and Christopher Hitchcock. 2003. "Explanatory Generalizations, Part I: A Counterfactual Account." *Noûs* 37 (1): 1–24.

Wysocki, Tomasz. In preparation. "The Underdeterministic Framework."

Ylikoski, Petri. 2013. "Causal and Constitutive Explanation Compared." *Erkenntnis* 78 (2): 277–297.