High-Fidelity Experiments, Situationism, and the Measurement of Virtue

Matthew Wilson[1,2]

Mark Alfano proposes an interesting distinction between high-fidelity and low-fidelity virtues.[1] The distinction purports to show that some virtues are more susceptible to empirical *disconfirmation* than others. In the context of Alfano's larger project, this distinction is consequential for his argument in favor of situationism, which, roughly, is the thesis that cross-situationally consistent or global virtues do not exist. But the distinction is also potentially important for those who think global virtues *do* exist, because it pertains to the question of how social science may attempt to measure virtue. This question is of significance to educators, politicians, philosophers, and even public health officials.[2]

In this essay, I argue that Alfano's proposed distinction is a distinction without a difference *with respect to virtues*, but I reconstruct the distinction as one that holds between types of *experiments* that measure for virtue. High-fidelity experiments, I propose, are those that can provide strong evidence against the presence of virtue in test subjects, merely through the observation of a single behavior that is contrary to a virtue. Low-fidelity experiments do not provide disconfirming evidence in this way. My proposed distinction between high-fidelity and low-fidelity experiments is important for two reasons. First, it shows that it is incorrect to think that one-time behavioral experiments cannot, in principle, provide strong disconfirming evidence of a virtue. This has been a complaint of some who have argued against situationism.[3] Although high-fidelity experiments may be difficult to engineer, they are not

---

[1] See Mark Alfano, *Character as Moral Fiction* (New York: Cambridge University Press, 2013).

[2] This is a recent development in public health. See, for example, Tyler J. VanderWeele, "On the Promotion of Human Flourishing," *Proceedings of the National Academy of Sciences of the United States of America* 114, no. 31 (2017): 8148–56.

[3] See Gopal Sreenivasan, "Character and Consistency: Still More Errors," *Mind* 117, no. 467 (2008): 607.

✉ Matthew Wilson
    mfwilson@jbu.edu; mfwilson@fas.harvard.edu

[1]   John Brown University, 2000 West University Street, Siloam Springs, AR 72761, USA

[2]   Human Flourishing Program, Institute for Quantitative Social Science, Harvard University, 129 Mt. Auburn St., Ste 1, Cambridge, MA 02138, USA

impossible, and this essay sketches out an example of what a high-fidelity experiment might look like, at least for the virtue of justice.

Second, the discussion illuminates further why most social psychology experiments cited by philosophers in the situationist debate fail to be disconfirming of virtue, *even for control groups*. Most of these one-off or cross-sectional experiments are low-fidelity because they fail to meet the desiderata of a high-fidelity experiment, which I outline in the second half of the essay. I recommend that empirical scientists take these desiderata into account as they attempt to more rigorously measure and test for the presence and absence of virtue.

1 Background

Over the past twenty years, a growing number of philosophers have concluded that social psychology experiments provide strong evidence that virtues do not exist in most human beings, at least not as virtues were conceived by Plato in his *Republic* or Aristotle in his *Nicomachean Ethics*.[4] These experiments are taken together to suggest that human behavior is mostly the result of situational factors rather than the stable, cross-situationally consistent character traits that were once thought to make up virtue.

Take the virtue of compassion, for example. Philosophers John Doris, Gilbert Harman, and Mark Alfano all cite the Milgram shock experiments, the Darley and Batson Princeton Theological Seminary hurry study, the Latané and Darley group studies, and the Isen and Levin dime phone booth studies as evidence that most people do not possess the virtue of compassion.[5] They claim that these experiments show us that the presence of an authority or bystanders, being in a hurry, or being in a positive mood are what determine the experimental subjects' behavior in each case. And compassion is not the only virtue under attack. Christian Miller provides a lengthy discussion of dozens of more recent experiments, which show that guilt,

---

[4] See John Cooper and D. S. Hutchinson, eds., *Plato: Complete Works* (Indianapolis: Hackett Publishing Company, 1997). Aristotle, *Nicomachean Ethics*, ed. and trans. Roger Crisp (New York: Cambridge University Press, 2000).

[5] See John Doris, "Persons, Situations, and Virtue Ethics," *Nous* 32, no. 4 (1998): 504–30; John Doris, *Lack of Character: Personality and Moral Behavior* (Cambridge: Cambridge University Press, 2002); John Doris, "Heated Agreement: Lack of Character as Being for the Good," *Philosophical Studies* 148 (2010): 135–46. Gilbert Harman, "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error," *Proceedings of the Aristotelian Society*, 1999, 315–31; Gilbert Harman, "The Nonexistence of Character Traits," *Proceedings of the Aristotelian Society* 100 (2000): 223–26. Alfano, *Character as Moral Fiction*. Stanley Milgram, *Obedience to Authority* (New York: Harper, 1974). John M. Darley and Daniel C. Batson, "'From Jerusalem to Jericho': A Study of Situational and Dispositional Variables in Helping Behavior," *Journal of Personality and Social Psychology* 27, no. 1 (1973): 100–108. Latane, Bibb and Darley, John M., "Group Inhibition of Bystander Intervention in Emergencies," *Journal of Personality and Social Psychology* 10, no. 3 (1968): 215–21; Bibb Latane and Darley, John M., *The Unresponsive Bystander* (New York: Appleton-Century Crofts, 1970). Alice Isen and Paul Levin, "Effect of Feeling Good on Helping: Cookies and Kindness," *Journal of Personality and Social Psychology* 21, no. 3 (1972): 384–88; Paul Levin and Alice Isen, "Further Studies on the Effect of Feeling Good on Helping," *Sociometry* 38, no. 1 (1975): 141–47.

embarrassment, negligible financial gains, personal introductions by strangers, grating noises and other seemingly trivial situational differences have significant effects on behavior.[6] He too concludes that few people possess the virtue of compassion, and he adds honesty and non-maleficence to the list.[7]

This situationist interpretation of the data is, of course, not uncontested. Sreenivasan, Kupperman, and Sabini and Silver all deny the situationist's claims for a variety of conceptual and methodological reasons.[8] Other philosophers like Adams and Snow, acquiesce to the situationist's conclusion that, at present, we only have empirical support for so-called local virtues.[9] But they argue for the possibility that people could develop global virtues from local virtue building blocks and cite their own experiments as support.

An important principle at stake in these debates is the permissibility of drawing conclusions about human *dispositions* from one-time behavioral experiments. Many philosophers agree that virtues are not *exceptionless* dispositions to act, think, and feel; they are only *reliable* dispositions to act, think, and feel.[10] It is therefore difficult to conclude from a single behavioral performance (or non-performance) that a person is truly disposed (or indisposed) to act in some way. Thus, to infer that a person is reliably disposed to act in some way – especially in the case of moral dispositions that involve choice – one would want to observe the same person over a repeated number of behavioral instances. Unfortunately, few of the cited experiments observe subjects at more than one point in time.[11]

---

[6] See Christian Miller, *Moral Character: An Empirical Theory* (Oxford: Oxford University Press, 2013); Christian Miller, *Character and Moral Psychology* (Oxford: Oxford University Press, 2014).

[7] For those unfamiliar with this debate, situationist philosophers argue that there is not sufficient empirical support for the existence of traditional virtues that extend across multiple situations or domains of human life, in a so-called global way. Thus, this aspect of the anthropology of Plato and Aristotle is wrong. Virtues such as courage, generosity, justice, and temperance were supposed to be causally efficacious in most life situations, but the experiments purport to refute this. Doris, Alfano, Miller and Upton all allow for the possibility of local virtues, however. These traits are narrowly indexed to situations. They could be described by phrases such as *courage-at-sea*, *test-taking-honesty*, or *compassion-at-home*. But because local traits are indexed to situations, there are nearly an infinite number of local virtues that might exist, and thus folk-talk of courage in general, or honesty in general, is just not accurate, according to situationist philosophers. Miller's view is slightly different than some others because he thinks that people can possess cross-situationally consistent character traits, but these are what he calls "mixed-traits" and are unlike anything we are familiar with in folk psychology. Unpacking Miller's novel account would take too long and is not relevant for the argument of this paper.

[8] See Gopal Sreenivasan, "Errors about Errors: Virtue Theory and Trait Attribution," *Mind* 111, no. 441 (2002): 47–68; Sreenivasan, "Character and Consistency: Still More Errors"; Joel Kupperman, "Virtue in Virtue Ethics," *The Journal of Ethics* 13 (2009): 243–55; John Sabini and Maury Silver, "Lack of Character? Situationism Critiqued," *Ethics* 115 (2005): 535–62.

[9] See Robert Adams, *A Theory of Virtue* (Oxford: Clarendon Press, 2006); Nancy Snow, *Virtue as Social Intelligence: An Empirically Grounded Theory* (New York: Routledge, 2010).

[10] This characterization of virtue is widely accepted, even by situationists. See, e.g., Doris, *Lack of Character: Personality and Moral Behavior*.

[11] The only longitudinal study commonly cited that specifically tests for a moral disposition like a virtue is Hartshorne and May's study on cheating and honesty. See Nicole Tausch, Jared Kenworthy, and Miles Hewstone, "The Confirmability and Disconfirmability of Trait Concepts Revisited: Does Content Matter?", *Journal of Personality and Social Psychology* 92, no. 3 (2007): 542; Dwight Riskey and Michael Birnbaum, "Compensatory Effects in Moral Judgment: Two Rights Don't Make up for a Wrong," *Journal of Experimental Psychology* 103, no. 1 (1974): 171–73. Sreenivansan, however, provides arguments

One thing, then, on which philosophers on both sides agree is the need for more longitudinal studies.[12] Future studies designed to measure virtue ideally would record repeated observations of the *same* people over extended periods of time. But longitudinal studies can be difficult and expensive to conduct. Is there no type of cross-sectional study design that could test for the presence or absence of a virtue? It is unlikely that the observation of a single behavior could ever establish the presence of a virtue. But philosopher Gopal Sreenivasan has also argued that "data from a one-time performance experiment [cannot] do anything to establish" the *non-existence* of a character trait.[13] If this is true, then, in principle, cross-sectional social psychology experiments cannot confirm or disconfirm the presence of virtue in either control or experimental groups. Alfano has termed this type of argument the *longitudinal data critique* of situationism.

Alfano defends against the longitudinal data critique by making a distinction between high-fidelity and low-fidelity virtues. According to Alfano, high-fidelity virtues are more susceptible to empirical *disconfirmation* than low-fidelity virtues. Thus, he states that a single observation from a one-off experiment may "provide strong evidence *against* (though only weak evidence for) the presence of a [high-fidelity virtue]."[14] In other words, high-fidelity virtues are supposedly susceptible to empirical disconfirmation by one-off experiments in a way that low-fidelity virtues are not. Unfortunately, Alfano does not argue for his distinction or provide a theoretical account. Nor does he examine the empirical evidence that might be relevant to the distinction.[15] But he does claim that this difference between virtues exists, and

---

Footnote 11 (continued)

for thinking the study does not properly test for honesty, despite its longitudinal design. See Sreenivasan, "Errors about Errors: Virtue Theory and Trait Attribution." Experiments conducted by Walter Mischel and Yuichi Shoda observed children over a six-week period at summer camp for traits related to verbal aggression, withdrawal, friendly, and pro-social behavior. See Walter Mischel and Yuichi Shoda, "A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure," *Psychological Review* 102, no. 2 (1995): 246–68; Yuichi Shoda, Walter Mischel, and J. C. Wright, "Intuitive Interactionism in Person Perception: Effects of Situation-Behavior Relations on Dispositional Judgements," *Journal of Personality and Social Psychology* 56 (1989): 41–53. These experiments have been cited as evidence for the existence of stable "if…then…" patterns in human behavior (i.e. dispositions), but Mischel and Shoda were not obviously testing for traditional moral virtues.

[12] Two recent calls for this are by Upton and Miller. See Candice Upton, "The Empirical Argument Against Virtue," *Journal of Ethics* 20 (2016): 355–71; Christian Miller, "Character and Situationism: New Directions," *Ethical Theory and Moral Practice* 20 (2017): 459–71. The failure to collect longitudinal data (along with inadequate sample sizes) is also a reason why many experiments in social psychology have failed to replicate. The often-cited dime in phone booth experiment is one example. See Levin and Isen, "Further Studies on the Effect of Feeling Good on Helping"; James Weyant and Russell D. Clark, "Dimes and Helping: The Other Side of the Coin," *Personality and Social Psychology Bulletin* 3, no. 1 (1976): 107–10. Yet despite making note of the replication failure, Doris and Miller (among others) continue to cite the experiment as evidence against virtue in their writings.

[13] Sreenivasan, "Character and Consistency: Still More Errors," p. 607. A similar complaint is lodged by Kristjansson Kristjan Kristjansson, "An Aristotelian Critique of Situationism," *Philosophy* 83 (2008): 76.

[14] Alfano, *Character as Moral Fiction*, p. 72.

[15] Some psychology experiments have been conducted showing that humans have a tendency, empirically, to infer that a character trait may be "disconfirmed" from a single negative behavior; but this is not the case with positive actions. This tendency is therefore referred to as negativity bias. There is a body of literature studying how people perceive the probabilistic relationships between behavioral cues and trait

that the property of disconfirmability purports to be the essence of the difference. He motivates our intuitions for the distinction through the use of examples.

The high-fidelity side of the distinction he motivates by considering the virtues of chastity and temperance. In the case of chastity, it seems obvious that a single instance of sexual unfaithfulness provides strong evidence that one does not possess the virtue. In the case temperance, he claims that a single observation of someone on a bender is strong evidence that he or she is not temperate. Without real argument, then, Alfano concludes that "fairness, fidelity, honesty, justice, and trustworthiness" are among the high-fidelity virtues.[16] This list, he notes, is not supposed to be comprehensive or uncontroversial.

On the low-fidelity side of the distinction, we are supposed to see that low-fidelity virtues are not susceptible to disconfirmation in the same way. Alfano remarks that "whether someone gives money to charity" is not very strong evidence for or against his generosity.[17] For virtues like generosity and benevolence, then, a single observation of behavior that is contrary to a virtue's characteristic behavior provides only *weak evidence* against the virtue's being possessed. Alfano therefore argues that, at least for these low-fidelity virtues, one-off experiments are not helpful for disconfirming their presence.

Before I proceed, a clarificatory point is in order. A charitable reading of Alfano's distinction would understand him as describing the *strength of evidence* that a single behavioral observation will provide in disconfirming the presence of a virtue (depending on whether it is high- or low-fidelity). If he were claiming that a single observation could disconfirm a virtue's presence absolutely, then, at least in the case of high-fidelity virtues, that would entail these virtues are exceptionless dispositions to act. Admittedly, Alfano sometimes talks as if high-fidelity virtues are exceptionless dispositions to act, and at other times he characterizes them as requiring only "near-perfect consistency."[18] Candace Upton has rightly criticized him for being inconsistent on this point.[19] But I do not think Alfano's distinction needs to be conceived as a distinction between exceptionless and non-exceptionless virtues. Alfano can simply be understood as talking about the *evidential strength* that a

---

Footnote 15 (continued)

categories. See, e.g., Nicole Tausch, Jared Kenworthy, and Miles Hewstone, "The Confirmability and Disconfirmability of Trait Concepts Revisited: Does Content Matter?", *Journal of Personality and Social Psychology* 92, no. 3 (2007): 542; John Skowronski and Donal Carlston, "Negativity and Extremity Biases in Impression Formation: A Review of Explanations," *Psychological Bulletin* 105, no. 1 (1989): 131–42; Dwight Riskey and Michael Birnbaum, "Compensatory Effects in Moral Judgment: Two Rights Don't Make up for a Wrong," *Journal of Experimental Psychology* 103, no. 1 (1974): 171–73. This literature shows that Alfano may not be alone in his intuitions regarding the possibility of "high-fidelity" virtues, but it does not undermine the conceptual and philosophical argument of this paper.

[16] Alfano, *Character as Moral Fiction*, pp. 31–32.

[17] Ibid, p. 73.

[18] Ibid, p. 31.

[19] See Upton, "The Empirical Argument Against Virtue," p. 362. On the one hand, Alfano claims that high-fidelity virtues only require "near-perfect consistency." See Alfano, *Character as Moral Fiction*, p. 31. At other times he says things like, "A single test can disconfirm someone's possession of a [high-fidelity] virtue." See Alfano, p. 73.

single disconfirming observation provides for disconfirming the presence of certain types of virtues.

The payoff for Alfano's distinction is twofold. First, it provides a nice rejoinder to those who employ the longitudinal data critique in the context of the situationist debate, at least for a subset of virtues. Second, the distinction could help future efforts to measure empirically the possession of virtues by focusing scientists on those virtues that are high-fidelity (whether local or global). This would allow social science to establish more confident boundary conditions on how widely high-fidelity virtues are held, if at all. Situationists and virtue ethicists agree that virtues are rare, but quantifying this rarity has proven to be extremely difficult. Alfano's distinction could thus help researchers identify virtues that can be measured more confidently with one-off, cross-sectional experiments.

2 A Distinction Without a Difference

Unfortunately, Alfano's distinction between high-fidelity and low-fidelity virtues is what philosophers call a "distinction without a difference."[20] The distinction is invalid or unreal. It purports to tell us something about the *nature of virtues* by dividing them into two distinct classes. But in reality, the distinction, if it exists, can only point to a difference in the evidential relation between a single virtue and a behavior as it is situated in a certain context (a "behavior-in-context"). In some cases, the behavior-in-context will be high-fidelity, because it will provide strong disconfirming evidence of the virtue in question. But in other cases, the behavior-in-context will not provide strong disconfirming evidence of that very same virtue. Alfano's distinction therefore does not provide a good defense against the longitudinal data critique as he intends it to.

Let us formulate Alfano's distinction more precisely. The high-fidelity / low-fidelity distinction purports to tell us something about the disconfirmability of a virtue V_x. It is supposed to tell us whether any single behavior that is *contrary* to V_x will always count as strong evidence against a person's possession of V_x. If the virtue is high-fidelity, then a single contrary behavior will always count strongly against V_x's possession. If the virtue is low-fidelity, then a single contrary behavior does not. To see why the distinction fails, let us examine some examples, first starting with the supposed low-fidelity virtue of generosity.

Alfano contends that generosity is low-fidelity. He contends that if we observe a person passing a homeless man on the street without giving him money, it is only weak evidence against the person's generosity. Alfano has the intuition – I think rightly – that witnessing this type of behavior does not speak strongly against a person's possession of the virtue. But he incorrectly thinks that *the virtue* of generosity must therefore be low-fidelity.

---

[20] Robert Sokolowski, "The Method of Philosophy: Making Distinctions," *The Review of Metaphysics* 51, no. 3 (1998): 520.

Consider a different case. Imagine there is a wealthy person living in NYC who is waiting to catch a bus because he is going to see his child perform in a school play. Imagine also that, as he sees the bus arriving, he is approached by someone whom he recognizes as the father of one of his child's classmates. This parent tells the wealthy man that he has forgot his wallet and needs money for the bus fare. Assuming the wealthy man has the cash on him, a refusal to spare a few dollars in this situation would be much stronger evidence that he is not generous than in the case of passing a homeless man.[21] Here the situational variables change the evidential significance of what is arguably the same type of behavior – in both cases we observe a person not giving a few dollars to a person in need. But the context of the bus scenario gives us much stronger evidence that the person is not generous. First, we know the man is wealthy, and the financial request is almost trivial. We also know that he is acquainted with the person asking, he knows what the need is for, and how the money will be used. There are relatively few explanations that could explain how a person could truly possess the virtue of generosity and fail to give money in this situation.[22]

Here we see that the *virtue* of generosity is neither high-fidelity nor low-fidelity. Depending on the context and the type of behavior observed, a single observation may provide strong or weak evidence against someone's generosity. Alfano has thus mistakenly ascribed the property of low-fidelity to the *virtue* when, in reality, the properties of high-fidelity and low-fidelity – if they are properties at all – tell us only whether the contextual variables surrounding a behavioral performance are sufficient (and sufficiently detailed) such that it would count strongly (or only weakly) against a person's possession of a virtue. Whether a wealthy person gives money in the bus scenario is a much higher-fidelity behavior-in-context than whether he gives money to one homeless person out of the many whom he may come across.

Consider next the virtue of justice, which Alfano presumes to be high-fidelity. Imagine an experimental setup where those at a supermarket cash register are given more change than what is owed. In this experiment, subjects would not be not getting their due, and an acceptance of the change (or failing to give it back) would be an act (or omission) contrary to justice. But would this one-off observation provide *strong* disconfirming evidence of these people's justice? Context seems to prevent such a conclusion. A person in this situation might feel rushed at the checkout counter if there is a long line behind her, and she may not want to hold other people up. Or she may worry that an attempt to rectify the situation will cause great delay for herself. Given that the monetary value of the change is very small, she may justify walking away without doing anything to correct the situation. She may even feel some guilt as she walks away, which itself would be indicative of a concern for justice and possibly possession of the virtue. It would be improper to conclude based on just this

---

[21] Many thanks to Matthew T. Lee for help with this example.

[22] I acknowledge that a creative reader might imagine a background story that would explain such a behavior without implicating the man's generosity, but the point remains: such an observation would provide strong disconfirming evidence against the man being generous, even if this observation would not be 100% conclusive.

one observation that she lacked the concern for justice necessary to possess the virtue or that she would not otherwise be disposed to behave in ways that are reliably just. The experiment would provide *some* evidence against a subject's possession of justice, but more observation would be necessary to conclude that each participant who to fails to give the change back is not otherwise characteristically just.

There are, however, unjust behaviors in other contexts that would provide much stronger disconfirming evidence. For example, if we were to observe a boss cheating an employee out of wages, or orchestrating insurance fraud, these behaviors would provide much stronger disconfirming evidence for justice than the supermarket case. Once again, the high-fidelity-low-fidelity distinction falls apart if considered to be a property of a virtue itself.

Finally, let us examine Alfano's primary motivating example – chastity. Surely this is a high-fidelity virtue if there ever was one. For it is true that cheating on one's spouse is pretty damning evidence that one is not chaste. But we might ask whether cheating is all there is to the virtue of chastity? Moral virtues traditionally have not been conceived as *merely* matters of abstaining from one or two behaviors. If that were the case, not smoking could be a virtue. Rather, virtues have been understood as complex dispositional states that regulate action, feeling, and desire in different domains of human experience, and these states are grounded in deep concerns and informed by practical wisdom.[23] The virtue of chastity, then, is not just concerned with the act of monogamous sex with one's spouse, though it is concerned with that too. The tradition, at least, has thought chastity to be a virtue that rightly orders and regulates sexual desire more broadly.[24] When understood in this broader sense, one can easily imagine chastity as being low-fidelity in some contexts. In other words, there are some behaviors in certain contexts that are both contrary to the right ordering of sexual desire and yet only provide *weak* disconfirming evidence of the virtue's presence. For example, if we were to observe a husband or wife taking a second glance at a sexually attractive stranger, that is *some* evidence against his or her chastity if it is sexually motivated, but that one act is not strong disconfirming evidence of the virtue.[25] Again, more observation is necessary. So even for the virtue of chastity, at least as it has been traditionally conceived, the high-fidelity / low-fidelity distinction falls apart.

---

[23] For an interesting challenge to this conception of virtue, considered as being a separate class or psychological kind that is distinct from depression, phobia's, and other mental pathologies, see Kate Abramson, "Character as a Mode of Evaluation," in *Oxford Studies in Normative Ethics*, vol. 6 (Oxford University Press, 2016), Chapter 3.

[24] See Romanus Cessario, *The Virtues, or the Examined Life* (Continuum USA, 2002), pp. 187 ff.

[25] Notice that if we specified the situational context in such a way that the second glance was due to the fact that the attractive person bore a striking resemblance to one's sister, then this behavior would not be any evidence *at all* against the person's chastity. The issue, one can see, is not the behavior; it is the motivation. Thanks to Robert Roberts for help on this point.

3 Could There be High-Fidelity Experiments?

The distinction between high- and low-fidelity *virtues* is ill-conceived, but it seems clear that some experimental contexts can in fact provide stronger evidence for the absence of a virtue than others. If this is correct, then perhaps it is possible to reconstruct Alfano's distinction as one applying to *experiments* that test for virtues, rather than the virtues themselves. High-fidelity experiments would be those one-off experiments where an agent's behavior provides strong evidence against the possession of a virtue, whereas low-fidelity experiments would be those in which this was not the case. If this is right, we may expect that the conditions that make an experiment high-fidelity or low-fidelity may be difficult to get clear about. Before attempting this, however, let us make a now common philosophical distinction between *behaviors* and *actions*.[26]

Roughly, a mere behavior is some observable movement of body parts, an utterance of words, or a combination of the two. Mere behaviors fall under descriptions like, "the man pulled his hand out of his pocket and put two dollars in that homeless man's hands." In observing a mere behavior, one does not know why people move or speak in the way they do. Let's distinguish this from an action, which is always described with reference to the intentional states that accompany a person's movements and utterances. "He helped the homeless man on the street" describes an action because the verb "help" has implicit intentional content. Rich descriptions of actions will refer to an agent's motivation, reasons for acting, and emotional states.

It is important here to notice that one and the same behavior may actually be two different types of action. "She spilled the coffee cup onto that man's lap," – a mere behavior – could be intentional or unintentional. If unintentional, it was an accidental spill; if intentional, it might have been an act of malicious revenge or a move of clever self-defense. Knowing the motivations and intentions that accompany a behavior is all-important in determining which action it is.

Strictly speaking, a *mere* behavior can never serve as evidence for or against a moral virtue. The observation of a mere behavior does not provide the observer access to the intentional states that are relevant for evaluating it as a moral action. Virtues are not dispositions to merely behave well; they are dispositions to act, think and feel well. Virtuous action by definition requires the right sorts of intentions, motivations, concerns, patterns of thinking and emotions to be virtuous. Observations of behavioral movements and utterances are therefore always underdetermined when considered as providing evidence for or against a virtue. Strictly speaking, the mere behavior of a man who hands ten dollars to a homeless man is not evidence for his generosity. If the *action* was performed out of a proper concern for the homeless man, then it would be evidence of a generous disposition. But if handing over ten dollars was an attempt to impress a woman walking beside him, *that* action would not be.

---

[26] For an overview of action theory and the debates in the metaphysics of action, see George Wilson and Samuel Shpall, "Action," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Metaphysics Research Lab, Stanford University, 2016), https://plato.stanford.edu/archives/win2016/entries/action/.

An important point to notice, then, is that anyone who claims a one-off experiment provides evidence against a virtue must be making implicit inferences about the intentions and motivations of agents. These inferences can be legitimate, but they are always probabilistic. Researchers must infer what was likely motivating an agent from the contextual details of a situation and what is generally known about human nature.[27] Likewise, if an experiment is to provide evidence against the possession of a specific virtue, the inferred action must be *contrary* to the virtue in question. If the subject's behavior were in accordance with the virtue in question, then it could not serve as evidence against one's possessing the trait.[28] In order to test for the absence of a virtue, then, experimental setups must elicit behaviors (or testimony of behaviors) that may be inferred as actions *contrary* to the virtue being tested-for. This may seem obvious, but as we shall see, giving an account of what it means to be contrary to a virtue is not as simple as one might think.

Having stressed the difference between observable behavior and action, we are now in position to notice an important difference in two measurement methods that could be used to assess the presence of virtue. For purposes of this paper, I call these methods *researcher observation* and *self-report*.[29] In researcher observation, what experimenters observe is the mere behavior of the experimental subject. The subject's motivations and intentions – and thus her actions – must be inferred from the context of the situation and what is probabilistically known about human nature. If we imagine observing a young man who turns his head to look at an attractive young woman, we might reasonably infer that he is doing so from sexual attraction, given what we know (probabilistically) about the nature of men and women. But there are other possible explanations. Perhaps the woman resembled his sister, or she was carrying an interesting looking book.[30] Competing explanations prevent these types of inferences, though inductively strong, from being conclusive. Importantly, such inferences require contextual information and presume lots of background information.

A second measurement method, self-report, involves querying a subject directly for his or her own perspective on her behavior or virtue. Such questioning may be done through interviews or survey questionnaires. This method gives more direct

---

[27] One might object that it is possible to give an account of virtue that is merely behavioristic. A purely behavioristic account of virtue, however, would not support the distinction between high-fidelity and low-fidelity virtues or experiments. If virtues are reliable dispositions merely to behave in certain ways and in certain circumstances, then a *single observation* of a behavior would not have any *predictive power* as to whether the individual would reliably behave that way. Without inferring the intentions or motivations behind someone's behavior, one would have no basis for making predictions from a one-off experiment. Longitudinal data would absolutely be required on the behavioristic picture of virtue.

[28] For more on the distinction between acting "from" and acting "in accordance" with virtue, see Robert Audi, "Acting From Virtue," *Mind* 104, no. 415 (1995): 449–471.

[29] Admittedly, many times these two methods are combined in a single experimental design, but they are nonetheless distinguishable as different types of measurement.

[30] Thanks to Robert C. Roberts on this point.

insight into the psychological states that are relevant to actions and virtue, and one's conclusions are therefore less reliant on background assumptions.[31]

For the rest of this paper, I focus on the researcher observation method because that is the method from which most have attempted to draw conclusions about the absence of virtue. Interpreters of such experiments commonly infer an action (or non-action) by observing a behavior that is supposed to be relevant to a virtue. In many cases, a manipulation or intervention is also inferred to be causally connected to the behavior observed. For example, a subject might be put into a room where there is a loud crash or scream next door. Or subjects might be put in a situation where they walk by a person in need.[32] The manipulation could even include a psychological component. For example, experimenters have attempted to induce guilt, embarrassment, and exert other sorts of psychological pressure prior to observing a subject's behavior.[33] The differences that these manipulations make are important for some of the situationist's arguments, but they are less relevant for the longitudinal data critique or testing for virtue. Regardless of the type of manipulation, what is observed is still a behavior that must be interpreted as a particular action, and this action must in turn be contrary to a virtue.

Let us return to the question that lies behind Alfano's intuitions. Why do some one-off behaviors obviously seem to provide very strong disconfirming evidence of virtue, while others do not? There seem to be at least three factors that weigh into judgements about an experiment's evidential strength. These are:

(i) the observer's *confidence level* in inferring the action that a subject undertook, compared to the mere behavior that was observed;

(ii) *how the inferred action is contrary to virtue*: specifically, whether an agent merely fails to act as one might expect (if she possessed V_1), or if she acts in a way that is overtly vicious (and thus contrary to V_1); and

(iii) the *moral seriousness* of the act as it relates to the virtue in question.

These three factors, in combination, are important for weighing the evidence that a single behavioral experiment can provide against a person's possession of a virtue. I discuss each in what follows.

(i) **From Behavior to Action: How Strong is the Inference?**

---

[31] In social pscyhology experiments, this type of self-report data has often been available but remains largely ignored. In both the Milgram and the Princeton hurry studies, experimenters conducted this kind of post-experiment questioning, but situationist philosophers largely ignore this data in making conclusions.

[32] See Clark Clark Russell D. and Larry E. Word, "Why Don't Bystanders Help? Because of Ambiguity?", *Journal of Personality and Social Psychology* 24, no. 3 (1972): 392–400; Darley and Batson, "'From Jerusalem to Jericho': A Study of Situational and Dispositional Variables in Helping Behavior."

[33] See Vladimir Konecni, "Some Effects," *Journal of Personality and Social Psychology* 23, no. 1 (1972): 30–32; Arnie Cann and Jill Blackwelder, "Compliance and Mood: A Field Investigation of the Impact of Embarrassment," *The Journal of Psychology* 117, no. 2 (1984): 221–26; Wim Meeus and Quinten Raaijmakers, "Administrative Obedience: Carrying Out Orders to Use Psychological-Administrative Violence," *European Journal of Social Psychology* 16, no. Oct–Dec 86 (1986): 311–24.

As discussed previously, observational research experiments observe behavior, not action; for the purpose of studying virtue, researchers must make assumptions about the intentions and motivations behind the behaviors they observe. The stronger and more probable the link between an observed behavior and an inferred action, the stronger the evidence that an observation may provide against a subject's possession of a virtue. If we are looking to disconfirm a virtue, then we must look for *actions* that are contrary to that virtue.

Having a strong inferential link between behavior and action is therefore an important desiderada of any high-fidelity experiment. Unfortunately, I see no principled or theoretic way to develop criteria that would determine the strength of an observer's inferential reasoning. These types of inferences are always situation specific; making good judgements requires good reasoning on the part of the observer *and* a good understanding of human nature. Some have even thought that, in order to recognize virtue, one must possess virtue oneself. Nevertheless, one thing that we can say is that the more thoroughly the context of a behavior is specified, and the more background information that we have about an experimental subject, the more confidently we can infer which action a subject performed. In many contexts, however, it will not be possible to judge with great confidence what action a subject performed because of competing explanations. (Think of the young man who turned to look at the attractive woman). We should also note, then, that one's ability to infer an action from an observation of behavior will decrease significantly if the observation is someone else's report – since many of the visual and other contextual clues will not be present. This seems to be what many situationist philosophers are doing when they use one-off social psychology experiments to make judgments concerning the absence of virtue. And notably, the experiments are not even designed to test for virtue.

(ii) **The Failure to Act vs. Acting Viciously**

Actions may be contrary to virtue in two ways. A person may *fail to act* in a way that we might expect of a virtuous person; or alternatively, a person might act in a way that is characteristic of one *who possessed a related vice*. An overtly vicious action provides stronger disconfirming evidence of a virtue than a mere failure to act, because a person cannot possess a global virtue and a corresponding global vice at the same time.[34] Let us examine each of these cases in turn.

An action may provide some disconfirming evidence of a virtue when an agent fails to act in the way we would expect of someone who possessed the virtue. In this case, one observes an experimental subject S acting in a way that fails to meet the expectations of what S would do if she possessed V_1, given the situation. But otherwise, S acts in a way that is not relatedly vicious. However, determining what

---

[34] Of course, neither could one simultaneously possess a so-called local virtue and a corresponding local vice. Here we should also note that the observation of one vicious action does not, necessarily, preclude a person from possessing a virtue, at least to some degree. While it is true that a person could not possess a vice and a virtue at the same time, it is possible that one could act poorly and from bad motivations and still possess a virtue. A theory of virtue should be able to accommodate for significant, although uncharacteristic, moral failings.

one should expect of a virtuous agent is often difficult. To claim that S fails to meet the expectations of V_1, one must be able to make a judgement (or a series of judgements) about what a person with S's constitution and characteristics, who possessed V_1, would likely have done in the experimental context. In many cases this requires that one determine a *set* of possible virtuous actions specific to the trait in question. The actions in the set may be *unique for each individual* given the same experimental circumstances; for as Aristotle reminds us, Milo the wrestler may eat significantly more than most people and still possess the virtue of temperance.[35] The observer must then judge that S's actual behavior fails to qualify as an action within the virtuous set, and the absence of virtuous action is then what counts as evidence for the absence of V_1.

This type of reasoning is present in much of science. We have a theory that predicts some outcome, and observations that are not congruent with that prediction count as evidence against the theory. This reasoning is what many situationists employ when they interpret social-psychological experiments as evidence for the absence of virtue. For example, in Baron's (1997) study, situationists expect a compassionate person to make change for a passerby in a shopping mall. The absence of this behavior (notably again, in both control and experimental groups) is taken as evidence for the absence of compassion.[36] Likewise, in the Darley and Batson hurry study (1973), situationists expect a compassionate seminarian to help a man slouched in an alleyway. If the seminarian passes without helping, situationists infer that he or she lacks compassion. Notice, however, that experimenters do not observe shoppers or seminarians acting in ways that one would consider overtly vicious. They merely fail to behave in ways that one might expect from a person with V_1.[37] For this reason, the observations do not give us much insight into what the subject understood himself to be doing. All one knows is that, in this one instance, the subject failed to perform an action that an outsider might expect of someone with the virtue. But without further insight into the subjects' actual psychology, it is nearly impossible to make any further inferences about the *lasting* concerns, motivations, or desires that this one behavior might be the result of – i.e. those that might be partially constitutive of a disposition or virtue. And because of this lack of insight, a single observation of a person's failure to behave in accordance with a virtue can provide only weak evidence, if any (depending on context), against a subject's possession of that virtue.

---

[35] See Aristotle, *Nicomachean Ethics*, 1106b1. Some philosophers have complained that situationists only unrealistically and implausibly allow for only one possible behavior. See, for example, Upton, "The Empirical Argument Against Virtue." To my knowledge, none of the experiments cited attempt to account for individual differences in people in relation to the measured outcomes.

[36] See Miller, *Moral Character: An Empirical Theory*; Alfano, *Character as Moral Fiction*.

[37] I grant that these subjects fail to meet the expectations of V_1 for sake of argument. However, these inferences are often made without taking into account important information about the experimental subjects. For example, if a shopping mall walker failed to make change because she was late to pick her child up from school, that contextual detail would significantly change one's evaluation of her compassion.

Here one might object that a failure to act can be relatedly vicious if the refusal derives from a vicious motivation or concern. The unwillingness to give money to a homeless man, or even to one's best friend, may be vicious if because of greed. But for a single observation to provide sufficient evidence for such a conclusion, one would need significant contextual detail to infer desires, concerns, motivations, patterns of thinking, etc. to be present in the subject that are related to the vice of greed. Context must therefore also rule out the possibility that the subject's failure to act was not caused by some other *non-related* vice. For example, we can imagine a wealthy friend being deeply envious of her best friend and refusing to help for that reason. That would indicate envy, not greed, and this omission would therefore provide much weaker evidence against her possession of generosity. Or imagine that our friend is angry at her best friend for some small offense that occurred many weeks ago. People who get too angry at small offenses, and those who stay angry for too long, suffer from ill-temper, not greed. One can see that there is a significant amount of contextual detail that would be needed to rule out such possibilities, if *omissions* to act are to count as strong evidence to disconfirm a virtue.

Let us now return to the second way that an action can be contrary to a virtue. This happens when we observe a person acting in a way that is both relatedly and overtly vicious. In other words, we observe S behaving in a way that provides direct evidence for a vice related to V_1. The observation here is not merely the *absence* of an expected action but is rather what we would expect if S possessed a V_1-related vice. Examples could include observing an overtly unjust act (e.g., stealing, fraud, embezzlement), an act of outright greed (e.g., selling an investor worthless CDOs to make a huge commission), adultery, or perjury. Such actions are more directly contrary to virtues because they are evidence of dispositions that would exclude the virtue.[38] The observation not only provides the same type of evidence discussed previously – the behavior fails to confirm a theory of how we expect the virtuous to behave – it also provides some evidence for a second, competing theory; namely, for the presence of a related vice. A theory of vice expects that people who are unjust, greedy, or unfaithful do things like steal, sell worthless CDOs and commit adultery. Thus, the same inferential steps are valid as in the failure to act cases, but now one also has evidence for the presence of possible concerns, desires, and other psychological structures that are directly contrary to V_1.

We should note, however, that even in this second set of examples the possibility of competing explanations is present. Acts of stealing, selling worthless CDOs, and even adultery can all be performed for reasons that are not motivated by a vice that one might expect. One can steal to feed the poor, sell CDOs to pay for cancer treatments, or commit adultery for political advantage. In each of these cases, the motivations behind such actions are not characteristic, nor expressions of, the related vice. So merely observing such vicious behaviors without truly knowing the person's motivations means that these inferences too could be mistaken. Knowledge of context and background information are still all important. On the other hand, we know that if such a person did possess the related virtue, her virtuous motivational

---

[38] Here I assume that virtue and vice cannot co-exist in the same subject at the same time.

dispositions were not strong enough to overcome whatever the competing reasons were. For we can imagine people who are so committed to justice that they would never steal, no matter what the cause, and chaste people who would remain faithful to their partner even if it cost them their life.

Observing an action that is both relatedly and overtly vicious with respect to V_1 therefore does not tell decisively against one's possession of a virtue. But considered from within this dimension of evaluation, behaviors that are more directly contrary to virtue do provide much stronger evidence against a virtue's presence than a mere failure to act.

(iii) **The Moral Seriousness of an Action (or Non-Action)**

A final factor that contributes to the strength of evidence that a one-off observation has against the presence of virtue is the moral seriousness of the (inferred) action (or non-action). There is a common intuition that egregious acts provide stronger evidence against the possession of virtue than those that are less morally serious.[39] Let us call this the *principle of moral seriousness*.

The principle relies on an assumption that morally egregious acts are often the result of a malformed (and thus non-virtuous) moral psychology. Non-egregious acts, by contrast, are more common, such that one cannot infer a particular type of moral psychology by observing a single act. Likewise, we commonly assume that a person with well-formed moral dispositions could not perform egregious acts, although such a person might sometimes commit minor moral transgressions.[40]

For example, people commonly think that a person who commits murder could not be compassionately-disposed. Because of the egregious nature of murder, we assume that a single act allows us to infer more about the person's dispositions and psychological makeup than, say, observing someone who keeps some extra change at the cash register. We assume that to commit murder one could not have a deep lasting concern for human life or love for others.

The intuition behind the principle of moral seriousness seems plausible, at least most of the time. We expect people who possess V_1 to have the characteristic concerns, motivations and habits of emotion related to V_1 deeply integrated into their personalities, and thus serious moral transgression should be unlikely. This principle seems especially plausible because it is hard to imagine exemplars of virtue committing morally serious actions. One can hardly imagine Mother Teresa, who I assume possessed the virtue of compassion, treating a child cruelly.

But there are some counter examples to the principle that should make us take pause. Consider the virtue of honesty. According to the moral seriousness principle, one should expect that an honest person would not lie if the stakes are very high

---

[39] This intuition seems to be supported in psychological studies on negativity bias. For references, see note 15.

[40] This type of intuition dates back at least to Aquinas, who distinguished between degrees of sins according to their "gravity" and the relation of a sin's gravity to virtue and vice. See Eileen Sweeney, "Vice and Sin," in *The Ethics of Aquinas* (Washington, D.C.: Georgetown University Press, 2002), p. 160.

(here I assume that higher stakes are correlated with increased moral seriousness), but that one might find an honest person, on occasion and uncharacteristically, lying in low stakes situations. But this doesn't always hold. It is also reasonable to think that a thoroughly honest person would *always* be honest in low stakes situations, where there is little or no cost, and only in high-stakes or morally serious situations uncharacteristically lie. For example, we can imagine a thoroughly honest parent who lies to the police to keep her guilty child from going to jail. Honesty, at least, seems to be a virtue where sometimes there is an inverse relationship with moral seriousness and the likelihood that one should expect an uncharacteristic behavior.

There are other cases too, where the uncharacteristic behavior of an agent can turn out to be quite morally serious. Exceptional but significant moral failings may be contrary to the way a person is otherwise disposed. Kupperman, for example, has argued that a virtuous person might err seriously and yet still be virtuous – a fact that situationists often fail to acknowledge.[41] Such cases show how the principle of moral seriousness cannot be relied upon exclusively in making judgements about virtue. Nevertheless, the principle retains a strong plausibility, even if it must be applied carefully.

To conclude this section, then, there are at least three factors that help to explain why some one-off researcher observation experiments may provide stronger disconfirming evidence for a virtue than others. These are (i) the observer's confidence level in inferring the *action* an agent performs from a mere behavior that is observed; (ii) whether an agent merely fails to act as we might expect or if she acts in a way that is overtly vicious; and (iii) the moral seriousness of the act as it relates to the virtue in question. These factors are important considerations in determining whether a particular experiment can be considered high-fidelity.

4 An Example

Assuming the possibility of creating a high-fidelity experiment exists, can we give an example? In other words, what type of experiment would, in principle, provide an opportunity for people to act in ways that are (i) directly contrary to a virtue, (ii) morally serious, and (iii) where we could have a high degree of confidence in the intentionality and motives behind their actions, without collecting self-report data? In what follows, I propose a hypothetical experiment that would fulfill the desiderata of a high-fidelity experiment for the virtue of justice.[42]

---

[41] See Kupperman, "Virtue in Virtue Ethics," 244. For an opposing view, see Nancy Schauber, "How Bad Can Good People Be?", *Ethical Theory and Moral Practice* 17, no. 4 (2014): 731–45.

[42] These hypothetical experiments are for illustration purposes only and have not been evaluated for their merits with respect to feasibility, cost, or ethical issues that concern research on human subjects. Many thanks to Dr. Robert Gahl, who provided ideas for these examples while discussing a previous draft of this paper.

i. **The Forgotten Wallet Experiment**

Imagine that subjects are secretly videotaped as they enter a hotel room. Staged in the room is a wallet sitting on the ground, on top of a purse. The wallet has $432 dollars in it, and some of the money is visible. Both items are located by the bed against the far wall.

What do the experimental subjects do? Do they report the wallet? Do they take it to the hotel's front desk? Do they keep the money, and/or the wallet? Do they leave it and do nothing? This type of setup has the *potential* to be a high-fidelity experiment. Stealing the money would be directly contrary to the virtue of justice, and such behavior would be *prima facie* vicious. The amount of cash is significant enough that stealing it would be morally serious. But this is still a complicated case, and there are many contextual issues that a designer of a high-fidelity experiment would need to address, especially if we are to infer with great confidence an experimental subject's intentions and motives.

Here are some example considerations: First, we would not want there to be a driver's license or photo ID in the wallet. If that were the case, any subject who ostensibly "steals" the wallet may actually be planning to return the wallet by mail or some other means. On the other hand, having a few credit cards in the wallet might increase the moral seriousness of the situation by directly linking the wallet and its contents to a real person.

Second, we might consider the best type of test subjects for this experiment. Would it be better to use hotel housekeepers or hotel guests? If we were to run the experiment with housekeepers, it could introduce some unwanted competing explanations. That is because, generally speaking, housekeepers are of a lower socio-economic status. What if a housekeeper who steals the wallet has a sick child at home and desperately needed to pay for medical care? Or maybe there is some other significant financial need inciting the housekeeper to keep the money, though it is otherwise uncharacteristic of her. The housekeeper's action might plausibly be explained by some overriding concern that is not directly related to justice. The experiment would be thus better designed by observing hotel guests, and preferably, guests who are checking into a nice, but not an ultra-luxury, hotel. This would make it more likely that the person finding the wallet is not in a place of unusual financial need (or at least not such that $432 would remedy), and this reduces the possibility of competing explanations. We also don't want multi-millionaires who stay in ultra-luxury hotels as test subjects, as they may have lost their appreciation for $432.

Finally, the location of the forgotten wallet is significant. Putting the location of the lost wallet in a hotel rather than, say, on a street corner, is an important consideration. There is a high probability the hotel would be able to contact the rightful owner, and there is very little effort required to performing the just action – i.e. one only has to call the front desk or walk it down to the lobby. It seems reasonable, then, to infer that any hotel guest who takes the money in this situation is knowingly stealing it. It is highly unlikely that a person who possesses a significant concern for justice (enough to qualify as possessing the virtue) would keep the money. Thus, his type of observed behavior-in-context would be strongly disconfirming of the virtue.

Of course, there are still other experimental design considerations that might be examined in even this example. The discussion here is illustrative. One important takeaway from the example is that designing a high-fidelity experiment is no simple or clear-cut task. It requires lots of forethought and planning. However, the example does illustrate that high-fidelity experiments are in principle possible.

5 Conclusion

I have argued that the high-fidelity / low-fidelity distinction, considered as a property of virtues, is a distinction without a difference. I provided counterexamples showing that virtues like generosity, justice, and chastity cannot have the property of being be high-fidelity or low-fidelity, as Alfano supposed. The distinction therefore does not provide a defense against the longitudinal data critique of situationism, nor will it help social scientists, policy makers, or educators in their goals to measure virtue.

Nevertheless, I have argued that certain types of *experiments* may be high-fidelity. These experiments provide strong disconfirming evidence against the presence of virtues from mere one-off observations of behavior. I offered three considerations that can help to determine whether an experiment is high-fidelity. First, behaviors must be understood within a rich contextual setting that allows researchers to infer, with great confidence, that agents are *acting* (and not merely behaving) in ways that are directly contrary to a virtue. Second, those actions must be relatedly-vicious to the virtue in question and not the result of some other non-related vice. The behaviors also must not be liable to some competing explanation, which would explain the behavior as uncharacteristic. Third, the subject's actions should be morally serious, according to the principle of moral seriousness discussed here. One-off experiments that meet all three of these desiderata will have a higher likelihood of providing strong disconfirming evidence for a virtue. I illustrated the ways these desiderata work in concert in *the forgotten wallet experiment*, which could reliably disconfirm the presence of justice from a single behavioral observation.

Unfortunately, most of the social psychology experiments cited by philosophers do not seem to be high-fidelity, because they fail to meet one or more of the desiderata outlined. A detailed review of this literature is beyond the scope of this paper, but my own impression is that many of these experiments actually fail on all three counts. Although the obedience experiments (e.g., Milgram 1974; Meeus and Raaijmakers 1986) do well on eliciting *behaviors* that are contrary virtue, and these behaviors are indeed morally serious, they don't allow us to infer with confidence that the subject's intentions and motivations are relatedly-vicious – i.e. that they are *acting* with the intentions and motivations of someone who likely possesses a vice related to the virtue in question. By contrast, the Zimbardo prison experiments would seem more likely to qualify as high-fidelity experiments on all three counts.[43]

---

[43] See Phillip Zimbardo, *The Lucifer Effect: Understanding How Good People Turn Evil* (Random House Trade Paperbacks, 2007). There has been some recent news which has called into question the integrity of these experiments, for example, in Scott McGreal, "Twilight of the Stanford Prison Experiment," Psychology Today, September 27, 2019, https://www.psychologytoday.com/us/blog/unique-everybody-else/201909/twilight-the-stanford-prison-experiment.

The examples in this paper illustrate the ways in which it may be possible for a one-time behavioral experiment to provide strong disconfirming evidence for a virtue. I think Mark Alfano was onto something – even if he mistakenly ascribed the property of being high-fidelity to virtues themselves. I hope that this discussion gives both philosophers and empirical scientists a much richer context from which to design and evaluate more rigorous experiments that empirically test for virtue.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Matthew F. Wilson is Assistant Professor of Philosophy at John Brown University and an Associate at the Human Flourishing Program at Harvard University's Institute for Quantitative Social Science.