Situationism and trait-eliciting situations

Adam Piovarchy

1. Introduction

Situationist experiments demonstrate that people's behaviour is influenced by seemingly irrelevant environmental factors much more than we expect. Doris (2002, 2005) famously argues that they show we lack the kinds of global character traits posited by many theories of virtue. He proposes that any empirically grounded theory of virtue ought to meet three criteria. A theory *is globalist* to the extent that it takes traits to be robust, implicating them in behaviour in a wider variety of contexts. A theory *has empirical content* to the extent that it forwards claims that may be substantiated by observation, particularly explanatory and predictive claims. Finally, a theory *is empirically adequate* to the extent that we have evidence matching its empirical content.

What evidence would make a theory empirically adequate? Doris's proposed test is this: 'If a person possesses a trait, that person will engage in trait-relevant behaviours in trait-relevant eliciting conditions with markedly above chance probability' (2002: 19). While a number of philosophers argue that the situationist studies do not show that global traits do not exist, Doris (2005) believes their arguments fail to meet this test in a way that satisfies the above criteria. Arguments that the situationist experiments do not threaten global conceptions of virtue because subjects experience the right emotional reactions, or because traits are only dispositions which need not result in actions, purportedly fail to have sufficient empirical content. Arguments that virtue is supposed to be rare purportedly only become empirically adequate by limiting their scope to a tiny proportion of humans.

I want to consider Doris on his own terms, accepting that virtue theorists should strive for a globalist theory of traits, that observable behaviour matters to our assessments and that theories of character should not have only the virtuous few in their scope.[1] My focus is on the most famous experiments that received the most philosophical attention. These are studies

1 Doris's scepticism has a few fronts. One concerns the ubiquity of situationist factors, given that there are thousands of situationist experiments. I will set this aside because most situationist factors cause only minor changes in behaviour in settings where agents' reasons for *Φ*-ing are weak, for example tipping waiters or agreeing to do a survey. Even if there are thousands of such studies, it remains to be shown that such factors influence us beyond a certain range, the boundaries of which might be grounded by something like global character traits. Another concern is that many people's behaviour doesn't seem evaluatively integrated (e.g. someone who scores high on one trait often scores low on another related trait). But as Webber (2006) argues, Hartshorne et al.'s (1929) studies showing low cross-situational consistency among children may have limited relevance for adults, and it is contentious whether we should think that not stealing and not cheating on a test both fall under the same trait.

Analysis Vol. XX | Number XX | XX 2022 | 1–9 doi: https://doi.org/10.1093/analys/anac035

in which slight social pressures generate disproportionately egregious outcomes, which Doris (2005: 657) emphasizes is a key problem for globalist trait theories: Latané and Darley's (1970) studies on the *Bystander* effect, Darley and Batson's (1973) 'Good *Samaritan*' study, and Milgram's (1974) *Obedience* to authority experiments.

I argue that difficulty satisfying Doris's test is what we ought to expect, without this constituting a problematic retreat for said theories regarding empirical content. In particular, what counts as a trait-relevant eliciting condition is inherently more difficult to identify than parties to this debate have realized. My argument proceeds by considering how we assess capacities. Although situationist subjects possess the general capacity to avoid wrongdoing, close attention reveals they may lack the specific capacity to avoid wrongdoing. This matters, because settings in which agents lack certain capacities are not fair tests of character trait possession.

2. *Capacities, traits and situation-specific inferences*

Doris and others take trait-eliciting situations to be settings in which there are clear moral reasons to *Φ*, and in which we would expect agents with a relevant trait to manifest it by *Φ*-ing. As he notes, it is possible to make nearly any pattern of responses consistent and thus evidence for *some* kinds of traits according to *some* standard (2002: 71–89), but some standards of consistency matter more than others. What should matter for us is consistency of behaviour relative to our 'shared ethical standards' (84). We do not want people to simply exhibit compassion-when-in-a-good-mood, we want them to regularly be compassionate in settings where our shared standards demand compassion.

Identifying which settings count as trait-relevant requires that we know quite a bit. For example, we need to avoid choosing settings where agents have good reasons for failing to meet our ordinary standards, such as a setting in which one ought not to display compassion, all things considered. More importantly, we need to avoid choosing settings featuring factors that compromise an agent's capacities or opportunities, and this is something that has not been adequately considered by parties to this debate. To show what I mean, suppose someone nearby requires aid, and I fail to help. We might infer that this counts against my having the global character trait of compassion. But if I were tied to a chair, my failure would not indicate that I lack the trait of compassion at all. The same applies if I was unconscious or excusably unaware; situations with such features are simply not trait-eliciting situations. To be sure, none of the situationist subjects experience coercion, nor does anything physically restrain them. And so, at first glance, it seems the situationist experiments are trait-eliciting settings, which has led to a lot of debate about whether the subjects are simply responding to different reasons – and thus manifesting different traits – to the ones Doris expects (e.g. Webber 2006). But

the argument I wish to make here is distinct, and it stems from a problem that arises when we notice two things: (A) capacity and opportunity talk is very context-sensitive, and (B) evidence taken to show that agents lacked a trait is quite similar to the kind of evidence showing they lacked a certain capacity.

Let us start with (A). Philosophers have long been aware that capacity ascriptions are context-sensitive, depending on our interest in making an assessment. Does LeBron have the capacity to play basketball when on an airplane with no access to a court? In one sense, yes, as he knows how. In another, no, as he lacks the means to do so. To clarify things, philosophers typically distinguish between general capacities and specific capacities. General capacities are assessed by abstracting away from the particulars of an agent's situation (e.g. their broken arm or lack of a basketball court). But in asking whether an agent has a specific capacity, we hold certain factors fixed. Some philosophers also distinguish between capacities and opportunities, where opportunities are typically taken to be factors external to the agent. Jaster (forthcoming) points out, however, that the distinction between general capacities, specific capacities and opportunities is really just a matter of which factors we hold fixed while varying others to observe rates of success in relevant settings. Some capacities depend on factors external to the agent (the capacity to baptize depends on being ordained), some opportunities are lacking in virtue of being internal features of agents (being bed-ridden prevents one from exercising one's capacity to play soccer) and the distinction between general and specific is more of a gradient that depends on the number of factors we want to hold fixed.

The upshot for us is that because capacity ascriptions are context-sensitive, establishing that an agent has a general capacity to Φ does not entail they have the specific capacity to Φ in the sense relevant to our discussion. Someone can have the general capacity to Φ, and yet, due to some factor *temporarily* compromising or masking that capacity, lack the specific capacity to Φ.

Turn now to (B). Character traits and capacities are often talked about as distinct phenomena. My capacity to read street signs does not seem to be related to my character traits in any significant way. The trait of kindness is also not inherently related to notions of capacity; kind people are not unable to be mean in the way that people tied up with rope are unable to run away. But a number of similarities are present when we consider the capacity to recognize reasons and the capacity to react to reasons, which are particularly important for assessing agency and responsibility. Traits are dispositions to act, think and feel, and we infer their existence (or lack) by examining patterns of behaviour.[2] Similarly, capacities are often analysed by examining patterns of responses. Someone who more frequently recognizes reasons in more settings has a higher capacity to recognize reasons than someone who

2 For a more thorough examination of the nature of dispositions, see Heil 2005.

recognizes reasons less often. Someone who more frequently acts on moral reasons having recognized them has a higher capacity to react to reasons.[3] On a standard way of thinking about capacities, if it is true that someone would *never* recognize a certain class of reasons (say, those pertaining to compassion), then that person lacks the capacity to act according to reasons of compassion.

Putting (A) and (B) together, we find that the trouble is this: some failures to Φ are attributable to a lack of a capacity or opportunity to Φ, rather than a lack of Φ-relevant traits, and it is difficult to distinguish the two. One difficulty concerns distinguishing between competing explanations: an agent who lacks a certain trait and an agent who has the trait but lacks the specific capacity to manifest it will both fail to Φ. Using observable evidence alone, how can we tell the difference? Another difficulty concerns what kind of settings should count as providing evidence in the first place, as some factors that count as capacity-undermining for one kind of trait are trait-diagnostic for others. Suppose Kim always says kind things, unless more than five people are present, in which case her social phobia makes her unable to speak. Her failure to speak kindly is not evidence that she lacks kindness. Compare her to Connor, who reliably displays great social confidence unless more than two people are present, in which case he becomes a nervous wreck. We do not say that Connor's social confidence is masked by his capacity-undermining anxiety prompted by the presence of others; rather, his tendency to become a nervous wreck in such settings *just is* what counts against attributing this trait to him. Although we can make these assessments about Kim and Connor easily enough, in other cases there might be substantive disagreement about whether a certain factor ought to count as trait-relevant or capacity-compromising, and thus disagreement about which observations will count as evidence that someone lacks a trait.

This matters for Doris because it is plausible that features of the situationist experiments diminish the (specific) *capacity* of subjects to do the right thing for the right reasons.[4] There is *a* sense in which they possess this capacity, as evidenced by the proportion of people who successfully do the right thing in the control conditions. But this may not be the sense relevant to our trait assessments of those who fail.

Doris might emphasize here that global traits are supposed to be *robust*: their manifestation in relevant settings should not disappear given minor changes in circumstance. It is true that if someone's trait-relevant behaviour

3 This terminology is based on Fischer and Ravizza's (1998) account of reasons-responsiveness, but I am endorsing only a capacitarian account more broadly. See Nelkin 2016 for an alternative account.

4 A number of philosophers have argued that the situationist settings are not capacity-compromising. For objections, and arguments that they are capacity-compromising, see Rudy-Hiller 2020 and Piovarchy 2022. If one likes, one can frame this section's argument as a conditional: *if* the situationist experiments are capacity-undermining, then they are not trait-eliciting settings, so scepticism of globalism is not warranted.

disappears when only minor changes to their environment are introduced, we typically take this to count against them having the trait. But such inferences rely on a whole host of background assumptions about how those changes affect us or the extent to which they make certain choices more difficult, and thus what counts as minor in the first place. And while our assumptions are typically accurate, the fact the situationist experiments are so *surprising* when one first hears about them may give us reason to question whether such assumptions can be relied upon when making inferences based on subjects' behaviour.

Trait manifestation requires successful exercise of a number of specific capacities, and those capacities – which are not all inherently trait-relevant – might be more fragile for some people than we assume. We readily accept that sleep deprivation, hallucinogenic drugs, various mental illnesses and phobias are the kinds of things that can prevent trait-relevant behaviour from manifesting without acting as evidence that one lacks the trait. But there may be other factors that are also capacity-compromising that we simply have not identified yet, and some agents are much more susceptible to certain factors than others. Not all fragility is alike.[5] Arguing that the situationist experiments are trait-eliciting conditions because global traits ought to be robust enough to withstand the factors present in those experiments requires situationists to provide a substantive account of which kinds of robustness matter and why, while also drawing the line between factors that are capacity-compromising and those that are trait-relevant (for any given trait) in the right place. Appealing to shared normative standards and a lack of coercion alone does not yet provide this.[6]

3. *Reconsidering the experiments*

The challenge raised above cuts both ways, however: without some positive way to distinguish capacity-compromising masks from trait-relevant features of the environment, global trait theorists are at risk of implausibly saying that, for example, Suzie's self-absorption compromises her capacity to care about others, and so settings where she is self-absorbed are not fair tests. How do we decide which factors count as capacity-compromising?

Here is how: if a certain factor not only reduces the chances of agents acting according to moral reasons in a particular setting, but also reduces the chances of agents acting according to *different* kinds of good reasons in *many* settings, that factor should be treated as capacity-compromising rather than simply part of the relevant environment. For example, it is because sleep

5 Some traits, like fortitude, plausibly do require manifestation in the presence of factors or circumstances that would be considered capacity-compromising if we were assessing most other traits.

6 Doris (2005) and others have pointed out that responses along these lines raise worries about subjects' level of practical wisdom. This might threaten some aspects of virtue ethical theories, but it does not necessarily threaten globalism, so I will set it aside.

deprivation degrades people's performance on a range of cognitive tasks, and makes them much more likely to crash their car (something even minimally rational agents would use their capacities to avoid), that we can infer that sleep deprivation compromises capacities. We can then work backwards and infer that slight sleep deprivation will slightly impede one's capacity to pay attention, for example. We can also apply this test to make sense of factors that degrade reasons-recognition. It is plausibly because people regularly fail to recognize features of their environment when they have limited time, have to multi-task or are intoxicated that we think reduced time, multiple tasks and alcohol compromise one's capacity to recognize features of one's situation. With this background knowledge about which factors generally impede one's capacity to act on reasons, we can then assess particular agents by seeing what best explains their patterns of behaviour across experimental variations and assess whether that explanation is the kind of thing that undermines one's agential capacities generally. I believe this is the case for the situationist experiments.

For example, if anything regularly undermines one's capacity to do the right thing, it is being unaware that something needs to be done.[7] Awareness is the kind of thing thar can be compromised by features of one's environment, and this seems to significantly explain the behaviour of subjects in *Bystander* and *Samaritan*. Multiple studies demonstrate that the bystander effect reduces when the need for help is more *recognizable*, for example when bystanders are clearly unable to help (see Fischer et al. 2011 for a meta-analysis). And interviews indicate that many subjects interpret their situation very differently from how we expect (Herdova 2016). In particular, they simply do not interpret it as one in which assistance is needed.

One might want to point to the success of those who do the right thing in the presence of unresponsive bystanders as evidence that those who failed to help could have helped too or, for *Samaritan*, to the fact that many subjects had noticed the person. But we have to be careful with such inferences – even if it is *possible* for one to do the right thing or recognize that someone needs aid, the *quality* of one's opportunity might be too low (Nelkin 2016). There are likely to be capacity-relevant differences between subjects that explain why some do the right thing and others fail. Subjects have different backgrounds and upbringing, and plausibly have different interpretations of their setting. These are things which are likely to affect the difficulty of doing the right thing, especially in confusing, unfamiliar contexts. Additionally, those who do the right thing might have been subject to a form of moral luck, in that their wandering attention just happened to notice certain features enabling them to interpret the situation in a particular way.

7 This is compatible with thinking that agents are sometimes culpably ignorant. See Piovarchy 2022.

At this point one might be willing to grant that *Bystander* and *Samaritan* are not trait-eliciting settings, but argue that ignorance is not a plausible feature of *Obedience*. That subjects were aware of what they were doing is precisely why they were so anxious. But in the way that recognition of reasons can be diminished due to bad evidence or unreliable social cues, other capacities required to do the right thing can be diminished too. Particularly relevant are the capacities to cease deliberating and form an intention, translate one's intentions into action and maintain one's resolve (Cohen and Handfield 2010). Like the capacity to recognize reasons, these come in degrees and can be temporarily compromised by some factors without counting as evidence that one is not compassionate. When we try to explain the relative variance in obedience rate between experimental conditions, a number of potentially capacity-compromising factors are present. Such factors include confusion, ambiguity, affordances and a lack of cues that we normally rely upon to make our way about the world (Piovarchy 2022). It seems particularly relevant that subjects are in a novel and confusing situation, and transcripts suggest that those who disobeyed had *realized* they could disobey, while many obedient subjects did not (Gibson 2013). Had subjects been forewarned what was going to happen, or been informed of the extent to which most people obey, many more would have disobeyed. This could be taken to show they had the specific capacity to disobey and simply did not exercise it. But it could also show that such factors make disobeying much less difficult, by enabling subjects to realize what options they had. Forewarned is forearmed.

Recall my suggestion that the surprisingness of the situationist studies gives us reason to question our assumptions about how particular factors affect us and the degree to which they make right action more difficult. Given subjects' success at avoiding wrongdoing in other situations (e.g. they would likely disobey if approached on the street and told to shock a stranger), it is tempting to think our surprise is insignificant and that claims the subjects temporarily lacked the capacity to avoid wrongdoing are ad hoc. But once we realize that our assessments and explanations must *also* consider patterns of failure in other situations (e.g. ambiguous settings more generally), claims of incapacity become more plausible and show that our initial surprise was indicative of mistaken assumptions. In confusing, unfamiliar and stressful situations, exercising our rational capacities is likely to be much more difficult than many of us realize.

4. *Reconsidering the situationist threat*

This argument does not fully deflate the situationist threat. A situationist could maintain that, given how virtue theorists typically talk about traits and capacities, the situationist experiments are trait-eliciting settings. Alternatively, they may argue that we have still reduced globalism's empirical content.

I do not deny that global trait theorists often talk as if they endorse a more robust conception of traits. But moving to an account that accepts that traits are not robust enough to withstand factors that compromise capacities is not much of a cost, especially if situationists have not yet provided an alternative method of distinguishing capacity-compromising factors from factors relevant to our trait assessments.[8] If we are in the business of predicting and explaining behaviour, accepting that we cannot count on people to exhibit the character traits our standards expect in novel, unfamiliar settings with high ambiguity seems to let global trait theorists maintain much of their original account.

To the extent that we have ended up with an account of traits that has weaker empirical content, this is not a concession for globalism. Unlike other replies to Doris, this has not occurred by weakening our theory of the nature of traits, virtue or how traits manifest. Instead, we have pointed to a difficulty that our proposed test failed to take into account: humans have varying and situationally contingent capacities, and this was the case before the situationist's argument was ever advanced.

University of Notre Dame
Australia
adam.piovarchy@nd.edu.au

References

Cohen, D. and T. Handfield. 2010. Rational capacities, resolve, and weakness of will. *Mind* 119: 907–32.

Darley, J.M. and C.D. Batson. 1973. 'From Jerusalem to Jericho': a study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology* 27: 100–8.

Doris, J.M. 2002. *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.

Doris, J.M. 2005. Replies: evidence and sensibility. *Philosophy and Phenomenological Research* 71: 656–77.

Doris, J.M. 2015. *Talking to Our Selves: Reflection, Ignorance, and Agency*. Oxford: Oxford University Press.

Fischer, J.M. and M. Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.

Fischer, P., J.I. Krueger, T. Greitemeyer, C. Vogrincic, A. Kastenmüller, D. Frey, M. Heene, M. Wicher and M. Kainbacher. 2011. The bystander-effect: a meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin* 137: 517–37.

Gibson, S. 2013. Milgram's obedience experiments: a rhetorical analysis. *British Journal of Social Psychology* 52: 290–309.

8 In later work, Doris (2015) provides an alternative and somewhat revisionist account of agency and examines how this intersects with questions of responsibility.

Hartshorne, H., M.A. May and J.B. Maller. 1929. *Studies in the Nature of Character*. New York: MacMillan.
Heil, J. 2005. Dispositions. *Synthese* 144: 343–56.
Herdova, M. 2016. What you don't know can hurt you: situationism, conscious awareness, control. *Journal of Cognition and Neuroethics* 4: 45–71.
Jaster, R. forthcoming. The ability to do otherwise and the new dispositionalism. *Inquiry*.
Latané, B. and J.M. Darley. 1970. *The Unresponsive Bystander: Why Doesn't He Help?* New York: Appleton-Century Crofts.
Milgram, S. 1974. *Obedience to Authority: An Experimental View*. London: Tavistock Press.
Nelkin, D.K. 2016. Difficulty and degrees of moral praiseworthiness and blameworthiness. *Noûs* 50: 356–78.
Piovarchy, A. 2022. Situationism, capacities and culpability. *Philosophical Studies* 179: 1997–2027.
Rudy-Hiller, F. 2020. Reasonable expectations, moral responsibility, and empirical data. *Philosophical Studies* 10: 2945–68.
Webber, J. 2006. Virtue, character and situation. *Journal of Moral Philosophy* 3: 193–213.