Taking Social Science Seriously: Grit Psychology as a Test Case for Theories of Causation, Explanation, and Counterfactuals

Kino Zhao

Simon Fraser University

kino_zhao@sfu.ca

Draft. Please do not cite. Comments welcome.

Abstract

Philosophy of science has historically had a physics focus, where the success of its theories is tested against paradigmatic examples in physics, and then modified to accommodate "special sciences" by adding caveats and exceptions. This paper aims to show that there is value for general philosophy of science to take the social sciences seriously by positing grit psychology as a testing case and treating it in much of the same way that philosophers of science have been treating physics examples. I consider theories of causation, explanation, and counterfactuals in the grit example and discuss the philosophical ramifications.

1. Introduction

Philosophy of science has historically focused on physics as the model science. That is, philosophical theories of scientific concepts, such as causation, explanation, and counterfactuals, are often tested on paradigmatic physics examples, and physics examples alone, for assessment of descriptive adequacy and prescriptive value. More recently, development in the philosophy of biology has shown that taking biology as the primary testing ground for philosophical theories of science is not only productive for understanding biology, but can also lead to novel perspectives in general philosophy of science. Similar efforts have also since occurred in areas such as climate science and economics.

Accordingly, many recent theories in the philosophy of science are careful about not restricting their use context to physics alone. For example, Woodward's interventionist account of causation explicitly advertises itself as allowing causal theorizing in non-physical contexts. It does so by leaving

the task of specifying appropriate causal variables to domain scientists. That is, instead of characterizing causation in such a way that only physical events can serve as causes (as, for example, the contact theory of causation does), the interventionist theory of causation takes the position that it is not the job of the theory of causation to specify what counts as an appropriate cause. When physicists apply interventionism to physics, they will distinguish appropriate from inappropriate physical causes. When social scientists apply interventionism to the social world, they will identify appropriate from inappropriate social causes. The theory of causation does not care if these causal variables are emergent or otherwise irreducible to physics.

While not all pluralistic theories in the philosophy of science target the social sciences as an area to include, many do. The present paper aims to put these theories to the test in a social psychology case study, much in the same way that old philosophical theories of science must pass the test example of Newtonian mechanics. The goal is to lay out, in a systematic way, what is required to give a philosophical account of what have been traditionally considered as core scientific concepts in a social science context.

Insofar as possible, I will hold the same testing standards physics-inspired theories are typically held to. On the one hand, I will allow the philosophical theories to diagnose certain scientific practices as misguided, even if these practices are unchallenged within the scientific community. On the other hand, if a philosophical theory attempts to rule out, from first principles, widely engaged scientific activities as impossible, or if its description of scientific activities seems to miss the point, then I will pronounce the philosophical theory inadequate.

This paper is organized as follows. In section 2, I provide background information for my testing example -- trait grit -- as well as some justification for this choice. In sections 3, 4, and 5, I discuss three major themes in the grit literature that would benefit from philosophical theorizing: whether grit counts as a proper cause, whether grit explains behaviour, and what it means to reason counterfactually about grit. Section 6 concludes.

2. A brief history of grit

In 2007, the psychologist Angela Duckworth and colleagues proposed grit as a personality trait that explains why "some individuals accomplish more than others of equal intelligence" (Duckworth et al., 2007). They developed a 12-item questionnaire that purportedly measures grit along two facets:

Consistency of Interests and Perseverance of Effort. The former asks questions such as whether the subject's "interests change from year to year", while the latter asks whether the subject is "a hard worker" (Duckworth et al., 2007).

Grit is thus proposed as a "noncognitive" determinant of success. The distinction between cognitive versus noncognitive attributes is somewhat blurry and does not track whether the construct involves a conscious, reflective component (Messick, 1979). The initial motivation behind the distinction is to differentiate between when low performance on an intelligence test results from lacking ability and when it results from lacking motivation to perform well or being distracted by something outside the test. In the context of education, noncognitive attributes also refer to less transient features such as study habits and educational ambition (see, e.g., Bowman et al., 2019).

Since 2007, a myriad of studies has been conducted around the efficacy of grit as a noncognitive determinant of academic (Duckworth et al., 2011; Strayhorn, 2014; Tang et al., 2021) and vocational (Mueller et al., 2017; Suzuki et al., 2015; Walker et al., 2016) success. Although psychologists are often very careful not to use the word "cause", there should be no doubt that grit is meant to stand in a causal relationship with success. While many studies cited in favour of grit are correlational, researchers often take their conclusions to motivate grit intervention programs. A U.S. Department of Education report (Shechtman et al., 2013) even goes as far as highlighting grit as an area that "should play an essential role in evolving educational priorities". It is clear that, in addition to whatever epistemic merit grit research might have as a purely scientific exercise, its sociopolitical stakes must be understood in causal terms.

As is typical in fields such as social psychology, education, and business management, studies around grit often suffer from a variety of methodological issues. These include sampling bias (Pendyala & Vyas, 2023) and measurement problems (Credé, 2018; see also Zhao, 2023). Curiously, inconsistent correlational evidence (e.g., Forbes & Fikretoglu, 2018; Direito et al., 2021) has not been the primary focus of grit skeptics. Instead, much emphasis has been placed on questioning the internal coherence of grit (Jachimowicz et al., 2018; Guo et al., 2019; Postigo et al., 2021) and the screen-off relationship between grit and adjacent constructs (Hicks, 2010; Meriac et al., 2015; Rimfeld et al., 2016). There is thus a variety of philosophical themes relevant in the grit debate that go beyond evidential inconclusiveness.

Grit is an ideal testing ground for philosophical theories for several reasons. First, whether grit is a legitimate construct that causes and explains behaviour is still an unsettled question within social psychology. This means that researchers are still debating issues such as how to understand the theoretical foundations of grit or interpret evidence around it. This gives philosophers of science an opportunity to check if our analysis respects scientific concerns in appropriate ways. Second, grit as a personality trait is neither purely macroscopic like a network's structure, nor plausibly reducible like pain. It therefore forces us to tackle the question of nonreductive, nonemergent causation seriously.

Third, the idea of noncognitive determinants of success in general, if not grit in particular, is difficult to dismiss as a case of causal reasoning. Many of us had to consciously develop strategies to cope with public speaking anxiety when we first started teaching. We needed to figure out ways to stay motivated after journal rejections or grant denials. We routinely teach these skills to our students in the hopes of helping them succeed. It seems unnatural to not use causal language to describe our own activities and reasoning. It seems unnecessarily pessimistic to say that such descriptions cannot be studied scientifically, even if we know such a study would be difficult.

Ultimately, I am open to the possibility that the notion of causation, explanation, or counterfactuals invoked in the grit case are *sui generis* in some sense and should not be analyzed in the same way as flagpoles and barometers. However, I would like to encourage philosophers of science to not start with this assumption.

3. Grit as a Cause

The theory of causation most naturally applied in the grit case is interventionism, which comes in several flavours (Menzies & Price, 1993; Woodward & Woodward, 2005; Pearl, 2009; Hitchcock, 2009). Instead of offering a comprehensive review of the vast literature on causation, I will discuss two components that are important to all interventionist accounts: precision intervention and the Markov screen-off condition. I will also offer some brief remarks on process and regularity theories in the grit context.

3.1 Precision Intervention

One common criticism of the interventionist account of causation is its reliance on precision intervention. The basic idea behind interventionism is that C causes E just in case intervening on C results in a change in E. This, however, is not enough, because an intervention in the ordinary sense

may also directly affect E or affects another cause of E, C', in addition to affecting C. Consequently, the intervention is typically considered to be precise or "surgical" (Pearl, 2009) in that it affects C alone. Every other change in the causal network must be a result of the change in C, and not the intervention.

The idea of precision interventions raises the question of feasibility. Here, theorists diverge. One strategy is to adopt a mathematical notion of intervention, where to intervene on C just is to set C to a particular value (Spirtes et al., 1993; Pearl, 2009). This strategy leads to a circularity worry concerning how we might come to know the effect of such an intervention on C without already assuming a causal relationship between C and E. The other camp is to assert that precision interventions are, in fact, fairly common (Menzies & Price, 1993; Weinberger et al., n.d.), and that examples of unmanipulable causes create problems for competing theories of causation as well. While this camp also faces the circularity challenge in the sense that it starts with the assumption that causation exists in the world, the challenge is much less worrisome because judgments of manipulability do not need to assume the existence of particular causal relationships (Woodward, 2023).

Before discussing the problem of precision intervention in the grit case, let me first highlight an interesting divergence in how psychologists theorize about personality traits as causes. As briefly summarized above, manipulability is essential to not only interventionism, but also many other theories of causation that invoke counterfactual dependence between a cause and its effect. In fact, causal theories that do not rely on counterfactual dependence, namely, regularity theories, are often taken as being antirealist about causation. Consequently, causal claims that involve unmanipulable causes are typically considered as challenging cases (e.g., Menzies & Price, 1993).

For personality psychologists, however, the relationship between manipulability and causation is much less immediate. Arguably, the most important aspect of the trait theory of personality is that traits, unlike habits or situations, are difficult to change (Allport, 1931; Johnson, 1997; see also Moreau et al., 2019). In fact, personality traits are considered appropriate causes of behaviours *because* their unmanipulability. During the so-called person-situation debate, trait psychologists debate with behaviourist-inspired psychologists over whether behaviours are best explained by long-term, trait-like tendencies or short-term stimulus-response analysis. In a critical review of situationism, Bowers calls the stimulus-response view of causation "misleadingly superficial" (Bowers, 1973). He then compares personality traits with laws of motion and argues that

understanding these laws provides us with genuine causal knowledge, even if they do not always increase the level of control we have at a day-to-day level.

We might reasonably wonder whether the conceptualization of grit as a personality trait (Duckworth et al., 2007) would conflict with the desire to develop grit-enhancing interventions. Curiously, I had not come across any skepticism over the causal power of grit along this line. On the one hand, while difficult to change, personality traits are not unchangeable (Roberts et al., 2017), and the occasional malleability of certain personality traits does not seem to deprive them of their trait status. On the other hand, the short-term experimental manipulability of grit has been taken as evidence that grit contains state-like features (DiMenichi & Richmond, 2015). Similarly, this finding has not caused the authors to question the causal efficacy of grit.

In addition to the desirability of manipulation, the feasibility question is similarly complicated in the case of grit. And this is true for both Pearl's *setting intervention* and Woodward's *possibility constrained intervention* (Woodward's 2023 terminologies), though in different ways.

Recall that Pearl's setting intervention is achieved by setting the value of a variable in a particular way, ignoring any feasibility constraints that might arise in the physical world. In the case of grit, the target of intervention would have to be the grit score, as measured by the Grit Scale. The Grit Scale, developed by Duckworth et al. (2007), is composed of two sub-scales, corresponding to the two factors, Consistency of Interests and Perseverance of Effort. The original recommendation in Duckworth et al. (2007) is to use the combined grit score as the predictor. However, researchers disagree over whether positing grit as a higher order factor is psychometrically legitimate (Credé, 2018; Tyumeneva et al., 2019) as well as whether both factors play a role in causing behaviour (Jachimowicz et al., 2018; Postigo et al., 2021). These concerns challenge the assumption that the overall grit score is the appropriate target of intervention. A proponent of set intervention may respond by saying that we should settle on a theory of grit before trying to precision-intervene on it. However, psychometric research typically involves extensive causal theorizing, and it would certainly be a weakness of our causal theory if it must be informed by, and therefore cannot inform, psychometric research.

The above worry is greatly alleviated in the case of possibility constrained intervention, because here, the manipulability of grit and its two components is determined by the world, rather than by our theory. If it is in fact the case that only Perseverance of Effort, but not Consistency of Interests, is

genuinely causal, then we would come to learn this fact when we discover that successful intervention programs target one but not the other. Indeed, this appears to be precisely what researchers are already doing when they claim that grit is better understood as unidimensional.

Understanding intervention as an actual activity faces a different set of challenges, however. Personality traits such as grit do not have a concretely defined set of behavioural manifestations or even psychological processes (Funder, 2001). Although the Grit Scale asks for specific behaviours such as maintaining focus or being diligent, there are often numerous ways of interpreting what counts as an instance of said behaviour. This makes it difficult to determine whether any given intervention program targets grit alone, or whether it targets a grit behaviour, which is the true cause for success. For example, programs that teach students study skills are often effective in improving their academic performance (Hattie et al., 1996), and it seems plausible that developing these skills may result in students classifying themselves as diligent and good at maintaining focus, thus scoring high on the Grit Scale. Unlike my worry with setting intervention, however, I believe it is reasonable, in this case, to put the burden on the grit theorists to determine the relationship between trait grit and grit behaviour. All I aim to do here is to flag the fact that the question of whether a certain intervention is precise is much more complicated in the case of social psychology.

3.2 The Markov screen-off condition

The Markov Condition (MC) is an important property in causal modeling and causal discovery. There are several equivalent formulations of it (Hitchcock, 2023), but I will focus on the screen-off formulation, as it is the most intuitive. Informally, MC states that, if C causes E, then the conditional probability of E on C is independent of every variable in the causal graph except for the effects of E.[1] That is, C "screens off" E from all other variables, except for the effects of E.

Although social psychologists are not quite thinking about causal structures as causal Bayes nets, many of their worries can be understood as worries about screening off. Since grit is a relatively late addition to traits that center around persistence and self-control (see Ryans, 1939 for a early review), skeptics have worried about whether, predictively, it adds anything new (Rimfeld et al., 2016; Credé

[1] This is, of course, an oversimplification. As (Steel, 2005) points out, the parent-descendent relationship in MC does not need to be interpreted as a causal relationship. A further Causal Markov Condition has also been formulated (Spirtes et al., 1993) and debated (Cartwright, 2002; Hausman & Woodward, 1999; Cartwright, 2006). Nevertheless, my analysis of grit does not invoke these subtleties. I will thereby proceed with this oversimplified gloss.

et al., 2017). In other words, critics argue that grit cannot be a genuine cause if its effects are screened off by variables such as conscientiousness.

This is an area where psychological debates about the causal structure around grit and related constructs can benefit from philosophical reasoning about causal models in several ways. First, the benefit of Markov screen-off is that it provides directionality for completed causal graphs. However, screen-off relations that occur in incomplete graphs are not always indicative of such relations in the complete graph (Sober & Steel, 2013). That is, variable A may screen off variable B in an incomplete graph, but fails to do so in a complete graph, where A and B share previously-unknown common causes. Since causal reasoning about human behaviour is very far from complete at present, the fact that conscientiousness screens off grit at present may not mean that it will continue to do so throughout future development of the causal graph.

Second, it seems that current skepticism of grit based on screen-off evidence is largely motivated by the fact that grit is a late comer when compared with conscientiousness and self-control. This is, of course, not a very convincing reason. Ideally, we would like to study whether conscientiousness screens grit off of all of its effects but not vice versa, so we can determine which is the more useful variable to keep in the final causal model.

Unfortunately, without precision intervention on almost every variable, it is often impossible to distinguish between several Markov equivalent structures (Eberhardt et al., 2005). Furthermore, even with precision intervention, there would still be structural underdetermination if the causal graph is incomplete (Eberhardt, 2013). Given the difficulties associated with precision intervention and causal completeness in the grit context discussed above, theorists should seriously consider alternative causal structures that are equally supported by the current screen-off evidence.

Third, there is sometimes a conflation between grit's (lack of) psychometric independence and its (lack of) causal independence. Several studies have challenged the construct validity of grit by pointing to its close connection with conscientiousness in structural equation models (Rimfeld et al., 2016; Schmidt et al., 2018; Ponnock et al., 2020), which is a different kind of challenge from that of grit's causal efficacy, and faces a different set of methodological issues unique to structural equation modeling. In a sense, this is a much deeper challenge to grit's legitimacy as a construct than causal screen-off. Nevertheless, critics (e.g., Credé et al., 2017) and defenders (e.g., Duckworth & Gross, 2014) alike cite grit's predictive edge as evidence in support of its construct independence.

3.3 Process and Regularity Theories

Compared with interventionist accounts of causation, process and regularity accounts tend to be much more physics-oriented. I will therefore not go into much detail on these accounts. Nevertheless, since interventionism also faces various challenges, I think we should not close the book on these theories yet. Below, I make a few remarks about how process and regularity theories may be adapted to the grit case.

Process theories of causation hold that causes produce their effects through causal processes. In the context of physics, this is usually cashed out in terms of the transference of some conserved quantity from the cause to the effect (e.g., Dowe, 1992). As such, the account clearly would not apply to the grit case. However, if we are willing to broaden the notion of causal processes, then, process theories may turn out to be surprisingly suitable to personality psychology.

The trait theory of personality has, by and large, emerged victorious from the person-situation debate (Funder, 2001; Moreau et al., 2019). That is, psychologists now generally accept personality traits as genuine causes of behaviour, even if situational factors and acute experimental manipulations also play a role. Although theorists disagree over the precise nature of personality, it seems reasonable to expect that any satisfactory account of personality would include a story about how personality-defining experiences will come to affect personality-specific behaviours. Given the difficulty associated with intervening on personality, a process account of causation with an appropriate broadening of the notion "process" may be more descriptively apt for personality psychologists.

Process theories are often considered as more metaphysically (and epistemically) demanding than interventionism, whereas regularity theories are typically considered as the least demanding. At its minimum, Humean regularity theory requires only that there be constant conjunction between cause and effect in the past, with no expectation that it would continue in the future. More substantively, we may require that the constant conjunction is a true universal generalization, or that it is governed by a law of nature, or that we are justified to infer the occurrence of the effect from observing the cause (Andreas & Guenther, 2021).

Perhaps surprisingly, the allegedly-minimalistic regularity accounts have a much harder time than the allegedly-demanding process accounts when it comes to adaptability to the grit case. As discussed above, trait theorists are often quite resistant to the stimulus-response view of causation. This

should not surprise us if we consider the fact that personality traits are theorized as more than their manifest behaviour – two people with the same personality trait may exhibit different behaviours in similar situations, a phenomenon called state-trait interaction. Consequently, trait psychologists are much more hesitant to take on the demand that personality traits must manifest in universal stimulus-behaviour pairs, as would be required by regularity accounts.

What this tells us is that the standard philosophical approach of understanding causation as, first and foremost, a relationship between two events, is ill-suited to the case of personality psychology. Nevertheless, not all aspects of the regularity theories are irrelevant. Bowers (1973), for example, argues that personality traits should be understood as universal causal laws much in the same way gravity is, from which behavioural regularities can be derived. It seems that the regularity theory's notion of inferability of effect from the cause holds just as much importance in personality psychology.

4. Does Grit Explain?

While psychologists are extremely reluctant to use the word "cause" when it is clear to philosophers that causation is what's at issue, they are perfectly happy to use the word "explain" when it is much less clear to philosophers that explanation is what's being achieved. Time and time again, defenders of grit support its legitimacy by citing evidence that grit provides explanatory values over and above other predictors (Duckworth et al., 2007; Duckworth & Quinn, 2009). In this section, I will first describe the technical sense of "explanation" invoked in these claims. Next, I discuss in what sense, if any, we might apply philosophical theorization about explanation to this technical notion. Finally, I ask whether grit researchers are warranted to draw the kind of conclusions they often draw on account of explanatory claims about grit.

4.1 Explaining the Variance

Those unfamiliar with the statistical reasoning behind these explanatory claims may be surprised to hear that researchers often use "explaining" interchangeably with "capturing" and "accounting for". Here, I briefly describe what is typically meant by a claim such as "[grit] explain[s] variance in GPA over and beyond that explained by intelligence" (Duckworth et al., 2007).

Suppose I have a dataset consisting of a group of students' GPAs, their intelligence scores as measured by some intelligence test, and their grit scores as measured by the Grit Scale. The explanandum is GPA, which we denote as Y. The possible explanans are X_1 = intelligence and X_2 = grit.

More precisely, it is not GPA *per se* that is being explained, but its variation. Let $\bar{Y}$ be the mean GPA of my sample. Most, if not all, data points Y_i will be different from $\bar{Y}$ because students' GPAs differ from each other. We can fit a degenerate model (M_0) $Y_i = \alpha + \varepsilon_i$ where the best-fitted value for α is $\bar{Y}$. In other words, we "predict" each student's GPA by citing the group average. Here, the variance of the error term ε is the same as the variance of the explanandum variable Y. No explanation is achieved.

We can improve upon this model by adding an explanatory variable, X_1. Assuming the relationship is linear, we can fit another model (M_1) $Y_i = \alpha + \beta_1 X_{1i} + \varepsilon_i$ where X_{1i} is the intelligence score of the i^{th} student. Every student's GPA can be predicted by multiplying their intelligence score by β_1 and adding α to that value. Unless Y and X_1 are perfectly correlated, this prediction will still make errors, which are collected in the residual error term ε. However, if Y and X_1 are correlated at all, the variance of ε in (M_1) will be smaller than that of (M_0). If the improvement is large enough, psychologists say that X_1 has explanatory value.

We can attempt to improve the situation even further by fitting (M_2) $Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$, resulting in even greater reduction in variance of the residual term ε. If the improvement is large enough when compared with (M_1), we say that X_2 has explanatory value beyond X_1.

This is, indeed, quite a specific and technical sense of explanation. Yet when (quantitative) social scientists say that one attribute explains another, they almost always mean it in this technical sense. Before trying to apply philosophical theories to this subject, I first highlight two interesting features of this technical sense of explanation.

First, when X_1 and X_2 are highly correlated, it is often (though not always) the case that neither X_1 nor X_2 has explanatory value beyond the other. In the extreme case where X_1 and X_2 perfectly correlate, having both variables in a model would result in no improvement over having only one. This means that the claim "grit adds no additional explanatory value beyond conscientiousness" should not be confused with the claim "grit has no explanatory value". Indeed, it may also be apt to say that conscientiousness adds no additional explanatory value beyond grit.

Second, although it is tempting to take "explaining the variance" as nothing more than simply "improving prediction", the statistical theory behind this approach is actually quite substantive. The

data modeling paradigm starts with the assumption that the data at hand is produced by a data generating process, which takes the form of a statistical model (such as a linear equation) plus some random errors. The goal of the researcher is to identify this true model and estimate its parameters.

A lot can be said about the unrealistic assumptions and idealizations that go into the data modeling story (see, famously, Breiman, 2001). However, theoretically, the variables retained in the true model are there because of the roles they play in the data generating process, which exists in the world. It makes sense to say that the variables responsible for the generation of the present data *explain* the data, in a way that is much more substantive than the simple claim that the variables are retained because they are predictively useful.

4.2 In What Sense Might Grit Explain?

Given the heavy reliance on statistical modeling, it seems natural to start with the Statistical Relevance (SR) model of causation (Salmon, 1971, 1984). Indeed, the SR model was designed quite explicitly with social scientific compatibility in mind and often invokes examples from the social world. Nevertheless, the sense in which the SR model is "statistical" is not quite the same as the sense described above. The SR model is statistical in the sense that it is probabilistic – C can be explanatorily relevant to E without being necessary or sufficient in E's occurrence. The Inductive Statistical (IS) model of explanation, which I will not go into much detail, is similarly statistical in the probabilistic sense.

In the context of linear regression modeling, it is sometimes the case that the response variable *Y* is a probability, such as when studying whether grittier Marine recruits are less likely to drop out (Dijksma et al., 2022). In this case, grit level was not found to change the probability of dropout, and hence was not taken to hold explanatory value, just as what the SR model would say. However, when studying the relationship between grit and performance at the National Spelling Bee (Duckworth et al., 2011), the response variable is the ordinal ranking of participants' spelling performance, which is not a probability.

It is not that the grit example cannot be made to conform to the SR or the IS theories – we may say of a subject who scores second place at the Spelling Bee that the probability of them scoring first place is very high – but doing so would neither capture the sense of "explanation" invoked to justify the construct legitimacy of grit nor be of much help in grit reasoning. One reason for this mismatch may be that the explanatory target in the grit case is not a particular behaviour, but that behaviour

when compared with other people's behaviours. That is, what we are explaining is not that Ada scored high at the spelling competition, but that Ada scored higher than all of her competitors.

This way of framing the explanandum is reminiscent of the contrastive theory of explanation (Van Fraassen, 1980; Lipton, 1991; Khalifa, 2010). In Lipton's words, "[t]o explain why Kate rather than Frank won the prize, it is not enough that she wrote a good essay; it must have been better than Frank's" (1991, p.689). The situation with explaining the variance is not exactly the same, but there may be sufficient parallels for fruitful application.

The three authors mentioned above approach contrastive explanations in quite different ways. Van Fraassen famously rejects the view that it is a feature of scientific theories that they explain. Instead, explanations, for van Fraassen, are answers to "why" questions, and they are successful just in case the questioner is satisfied with the answer. This means that what constitutes a satisfactory answer will vary depending on context.

Although van Fraassen's (1980) pragmatic account of explanation is meant to occur outside of scientific theorizing, it fits surprisingly well with the grit example. Colloquially, personality traits are often considered as adequate answers to why questions about behaviour. Many people would accept "Ada works very hard because she is perseverant" as an adequate explanation, even though it is, in some sense, circular. Similarly, we may answer the question "why did Kate score two standard deviations above Frank?" by citing Kate's grit score as compared to Frank's, which seems to be a good description of what grit researchers are doing when they say that grit explains variance in spelling performance.

In quite a different vein, Lipton (1991) takes contrastive explanations to be attempts at specifying particular causal information that interests us. I will not go into much detail here because I have already discussed causation in the previous section.

More recently, Khalifa (2010) proposed a non-causal account of explanation inspired by theorizations about obligations and commitments, called accountabilism. The idea is that a claim p *rather than q* is an appropriate explanandum just in case the circumstances that lead one to believe p could have easily led someone to believe q instead. The person making the claim therefore owes it to listeners to explain why they should not conclude q. The claimant then satisfies the demand for explanation by offering an account for choosing p but not q.

Khalifa's requirement that p and q constitute an appropriate contrast class only if the circumstances that lead one to believe p would have led one to believe q is quite interesting, and seems to capture something important about the way researchers reason about variance capture. Recall that explanatory values of variables are measured in comparison with the degenerate model (M_0) $Y_i = \alpha + \varepsilon_i$, where the prediction is just the sample mean. Given only the sample mean, we may reasonably ask "why should I expect Kate to score higher than Frank, given that they belong to the same sample, with sample mean $\bar{Y}$?" The researcher then provides an account by adding an explanatory variable and pointing to the improvement in fitness.

4.3 Inference to the Best Explanation

One reason I have not been discussing accounts of causal explanations is that, in both the grit example specifically and social psychology at large, explanatory claims are typically taken as evidence for causal claims. In fact, explanatory claims are often taken as the first line of defence in the entire enterprise.

Recall that the motivating question for grit research is a why question: "Why do some individuals accomplish more than others of equal intelligence?" (Duckworth et al., 2007) After six studies, the answer is reported as follows: "individual differences in grit accounted for significant incremental variance in success outcomes over and beyond that explained by IQ, to which it was not positively related" (Duckworth et al., 2007). That is, grit is a legitimate construct *because* it explains success beyond IQ.

Can philosophical theories of explanation justify the legitimacy of the invoked explanans? When framed in this way, philosophers can immediately recognize this question as one about realism through inference to the best explanation (IBE). The "no-miracle" argument for scientific realism (Putnam, 1975a; Musgrave, 1988) holds that the best explanation for the predictive success of scientific theories is that they are (approximately) true, and that their theoretical terms refer. The indispensability argument for mathematical realism (Putnam, 1975b; Maddy, 1992; Colyvan, 2010) has similarly relied on the premise that explanatory power entails realism.

Might the same arguments be adapted to the case of grit research? One challenge is that, in the philosophy of science and of mathematics, IBE arguments tend to be used to support realist attitudes in broad strokes, rather than in individual cases. That is, the success of science in general should warrant a realist attitude towards the existence of unobservable entities, but the particular

(and sometimes temporary) predictive edge of Ptolemaic epicycles and Copernican elegance should not be taken too seriously. In fact, the ability to resist overemphasizing an immediate predictive edge is exactly what is attractive about the realist perspective for some (e.g., Forster & Sober, 1994; Hitchcock & Sober, 2004).

What this means is that, while we may have good IBE reasons to believe that personality traits exist because they successfully explain, the same line of reasoning will not automatically extend to the comparison of one trait theory over another. Again, this seems to be an area where philosophers of science can make genuine contributions to psychological theorizing.

5. Counterfactual Reasoning About Grit

Some studies in grit employ randomized controlled trials (RCTs) (e.g., Leppin et al., 2014), which are taken as the gold standard for establishing causal claims. Most studies, however, are association studies, comparing people who measure high on grit with people who measure low. This is not uncommon in the social sciences and medicine, as RCTs are not always possible or even desirable (see, e.g., Cartwright, 2007). What is clear is that grit intervention programs rely on counterfactual claims about grit. It is therefore important that we assess the nature and quality of the counterfactual evidence.

The most popular semantics for counterfactuals is the Lewis-Stalnaker possible world semantics (Stalnaker, 1968; Lewis, 1973). Very roughly, the account says that the counterfactual "if A were to occur, B would occur" is true just in case, in the nearest possible world in which A occurs, B also occurs. The benefit of this approach is that it reduces counterfactuals to truth-functional classical logic. The challenge is to give an account of "nearness" of worlds that does not presuppose counterfactual knowledge. While the metaphysics of possible worlds is a thorny issue, it turns out that counterpart theory is surprisingly well suited to the case of personality psychology.

One important motivation for Lewisian possible worlds is the prospect of providing a Humean account of modality. That is, instead of treating modality as a metaphysically loaded properties that distinguish "essential" from "accidental" attributes, the Humean account reduces all modal claims to factual claims about a metaphysically "flat" Humean mosaic. When we say that Frank could have won the award instead of Kate, we are making a factual claim about what happens in a nearby possible world, rather than about essential or nonessential properties of Kate and Frank.

One challenge with this account is how to know who is Kate in that possible world, since she differs from the real Kate in her award winning status. According to Lewis's counterpart theory, the person Kate refers only to Kate in the real world and nowhere else. But the real Kate has counterparts in other worlds, who are the bearers of counterfactual claims about the real Kate. To figure out who is Kate's counterpart, Lewis suggests that we look at qualitative similarities between Kate and her candidate counterparts, and he allows similarity judgments to differ across contexts.

Unless one is already committed to a Humean metaphysics, Lewisian counterpart theory appears way too strange to be plausible. For one, a counterfactual claim about Kate turns out not to be a claim about Kate at all, but about another person who is "qualitatively similar" to Kate. Second, what does it mean for two people, who necessarily differ in one aspect, to be qualitatively similar? What if observers disagree over which candidate counterpart is more similar than the other? For those of us who are already on board with robust causal claims, counterpart theory may appear unnecessarily complicated for unclear theoretical gain. However, there is a surprising sense in which counterpart theory can be fruitfully adapted to counterfactual reasoning about personality traits.

Consider, again, the combined claim that personality traits cause behaviours and that personality traits are unmanipulable. What might be the evidence for the first claim, if not manipulation evidence? As discussed in the previous section, the evidence is comparisons of individuals who share or do not share personality traits. In other words, the claim is that grit is the difference maker because those who are similar in grit behave similarly, and those who are not, do not.

The same reasoning can also be observed in everyday discussions about personality. Imagine you are deciding whether you will go to an event, and two of your friends, Ada and Bri, give opposite recommendations – Ada says she has gone to a previous iteration of the event and loved it, whereas Bri says she has gone to the same event and hated it. Personality traits, for those who believe in them, are exactly the kind of thing invoked to settle disputes like this. If you are more like Ada, you take Ada's advice, even if she does not provide any reason for her prediction of your enjoyment.

When a Lewisian claims that the statement "Kate could have lost the competition" is about what happens to a counterpart who is very similar to Kate in relevant ways, he is often met with skepticism over whether the person he identifies is indeed Kate's counterpart, or where could he possibly have gotten this trans-possible-world knowledge from. My goal here is not to make progress on these difficult philosophical questions. However, insofar as it is possible for the Lewisian

to provide an answer without referring to some essential connection between Kate and her counterpart, the lesson may be fruitfully exported to counterfactual reasoning about personality traits. That is, when a personality psychologist claims "if Kate were not as gritty, she would not have won the competition" based on evidence obtained from other people who share Kate's level of intelligence, skeptics may likewise object that these apparent similarities are insufficient to justify this counterfactual claim, and that there should be some essential connection between grit-similarities and behaviour-similarities. The psychologist should be allowed, just as the Lewisian is, to decline this request.

6. Conclusion

My main goal in writing this paper was to convince philosophers of science that there is value in making a genuine effort at applying philosophical theories to the social world that goes beyond the typical caveats around indeterministic events or the possibility of social laws. In pursuit of this goal, I posited grit psychology as a test case and treated it in much of the same way that philosophers of science have been treating physics examples. I aimed to cover a number of different topics as a way to show the breadth of my claim. Unfortunately, this means that I had to gloss over many interesting details and important subtleties of these philosophical views. Similarly, while it is reasonable to take the grit case as exemplary of many debates in social psychology, education, microeconomics, management science, or even medicine, it is clearly only one of many social-scientific contexts that would benefit from increased philosophical attention. I hope future work in the philosophy of social science can address both kinds of limitations.

Nevertheless, I take my analysis to support several preliminary observations. First, the usual judgment of which theories have "substantive" or "minimalistic" commitments does not always hold when applied to the grit example. For example, regularity theories of causation are typically considered too minimalistic to count as genuine causation, while process theories are often criticized as too substantive to be plausible. Yet, as discussed in section 3, the amount of regularity typically assumed by regularity theories is in fact much more demanding than giving a satisfactory account of causal processes in the case of personality psychology.

Second, many philosophical accounts of causation, explanation, and counterfactuals are based on the assumption that there is a scientific consensus on what causes what, and that the job of the

philosopher is to explicate this consensus. This attitude has somewhat changed recently, with an increasing amount of philosophical theorization aiming to be both descriptively adequate and prescriptively useful for science (see Ruiz & Schulz, 2023, for an example in economics and Li, forthcoming, for an example in climate science). Insofar as grit psychology is an active area of debate, it would benefit from more prescriptively focused philosophy of science.

Lastly, I would like to point out the extent to which scientists engage in philosophical argumentation. Granted, there are important empirical and methodological issues surrounding grit research that are perhaps better left for scientists and statisticians to critique. Yet, as I highlighted throughout the present paper, there are also substantive philosophical themes at stake. Issues such as realism through IBE and counterfactual claims without manipulability are topics that philosophers have spent a lot of time debating and clarifying. Both philosophy and psychology would benefit from a kind of philosophy of science that takes the social sciences seriously.

References

Allport, G. W. (1931). What is a trait of personality? *Journal of Abnormal and Social Psychology*, *25*, 368–372.

Andreas, H., & Guenther, M. (2021). Regularity and Inferential Theories of Causation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entries/causation-regularity/

Bowers, K. (1973). Situationism in psychology: An analysis and a critique. *Psychological Review*, *80*, 307–336.

Bowman, N. A., Miller, A., Woosley, S., Maxwell, N. P., & Kolze, M. J. (2019). Understanding the Link Between Noncognitive Attributes and College Retention. *Research in Higher Education*, *60*(2), 135–152. https://doi.org/10.1007/s11162-018-9508-0

Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, *16*(3), 199–215.

Cartwright, N. (2002). Against Modularity, the Causal Markov Condition, and Any Link between the Two: Comments on Hausman and Woodward. *The British Journal for the Philosophy of Science*, *53*(3), 411–453.

Cartwright, N. (2006). From Metaphysics to Method: Comments on Manipulability and the Causal Markov Condition. *The British Journal for the Philosophy of Science*, *57*(1), 197–218.

Cartwright, N. (2007). Are RCTs the gold standard? *BioSocieties*, *2*(1), 11–20.

Colyvan, M. (2010). There is No Easy Road to Nominalism. *Mind*, *119*(474), 285–306.

Credé, M. (2018). What Shall We Do About Grit? A Critical Review of What We Know and What We Don't Know. *Educational Researcher*, *47*(9), 606–611. https://doi.org/10.3102/0013189X18801322

Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*, *113*(3), 492–511. https://doi.org/10.1037/pspp0000102

Dijksma, I., Lucas, C., & Stuiver, M. (2022). Grit was not associated to dropout in Dutch Marine recruits. *Military Psychology*, *34*(5), 616–621. https://doi.org/10.1080/08995605.2022.2028518

DiMenichi, B. C., & Richmond, L. L. (2015). Reflecting on past failures leads to increased perseverance and sustained attention. *Journal of Cognitive Psychology*, *27*(2), 180–193. https://doi.org/10.1080/20445911.2014.995104

Direito, I., Chance, S., & Malik, M. (2021). The study of grit in engineering education research: A systematic literature review. *European Journal of Engineering Education*, *46*(2), 161–185. https://doi.org/10.1080/03043797.2019.1688256

Dowe, P. (1992). Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory. *Philosophy of Science*, *59*(2), 195–216.

Duckworth, A. L., & Gross, J. J. (2014). Self-Control and Grit: Related but Separable Determinants of Success. *Current Directions in Psychological Science*, *23*(5), 319–325. https://doi.org/10.1177/0963721414541462

Duckworth, A. L., Kirby, T. A., Tsukayama, E., Berstein, H., & Ericsson, K. A. (2011). Deliberate Practice Spells Success: Why Grittier Competitors Triumph at the National Spelling Bee. *Social Psychological and Personality Science*, *2*(2), 174–181. https://doi.org/10.1177/1948550610385872

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Duckworth, A. L., & Quinn, P. D. (2009). Development and Validation of the Short Grit Scale (Grit–S). *Journal of Personality Assessment*, *91*(2), 166–174. https://doi.org/10.1080/00223890802634290

Eberhardt, F. (2013). Experimental Indistinguishability of Causal Structures. *Philosophy of Science*, *80*(5), 684–696. https://doi.org/10.1086/673865

Eberhardt, F., Glymour, C., & Scheines, R. (2005). On the Number of Experiments Sufficient and in the Worst Case Necessary to Identify All Causal Relations Among N Variables. In F. Bacchus & T. Jaakkola (Eds.), *Proceedings of the 21st Conference on Uncertainty and Artificial Intelligence* (pp. 178–184).

Forbes, S., & Fikretoglu, D. (2018). Building Resilience: The Conceptual Basis and Research Evidence for Resilience Training Programs. *Review of General Psychology*, *22*(4), 452–468. https://doi.org/10.1037/gpr0000152

Forster, M., & Sober, E. (1994). How to Tell When Simpler, More Unified, or Less *Ad Hoc* Theories will Provide More Accurate Predictions. *The British Journal for the Philosophy of Science*, *45*(1), 1–35. https://doi.org/10.1093/bjps/45.1.1

Funder, D. C. (2001). Personality. *Annual Review of Psychology*, *52*, 197–221.

Guo, J., Tang, X., & Xu, K. M. (2019). Capturing the multiplicative effect of perseverance and passion: Measurement issues of combining two grit facets. *Proceedings of the National Academy of Sciences*, *116*(10), 3938–3940. https://doi.org/10.1073/pnas.1820125116

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of Learning Skills Interventions on Student Learning: A Meta-Analysis. *Review OfEducational Research*, *66*(2), 99–136.

Hausman, D. M., & Woodward, J. (1999). Independence, Invariance and the Causal Markov Condition. *The British Journal for the Philosophy of Science*, *50*(4), 521–583.

Hicks, R. (2010). Do Time Management, Grit, and Self-Control Relate to Academic Achievement Independently of Conscientiousness? In *Personality and Individual Differences* (pp. 79–90). Australian Academic Press.

Hitchcock, C. (2009). Causal Modelling. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *The Oxford Handbook of Causation* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199279739.003.0015

Hitchcock, C. (2023). Causal Models. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2023/entries/causal-models/

Hitchcock, C., & Sober, E. (2004). Prediction Versus Accommodation and the Risk of Overfitting. *The British Journal for the Philosophy of Science*, *55*(1), 1–34. https://doi.org/10.1093/bjps/55.1.1

Jachimowicz, J. M., Wihler, A., Bailey, E. R., & Galinsky, A. D. (2018). Why grit requires perseverance and passion to positively predict performance. *Proceedings of the National Academy of Sciences*, *115*(40), 9980–9985. https://doi.org/10.1073/pnas.1803561115

Johnson, J. A. (1997). Units of analysis for the description and explanation of personality. In *Handbook of personality psychology* (pp. 73–93). Elsevier.

Khalifa, K. (2010). Contrastive Explanations as Social Accounts. *Social Epistemology*, *24*(4), 263–284. https://doi.org/10.1080/02691728.2010.506960

Leppin, A. L., Bora, P. R., Tilburt, J. C., Gionfriddo, M. R., Zeballos-Palacios, C., Dulohery, M. M., Sood, A., Erwin, P. J., Brito, J. P., Boehmer, K. R., & Montori, V. M. (2014). The Efficacy of Resiliency Training Programs: A Systematic Review and Meta-Analysis of Randomized Trials. *PLoS ONE*, *9*(10), e111420. https://doi.org/10.1371/journal.pone.0111420

Lewis, D. (1973). Counterfactuals and comparative possibility. *IFS: Conditionals, Belief, Decision, Chance and Time*, 57–85.

Li, D. (forthcoming). Machines Learn Better with Better Data Ontology. *Minds & Machines*.

Lipton, P. (1991). Contrastive Explanation and Causal Triangulation. *Philosophy of Science*, *58*(4), 687–697.

Maddy, P. (1992). Indispensability and Practice. *The Journal of Philosophy*, *89*(6), 275–289. https://doi.org/10.2307/2026712

Menzies, P., & Price, H. (1993). Causation as a Secondary Quality. *The British Journal for the Philosophy of Science*, *44*(2), 187–203. https://doi.org/10.1093/bjps/44.2.187

Meriac, J. P., Slifka, J. S., & LaBat, L. R. (2015). Work ethic and grit: An examination of empirical redundancy. *Personality and Individual Differences*, *86*, 401–405. https://doi.org/10.1016/j.paid.2015.07.009

Messick, S. (1979). Potential Uses of Noncognitive Measurement in Education. *Journal of Educational Psychology*, *71*(3), 281–292.

Moreau, D., Macnamara, B. N., & Hambrick, D. Z. (2019). Overstating the Role of Environmental Factors in Success: A Cautionary Note. *Current Directions in Psychological Science*, *28*(1), 28–33. https://doi.org/10.1177/0963721418797300

Mueller, B. A., Wolfe, M. T., & Syed, I. (2017). Passion and grit: An exploration of the pathways leading to venture success. *Journal of Business Venturing*, *32*(3), 260–279. https://doi.org/10.1016/j.jbusvent.2017.02.001

Musgrave, A. (1988). The Ultimate Argument for Scientific Realism. In R. Nola (Ed.), *Relativism and Realism in Science* (pp. 229–252). Springer Netherlands. https://doi.org/10.1007/978-94-009-2877-0_10

Pearl, J. (2009). *Causality*. Cambridge university press.

Pendyala, A. G., & Vyas, M. (2023). Moving away from the WEIRD Grit: A Critical Review. *Indian Journal of Positive Psychology*, *14*(1), 67–71.

Ponnock, A., Muenks, K., Morell, M., Seung Yang, J., Gladstone, J. R., & Wigfield, A. (2020). Grit and conscientiousness: Another jangle fallacy. *Journal of Research in Personality*, *89*, 104021. https://doi.org/10.1016/j.jrp.2020.104021

Postigo, Á., Cuesta, M., García-Cueto, E., Menéndez-Aller, Á., González-Nuevo, C., & Muñiz, J. (2021). Grit Assessment: Is One Dimension Enough? *Journal of Personality Assessment*, *103*(6), 786–796. https://doi.org/10.1080/00223891.2020.1848853

Putnam, H. (1975a). *Mathematics, Matter and Method* (Vol. 1). Cambridge University Press.

Putnam, H. (1975b). What is mathematical truth? *Historia Mathematica*, *2*(4), 529–533.

Rimfeld, K., Kovas, Y., Dale, P. S., & Plomin, R. (2016). True grit and genetics: Predicting academic achievement from personality. *Journal of Personality and Social Psychology*, *111*(5), 780–789. https://doi.org/10.1037/pspp0000089

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin*, *143*(2), 117–141. https://doi.org/10.1037/bul0000088.supp

Ruiz, N., & Schulz, A. W. (2023). Micro-foundations and Methodology: A Complexity-Based Reconceptualization of the Debate. *The British Journal for the Philosophy of Science*, *74*(2), 000–000.

Ryans, D. G. (1939). The measurement of persistence: An historical review. *Psychological Bulletin*, *36*(9), 715–739. https://doi.org/10.1037/h0060780

Salmon, W. C. (1971). *Statistical explanation and statistical relevance* (Vol. 69). University of Pittsburgh Pre.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.

Schmidt, F. T. C., Nagy, G., Fleckenstein, J., Möller, J., & Retelsdorf, J. (2018). Same Same, but Different? Relations between Facets of Conscientiousness and Grit. *European Journal of Personality*, *32*(6), 705–720. https://doi.org/10.1002/per.2171

Shechtman, N., DeBarger, A. H., Dornsife, C., Rosier, S., & Yarnall, L. (2013). Promoting grit, tenacity, and perseverance: Critical factors for success in the 21st century. *Washington, DC: US Department of Education, Department of Educational Technology*, *1*, 1–107.

Sober, E., & Steel, M. (2013). Screening-Off and Causal Incompleteness: A No-Go Theorem. *The British Journal for the Philosophy of Science*, *64*(3), 513–550. https://doi.org/10.1093/bjps/axs021

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, Prediction, and Search* (Vol. 81). Springer New York. https://doi.org/10.1007/978-1-4612-2748-9

Stalnaker, R. C. (1968). A theory of conditionals. In W. L. Harper, R. C. Stalnaker, & G. Pearce (Eds.), *Ifs: Conditionals, belief, decision, chance and time* (pp. 41–55). Springer.

Steel, D. (2005). Indeterminism and the Causal Markov Condition. *The British Journal for the Philosophy of Science*, *56*(1), 3–26. https://doi.org/10.1093/phisci/axi101

Strayhorn, T. L. (2014). What Role Does Grit Play in the Academic Success of Black Male Collegians at Predominantly White Institutions? *Journal of African American Studies*, *18*(1), 1–10. https://doi.org/10.1007/s12111-012-9243-0

Suzuki, Y., Tamesue, D., Asahi, K., & Ishikawa, Y. (2015). Grit and Work Engagement: A Cross-Sectional Study. *PLOS ONE*, *10*(9), e0137501. https://doi.org/10.1371/journal.pone.0137501

Tang, X., Wang, M.-T., Parada, F., & Salmela-Aro, K. (2021). Putting the Goal Back into Grit: Academic Goal Commitment, Grit, and Academic Achievement. *Journal of Youth and Adolescence*, *50*(3), 470–484. https://doi.org/10.1007/s10964-020-01348-1

Tyumeneva, Y., Kardanova, E., & Kuzmina, J. (2019). Grit: Two Related but Independent Constructs Instead of One. Evidence From Item Response Theory. *European Journal of Psychological Assessment*, *35*(4), 469–478. https://doi.org/10.1027/1015-5759/a000424

Van Fraassen, B. C. (1980). *The scientific image*. Oxford University Press.

Walker, A., Hines, J., & Brecknell, J. (2016). Survival of the Grittiest? Consultant Surgeons Are Significantly Grittier Than Their Junior Trainees. *Journal of Surgical Education*, *73*(4), 730–734. https://doi.org/10.1016/j.jsurg.2016.01.012

Weinberger, N., Williams, P., & Woodward, J. (n.d.). *The Worldly Infrastructure of Causation*.

Woodward, J. (2023). Causation and Manipulability. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Summer 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2023/entries/causation-mani/

Woodward, J., & Woodward, J. F. (2005). *Making things happen: A theory of causal explanation*. Oxford university press.

Zhao, K. (2023). Measuring the Nonexistent: Validity before Measurement. *Philosophy of Science*, *90*(2), 227–244. https://doi.org/10.1017/psa.2023.3