Against Reductivist Character Realism

Abstract: It seems like people have character traits that explain a good deal of their behavior. Call a theory character realism just in case it vindicates this folk assumption. Recently, Christian Miller has argued that the way to reconcile character realism with decades of psychological research is to adopt metaphysical reductivism about character traits. Some contemporary psychological theories of character and virtue seem to implicitly endorse such reductivism; others resist reduction of traits to finer-grained mental components or processes; and still others remain silent on the metaphysics of traits. In this paper we argue that character realists do not have to, and in fact *should* not, be reductivists. We introduce a theoretical dilemma for reductivist character realism. Then we explain how nonreductivists can meet the standards for empirical adequacy laid out by Miller and others. Further, we argue, hylomorphic nonreductivism avoids the theoretical dilemma that threatens reductivism. It also fits nicely the major commitments of recent models of virtue in psychology. Thus, character realists should not be reductivists.

Keywords: virtues, situationism, mixed trait theory, generosity

Mother Theresa was compassionate; Martin Luther King, Jr. was just; Malala Yousafzai is courageous. We commonly attribute character traits like virtues to people, and not just to the morally exceptional: my friend is honest; your students are persevering; our colleague is fastidious. The folk assumption is that people have character traits and those traits largely explain their behavior. Call a theory *character realism* just in case it vindicates this assumption.

Whether scientific study corroborates this folk assumption is a matter of debate. Psychological studies on behaviors associated with character traits like honesty have cast doubt on character realism. The cumulative weight of some empirical studies led many psychologists in the late 20th century to abandon character realism in favor of situationism. The latter is the view that behavior is more a function of features of our situation than something flowing from stable character traits like traditional virtues or vices. Situationism is troubling not just because of what it would mean for our folk intuitions or realist theories of character: it threatens to disturb normative ethical theories that depend on the supposition that a certain set of character traits, namely the virtues, are psychologically realistic goals attainable and necessary for living a good life.[1] If we want to build a normative ethical theory on rock instead of sand, we should make sure the underlying theory of character is not undermined by our best psychological evidence.

Simultaneously, character realists should be careful not to introduce *other* problems into the foundation in constructing an empirically adequate theory of character. Christian Miller has argued, in one of the most impressive and sustained recent attempts to defend character realism, that the most viable way for character realists to maintain empirical adequacy is to adopt a certain metaphysical picture of character traits. In this essay we will argue that in fact, character realists in psychology and philosophy should not be reductivists.

First we offer a taxonomy of metaphysical views a character realist could have, since this has yet to be done in the literature and yet such a taxonomy demonstrates that there are viable options besides reductivism. Second, we show that two influential accounts of character, i.e., Miller's Mixed Trait Theory (Miller, 2013, 2014, 2018) and Fleeson and Jayawickreme's Whole

[1] Virtue ethics is not the only normative ethical framework that requires the claim that humans can acquire settled character traits. For instance, recent work on Kant's *Metaphysics of Morals* highlights the importance of virtues for Kantian deontology (e.g. Stohr 2018). For this paper, we put to the side the suggestion made by Alfano (2013) that false attributions of character traits might actually serve as a moral technology for virtue development.

Trait Theory (Fleeson and Jayawickreme, 2015), embrace the reductivist metaphysics of character. Third, we introduce the theoretical dilemma for reductivist realism which threatens to make the account of character either circular or unable to account for commonsense judgements about cases of character traits in action. Fourth, we argue that the hylomorphic nonreductivist model of character provide a more viable alternative to reductivist models while being able to meet the standards for empirical adequacy of a theory of character. We support our argument by demonstrating that several other empirically respectable models of virtue and character in recent psychology can be combined with nonreductivism (Schnitker et al. 2019, Darnell et al. 2019, Cole-Wright, et al. 2020, Fowers et al. 2021, Ng and Tay 2020).

1 The Metaphysics of Character Traits: A Taxonomy

Character realism holds that people have personality traits that largely explain patterned behaviors subject to moral evaluation. *Character traits*, traditionally understood, are coarse grained in the sense that they cover a wide range of behaviors. Some have called these global traits (Doris 2002). For instance, a global trait of honesty is associated with behaviors such as promise-keeping, truth-telling, not stealing, not cheating. Character realists attuned to empirical research also talk about more fine-grained mental state dispositions that play a role in generating morally salient behaviors. These are dispositions to form desires and beliefs about a narrower domain (e.g., cheating) or in more specific circumstances (e.g., under threat of being caught) (Miller 2014, 5).

How do these narrow and global traits relate to each other? As we will see, a lot in virtue theory and measurement hangs on how we answer this question. The problem is that this relationship can be construed in a variety of ways, and, so far, the literature has no explicit taxonomy of these possibilities. Before we move on with our argument, we create a taxonomy of

metaphysical views about character. We use two questions to generate the relevant distinctions – the Grounding Question and the Counting Question.

The first question is:

Grounding Question: What is the order of grounding between dispositions to believe $p_1 ... p_n$ and desire $d_1 ... d_n$ and character trait T, where both are involved in the production of some behavior typical of persons with T?

The Grounding Question helps us distinguish *what underlies what* when it comes to trait-level dispositions and fine-grained mental state dispositions.

Character realism is compatible with several different answers to the Grounding Question, such as:

Reductivism: Dispositions to desire $d_{1...n}$ and believe $p_{1...n}$ *constitute* T; the person's having those dispositions explains the person's having T.

On the reductivist view, the fine-grained mental state dispositions underlie character traits. For example, say that Krista is aggressive. Reductivism claims that what grounds Krista's aggression is her fine-grained mental state dispositions, like the disposition to form beliefs that using physical force will get someone to do what you want, that it is okay to fight someone if they make you mad, or that it is okay to fight someone if they call you names or tease you (Farrell et al. 2008, 2010).

A character realist could also answer the Grounding Question with:

Nonreductivism: A person's possession of trait T explains fine-grained dispositions to desire $d_{1...n}$ and believe $p_{1...n}$.

On this view, even if dispositions to form specific desires and beliefs proximately cause behavior, what *ultimately* causes the behavior is the trait. If Krista has the trait of aggression, she is disposed to form beliefs about needing to fight and to form desires to fight when someone teases her in

virtue of having aggression. The nonreductivist might not even think of these fine-grained mental state dispositions as having any real ontological status. What's distinctive about nonreductivism is the commitment to the coarse-grained character trait being the ultimate explainer of the associated behaviors. Aggression is not a label for, or mental thing that supervenes on, a cluster of more fine-grained mental state dispositions; it is more metaphysically basic.

Here is a second important metaphysical question about character traits:

Counting Question: How many things are there in the mind in the final metaphysical account?

This question requires the character theory to make her ontological commitments about traits and more fine-grained mental states explicit. Does she truly think character traits *exist* as something over and above fine-grained mental state dispositions? Or does she think trait just is the cluster of certain dispositions, such that it is appropriate to *label* certain clusters of mental state dispositions using trait terms?

We might hold:

Monism: Trait T is *identical to* those dispositions to desire $d_{1...n}$ and believe $p_{1...n}$ that compose T.

Combined with reductivism, this is quite an ontologically parsimonious view. If the trait, e.g. aggression, does no causal work beyond what the fine-grained dispositions do, then why countenance the trait in our final metaphysical count as anything distinct from those dispositions? We might think that a commitment to the existence of aggression just is a commitment to the existence of all the fine-grained dispositions that compose it. Further, if psychologists are already happy to countenance fine-grained mental state dispositions, this version of realism seems more likely to deliver on a promise of empirical adequacy.

Contrast monism with dualism:

Dualism: Trait T is *constituted by*, but *not identical to* the dispositions to desire $d_{1...n}$ and believe $p_{1...n}$.

Dualism claims that a trait like aggression exists when and only when certain fine-grained mental state dispositions, like the disposition to form beliefs about the efficacy of physical force to get what you want and the disposition to find the goal of power salient, exist in someone's psychology. Depending on the dualist's answer to the grounding question, she might hold that the trait exists *in virtue* of the fine-grained dispositions but is an emergent entity or property of persons, or that the fine-grained dispositions exist and the trait-level property exists but neither is identical to the other or grounds the other.

Here is a third view:

Hylomorphism: Trait T exists in its own right; component fine-grained mental states or mental state dispositions *exist as a material parts* of T; component goals or orientations of the agent *exist as formal parts* of T.

The hylomorphic view gives an idiosyncratic response to the Counting Question. This is because it assumes that the question "what exists?" is somewhat malformed. There are various sortals or categories of being in the final metaphysical count even though entities in some category might be privileged. There are entities like traits, or to take a standard example from metaphysics, a clay statue; but there are also things that exist *qua matter*—e.g. the clay, which cannot exist by themselves but do give the existence and persistence conditions of other entities-- and things that exist *qua form*—e.g the statue's shape, which do not exist independently but really exist as parts of the entities they structure. So in one sense, three things exist: T (e.g. statue), the form of T (e.g.

the statue's shape), and the matter en-formed by T (e.g. the clay). But they exist in very different ways-- qua object, qua form, qua matter.

All this metaphysical taxonomizing might seem taxing, but it matters more than has been noticed. For the metaphysical view of character traits one holds, implicitly or explicitly, will determine how one goes about conceptualizing, individuating, and constructing measures of those traits in the process of operationalizing them. A monist reductivist can start developing a construct by collecting instances of fine-grained dispositions to desire and believe specific things in particular circumstances, and aggregating them into a group to form a construct (e.g. gratitude) and then generating items testing for the fine-grained dispositions as a way to measure that construct (e.g. "list the things you are grateful for"). But nonreductivist will go about things in the opposite direction, first defining the features of the trait and generating items by looking at clusters of beliefs, desires, and behaviors we could expect to flow from that trait as defined. Nonreductivism would allow for more variability in predicted behaviors across situational contexts and social roles than reductivism.

2 The Reductivist Realism of Mixed Trait Theory and Whole Trait Theory

Most recent theories of character and virtue do not engage directly in discussions about the metaphysical questions above. In fact, Miller says such issues are "almost entirely neglected" in the literature, and we think he is right. Miller's Mixed Trait Theory (MTT) offers a refreshing exception, which allows us to characterize his account of character as fitting neatly into the monist reductivist categories. Fleeson and Jayawickreme also make many remarks in discussing Whole Trait Theory (WTT) that are strongly indicative of a reductivist metaphysics. The measures WTT uses to test for character traits also suggest a reductivist view.

Miller unambiguously defends monist reductivist account of character (Miller 2014, 24). There he addresses what we are calling the Grounding Question and the Counting Question. First, he claims that fine-grained mental state dispositions metaphysically ground character traits, and are causally responsible for the characteristic behaviors we tend to attribute to them:

> What immediately underlie a trait disposition are further dispositions, more precisely the dispositions to form one or more occurrent mental states which are relevant to the character trait. This underlying causal base of the disposition, in other words, includes dispositions to form certain trait-specific beliefs and/or desires in the person's mind. (ibid, 26)

We see this reductivist commitment emerge again in *Moral Character: An Empirical Theory*:

> A character trait disposition which is had by a person consists of some cluster of her relevant interrelated mental state dispositions such that necessarily, if she has this cluster of dispositions, then she instantiates that character trait as well. (Miller 2013, 10-11)

He then carefully explains how we can reduce traditional moral character traits and morally "mixed traits" to their fine-grained dispositional constituents.

For example, he describes a Global Helping Trait (GHT) as "a disposition constituted by a certain cluster(s) or mental states - beliefs, desires, intentions, and the like - which mediates the relationship between the presence of a trigger on the one hand, and elevated or reduced helping behavior on the other" (Miller, 2010, 4). He goes on to explain that although various psychological factors - like being in a good or bad mood, feeling of embarrassment, or empathy, or anger - can

trigger or inhibit the GHT, they should not be considered constituents of the GHT, but rather external enablers or inhibitors.

What motivates this reductivism? Miller argues that it provides the *only serious answer* to questions of metaphysical grounding. Regarding explanatory power, he writes, "Explaining Jones's compassionate thoughts and actions in terms of his being a compassionate person does not count as a very helpful or enlightening explanation by itself," (ibid, 25). To really explain behavior in terms of character traits, he says, we need to hold "that trait dispositions are not ungrounded themselves, but rather *exist in virtue of certain underlying* properties which enable them to be instantiated…" (ibid). With respect to empirical respectability, while Miller is committed to there being some entity that vindicates our folk assumptions about character, he is also concerned to make sense of the best of our psychological data (2013, 24). The way to make good on both commitments in his estimation is to account for character traits in terms of *just those psychological units* that seem immediately causally responsible for observed behaviors in specific situations according to psychological studies (2014, 139).

Miller claims that such dispositions provide an appropriate reduction base in a theory of character because they are "familiar mental categories" in mainstream contemporary psychology and so constitute a "satisfying explanation" (2014, 28). The standard for empirical adequacy at play here is explanation of the distribution of behaviors in various psychological studies of trait-behaviors such as honesty, compassion, and helping.

With respect to the Counting Question, Miller espouses monism. The trait constituted by a cluster of fine-grained mental state dispositions, he says, is "simply *identical to* the mental state dispositions which underlie them," (2014, 29, our emphasis). Later, he introduces dualism only to

dismiss it quickly as objectionably mysterious, and hylomorphism gets no consideration. He argues that metaphysical parsimony favors monism.

Fleeson and Jayawickreme's WTT also implies reductivism about character. However, theirs is not a monist, but a dualist version of reductivism. WTT of personality traits is meant to extend to character traits, including the traditional virtues (2017). WTT first posits personality *states*, which resemble traits in their descriptive content. For instance, Emma might exhibit an extraverted state such that she is talkative, assertive, and energetic. Unlike traits, which are supposed to be temporally stable over a long period of time, states last for a short period of time. Emma will "enact" various states over time, and we can aggregate the states enacted to form what Fleeson and Jayawickreme call a *density distribution* of Emma's states. If we plot all of Emma's states with respect to a description like generosity over the course of two weeks, we might see her fluctuating between very high and medium states over time, and we might be tempted to infer a lack of stability required for possessing a high degree of generosity. If we compare Emma's density distribution to someone else's, however, we probably can see significant differences between the centers of their density distribution of states. Ingrid's density distribution center is between a medium and low state of generosity. This indicates a real between-person difference in how generous Emma and Ingrid are. Whole Trait Theory holds that traits are best described as someone's density distributions with respect to some dimension of personality over time (Fleeson and Jayawickreme 2015).

On WTT, character traits have a descriptive side and an explanatory side. The descriptive side is comprised of a person's "moral behaviors, feelings, and thoughts" while the explanatory side consists in the "various mechanisms… undergirding" these behaviors, feelings, and thoughts (2017). Remember that the Grounding Question asks *what explains what* when it comes to

character traits and fine-grained mental states or processes. WTT's answer to the Grounding Question is this: social-cognitive mechanisms and processes underlie and cause personality states that make up density distributions, or traits (Fleeson and Jayawickreme 2015).

Since proponents of WTT say the fine-grained and temporally short mechanisms and processes explain traits, they are reductivists. This commitment is evidenced further by claims they make about causation: "The existence of stable differences in descriptiveness implies logically that there must be some explanation *responsible for* those individual differences," (Fleeson and Jayawickreme 2015, 84, our emphasis). WTT holds that what explains the differences between individuals at the trait level are the finer-grained social-cognitive processes and mechanisms, such as particular goals or self-regulatory plans:

> Several processes, including social-cognitive processes described (but not identical to) CAPS, are the determinants of states (Fleeson & Jolley, 2006). These processes include interpretative processes, motivational processes, stability-inducing processes, temporal processes, and random error processes. (ibid., 86-87)

The goals, plans, and beliefs that issue in mental processes and resultant behavior described here can be momentary and subject to drastic change due to situational influences over time (ibid. 87). "In fact, McCabe and Fleeson (2012) found that 50–75% of the variance in personality states was predictable from the goals people were working on at the moment," (ibid.). On this picture, then, character traits are grounded in mental properties— like having a goal of being liked by people at a party—which can activate processes that together issue in a personality state of, say,

extraversion. Over time, the many fine-grained personality *states* generated by that and similar processes aggregate to form a coarser-grained pattern we can label as a trait like high extraversion.

Fleeson and Jayawickreme's reductivism, like Miller's, seems motivated largely by a desire to fit the empirical data used by situationists. Fleeson and Jayawickreme argue that the explanatory side of traits helps account for the situationist observations and the descriptive side allows them to countenance real character traits: "While the evidence for Whole Trait Theory verifies some arguments proposed by situationists, it also strongly supports the trait perspective. That is, consistency does exist, but in the distribution as a whole and not in single states," (2017, 523). So again, the bar for empirical respectability is set at being able to predict or explain the behaviors observed in the psychological studies typically used to support situationism.

When it comes to the Counting Question, WTT's defenders espouse dualism. They write:

> Adding an explanatory account … creates two parts to traits, an explanatory part and a descriptive part, and these two parts are distinct entities that nevertheless can be joined into whole traits because one of the parts is the causal consequence of the other part. (2015, 83)

> The explanatory side of traits causes the descriptive side of traits. They are separate parts of the same trait. Nonetheless, they belong together, because they stand in a very direct and mutually logically necessary relationship to each other. The explanatory part implies a descriptive part as its direct output, and the descriptive part implies an explanatory part that produced it. A full account of traits must include both parts, separate and individually important, but working together. (Fleeson and Jayawickreme 2015, 84)

They also say that when we talk about traits and trait levels, we "refer to more than the mechanisms explaining the level, but also to the distributions of states on which people differ," (ibid.). That is, if I say that Ingrid is introverted, I aim to represent a reality in the world, the density distribution of her extraversion states, and not simply to point to all the social-cognitive mechanisms at various points in time that cause the states composing the distribution. That means there must be something distinct from, but reducible to, the mechanisms that I can refer to by talking about her introversion. And that is exactly what the dualist reductivist would say: the traits really exist, as do their underlying finer-grained components, but that while the fine-grained mechanisms constitute the trait, the trait is not identical to the parts that compose it.

In this section we identified MTT and WTT as versions of metaphysical reductivism. We also saw that the impetus for reductivism is that it is supposedly *uniquely affords empirical adequacy to character realism*. We will return to the question of whether this is true in section 4. For now, we turn to the theoretical problem for reductivist realism.

3 The Dilemma for Reductivist Realism

Here is the dilemma for realist reductivism in a nutshell: The reductivist is committed to the claim that whether a person has a trait depends only on whether she has all the mental dispositions, or engages in the psychological processes, that form the reduction base of a character trait. This results in either a circular account, or an account with counterintuitive verdicts on cases where the character trait has normative or moral significance. Both are devastating problems for a model of character meant to be used to support normative theories like virtue ethics.

3.1 The Dilemma

Let us now fill out the dilemma. We begin by reiterating the reductivist realist's response to the Grounding Question:

1. That someone has a character trait T is explained by the existence of more fine-grained mental states dispositions to form desires and beliefs which constitute T.

Next, we note that any good theory of character needs to be able to explain why some set of dispositions (or distribution of processes, mechanisms, and personality states) composes the trait it does, rather than some other trait. So, the reductivist must be able to provide such as explanation, as well:

2. The reductivist must have an account of why *those* dispositions constitute *that* trait. (Differentiation Requirement)

Here is an example to flesh out the Differentiation Requirement. Suppose for example that Mother Theresa is honest. The reductivist understands her honesty in terms of her fine-grained mental state dispositions, like the dispositions to form desires to avoid stealing even when she can get away with it, dispositions to form beliefs that she should not lie when it promotes good feelings in another person, and so on, or in the case of WTT, various personality states indicative of medium to high honesty generated by psychological mechanisms, like the goal of gaining trust of someone she is serving, over time. The reductivist will need to explain why *these* dispositions or mechanisms belong to *honesty* and not to some other trait, like compassion. In general, the theory should provide a principled way to partition the fine-grained dispositions in the reduction base of a trait, as well as an explanation for the partition (Russell 2012, West 2016, Beary 2019).

When providing an explanation that satisfies this Differentiation Requirement, the reductivist will either make reference to the trait supervening on the fine-grained mental state dispositions or not. That is:

3. Either the reductivist explains why dispositions (or mechanisms) $d_1 \ldots d_n$ belong to a trait T by referencing T or without referencing T.

This leads to the first horn of the dilemma. For if the reductivist explains why the dispositions $d_1 \ldots d_n$ compose a character trait T by referring to T, the account of T will be circular. The definition of honesty, if it includes having a disposition to tell the truth, cannot explain why Mother Theresa's disposition to tell the truth is part of honesty. But if the definition of the trait non-circularly explains why Mother Theresa is disposed to tell the truth, the account will certainly not be reductive.

Put succinctly:

4. If the reductivist appeals to T in her account of why certain mental state dispositions (or mechanisms) $d_1 \ldots d_n$ constitute T rather than some other trait, the account is circular. (First Horn)

However, if the reductivist attempts to avoid circularity, she will run into the dilemma's second horn:

5. If the reductivist identifies the dispositions without reference to T her account will face counterexamples where the specified combination of $d_1 \ldots d_n$ fails to produce T or where a person seems to have T yet lacks one or more of $d_1 \ldots d_n$ that compose it. (Second Horn)

This second horn suggests two types of counterexamples: *false positives* and *false negatives*. The false positive counterexample would show someone with all the dispositions in the reduction base for the trait who nevertheless fails to have the trait; thus, the positive result given by the reductive account would be false. The false negative counterexample would show someone who appears clearly to have the trait but does not have all the previously specified dispositions that are supposed to compose it; thus, a negative result given by the reductivist would be false.

Consider an example of a false negative issued by MTT. Miller might analyze the virtue of temperance in terms of dispositions to desire moderate amount of food, form beliefs about when to eat and stop eating based on when one has had more than the moderate amount, desires to not eat when one is not hungry, desires to consume simple and nutritious food and drink as opposed to indulgently decadent food and drink. The vice of gluttony, on the other hand, would be constituted by dispositions to indulge in the pleasure of eating, to continue eating beyond the point when one is full, and so on. What would such an account make of the following character?

> **Hermit:** John Cassian in *Apophthegmata Patrum* 13.3 recalls that, during his visit to the Egyptian Desert Fathers, he was surprised to find that the famous hermits were willing and ready to break their usually strict fasting habits for the sake of showing the travelers hospitality. He remembers one hermit in particular: "He invited us to eat, and though we had eaten he urged us to eat more. I said I could not. He replied, '*I had already given meals to six different visitors, and I have eaten with each of them, and I am still hungry. And you who have only eaten once are so full that you cannot eat with me now?* '"

Cassian tells this story precisely because he finds Hermit admirable in the way he uses his control over his appetite in such an unusual circumstance; that is, he holds up Hermit as having extraordinary temperance. According to the analysis that would be given by the reductivist theory, however, we would need to conclude that Hermit lacks temperance. For Hermit's behavior in the case shows that he fails to have the disposition to eat the moderate amount of food and the desire to not eat when he is not hungry. It looks, therefore, that the reductivist theory gives a false negative, on a very important instantiation of temperance no less.

In response to the second horn, a reductivist may say that Hermit's actions manifest not the virtue of temperance, but the virtue of hospitality. This response, however, remains unsatisfying because it seems like the dispositions that constitute temperance are still there, even if the arrival of the guests activates dispositions that constitute hospitality. In other words, the hermit does not become a glutton by acting hospitably. Here, reductivism struggles to meet the Differentiation Requirement. Cassian wants his readers to re-imagine *temperance* as a virtue that enables the monk to fast *and* to feast when guests are present. Reductivism specifies *in advance* which combination of fine-grained mental states dispositions constitutes trait T such that it cannot recognize the instantiation of that trait that is somewhat unexpected in the story—the activity of feasting and not just of fasting. Note, too, that the reductivist will be forced to admit conflict between the virtues (see Russell 2009, Darnell 2019).[2]

To avoid the second horn, the reductivist might be tempted to specify which dispositions count as parts of the trait temperance by reference to some normative standard of appropriateness that will allow her to say that Hermit's consumption of the meals in this unusual circumstance is appropriate and so temperate (e.g., Hermit's temperance is composed of the disposition to eat the appropriate amount of food in any situation). But how are we to analyze appropriateness and overindulgence without reference to the trait, temperance, or the normative standard embedded in a broad goal that is best characterized at the trait level? Any normative standard that will allow us to countenance Hermit's dispositions to desire and consume "appropriate" amounts of food or drink are just going to end up being either equivalent to or derived from the normative standard

[2] Thanks to [removed for anonymous review] for pointing out this further problem with reductivism created by such examples.

for the *trait*, temperance.[3] Thus, the character trait is doing ontological work, rather than the fine-grained dispositions that are supposed to ground the trait.

Miller's concerns about theoretical parsimony might even invite a worry about countenancing fine-grained dispositions at all. For if we can explain Hermit's behaviors by reference to the trait-level orientation toward, or goal of, maintaining a healthy body and control of appetites, then it is unclear what causal work is done by finer-grained states that could as easily be described as situationally specific enactments of the broad goal.[4]

3.2 Applying the Dilemma: Generosity

We have presented the dilemma for reductivist realism and offered just one concrete example. But suppose the reductivist replies, "Hermit shows that the proposed reductive base for temperance is incorrect, *not* that it is not possible to give a correct, non-trivial reductive base for temperance." Fair enough. There is probably a better reductive account of temperance than our toy account. But could a better reductivist account escape the dilemma? We think the prospects are bleak. In support of this reply, let us show how Miller's carefully developed and defended account of generosity falls prey to our dilemma.[5]

---

[3] For instance, suppose the reductivist says Hermit has temperance because he has the disposition to form desires for the appropriate amount of food when hosting guests. The "appropriate" amount cannot simply be a quantity equivalent to three meals (since there are obviously situations where that would be gluttonous– for instance, when he hosts only one guest, or a guest who is fasting for Ramadan). Instead, "appropriateness" will need to be cashed out in reference to some goal like maintaining a healthy body and control over one's appetites. But once we refer to this broad goal to draw the line between gluttonous and temperate desires, it seems we are simply applying the trait-level normative standard for temperance; namely, being oriented toward the goal of control of the appetites in the domain of food and drink.

[4] Here we might draw on psychological models of characteristic adaptations of dispositional traits from McAdams and Pals (2006).

[5] Miller crafts a perspicuous account, making compelling arguments for each element and revising parts of the analysis along the way to accommodate pre-theoretic intuitions about a wide range of cases (Miller 2018). If any reductive account of generosity is going to work, we think, this one should. We will show, however, that it too admits of both false negative and false positive counterexamples due to the reductivist commitment and rejection of the idea that phronesis, or practical wisdom, serves as a controlling or architectonic virtue.

Miller's account of generosity starts with the behaviors the trait is supposed to produce, that is, generous acts. Generosity then gets analyzed in terms of the dispositions to form the mental states that will result in generous acts in a stable and cross-situationally consistent manner (ibid 216-217). Miller explicitly distinguishes his preferred analysis from the method of analyzing virtuous acts in terms of what the fully virtuous person would do in some circumstance (ibid 219).[6]

Here is the final analysis of the virtue of generosity:

> A person has the virtue of generosity only if she has the dispositions to form beliefs and desires that generate actions in a variety of different relevant situations and stably over time with the following qualities:
>
> (i) What is bestowed by the action is of value to the giver.
>
> (ii) The actions are motivated by an ultimate desire that is altruistic, and in the case of mixed motives this desire is primary and capable of leading to these actions even in the absence of other motives.
>
> (iii) The actions are morally supererogatory and not morally required. (ibid 231)

Miller's account of generosity avoids the first horn of the dilemma. It does not make mention of generosity in any of the dispositions to which the trait is reduced, nor does he appeal to generosity explicitly in explaining why these fine-grained dispositions belong to generosity

[6] One announced motivation for parting ways with the common method here is that "there are conditions that must be met by [the agent performing a generous act] that have nothing to do with how a virtuous person would behave in her circumstances," (ibid). We, too, think it is possible for an agent to perform a generous act and not do it from the virtue of generosity. However, we do not think this enables us to understand what makes an act type generous without reference to the virtue of generosity. On our reading of Aristotle, it must be the case that the moral learner can perform generous acts before having the virtue of generosity just like it must be the case that an unskilled flautist can perform musical acts before having the musical skill or excellence she builds by practicing those very acts. The difference between the learner and the person with the excellence consists in the end or goal they have internalized and from which they act. Hence, as Aristotle says, "virtue makes the goal right."

rather than others, though he does appeal to "generous" acts. We think the appeal to generous acts is unproblematic so long as he analyzes acts as generous by reducing the property of being generous to (i)-(iii). He does not build in the normative standards of generosity in the fine-grained dispositions that compose generosity (e.g. "the disposition to give to charity when appropriate"). The defense of generous acts and generosity proceeds mostly by way of examples and intuitions.

Unfortunately, Miller's account doesn't fare as well when it comes to the second horn. He claims to have given a preliminary account of generosity by providing necessary, but not sufficient, conditions for the trait of generosity (235). This means that the primary worry for his account will be with its yielding false negative counterexamples.

Such counterexamples are not hard to find, though. Consider:

> **Mother Theresa:** Mother Theresa is convinced that God has called her to give up a normal life and ordinary comforts to serve the poor on the streets of one of the most impoverished parts of India. She experiences many demands on her time, and thus is not disposed to act in ways that seem to her to go beyond what is required of her, as it might lead her to shirk her duties. Nonetheless, she gives nearly all of her waking hours to caring for the poor in this province.

It seems obvious that Mother Theresa is generous. But she fails to meet Miller's necessary condition (iii) because she does not form desires and beliefs that lead her to do what she considers *morally optional*. She sees her own vocation not only as supremely important, but as something that is directly commanded by God and thus required.[7]

---

[7] To be fair, Miller allows the possibility that "acting from generosity, at some point or other during one's life, is morally required" (227, note 23). He even allows that, in particular circumstances, generosity might require a person

We can generate another false negative counterexample based only on conditions (i) and (ii):

> **Sergio:** A local restaurant owner, Sergio, is fiscally conservative with his business's earnings. He is committed to helping people in his community but wants to make sure he does not give so much of their profits to local charity or individuals in need that it is costly to his business. Further, he does not himself value a life with luxury items, as he thinks a good life is one in which a person contributes significantly to the community. Sergio comes up with a plan to give small amounts on a regular basis, never giving an amount of great value to him or his business.

Sergio has orchestrated both personal life and business plans around benefiting others for their own sakes. Yet he does not ever perform single acts such that he gives what is of significant value to him. Compared with other local business owners, Sergio clearly looks generous, but not so on Miller's account since he fails to meet condition (i).

Sergio and Mother Theresa provide false negative counterexamples to Miller's judiciously-fashioned reductive analysis of generosity. The dispositions Miller includes in that analysis are well defended. In fact, we do not think another reductive theory could do better. Instead, we think the problem is not with the way he constructs the reduction base of generosity, but with *reductivism*.

Eventually, a complete reductive analysis of generosity must supply sufficient conditions for having a trait in terms of the underlying social-cognitive mechanisms or fine-grained mental

---

to help, "*so long as this does not entail that one is morally obligated to help*" (Ibid., our emphasis). Notice, however, that Mother Theresa performs acts she clearly takes to be morally obligatory, yet we would be hard-pressed to deny that her acts are acts of generosity and mercy. Moreover, Mother Theresa is not unique in believing that she is morally obligated to help the starving person on her doorstep. Many Christians, as well as people of other faiths, see it as their moral duty to help those in immediate, real need if they can do so without endangering the wellbeing of their dependents (see, for example, Aquinas, ST II-II.32.5). Yet these same faith practitioners understand themselves to be pe rforming acts of almsgiving - and so of generosity - when providing for the starving persons.

state dispositions. Once sufficient conditions are proposed, we suspect the analysis will fall prey to counterexamples of the false positive type because we can always imagine cases analogous to Wrong Kind of Reason cases.[8] Consider an illustration of our point from Dickens's novel *Bleak House*:

> Mrs. Jellyby: Mrs. Jellyby spends her time in various philanthropic pursuits, one of which is setting up a new colony in Africa. Meanwhile, her family is completely neglected. Her children are hungry, dirty, wild, and ignorant, and her husband is driven almost to distraction by financial and domestic turmoil that stems from Mrs. Jellyby's "telescopic philanthropy."

Mrs. Jellyby meets all the necessary conditions for generosity, but nobody who has read Dickens' account would call her generous, since her supposedly great philanthropic actions bring immense misery to those who rely on her the most. [9]

4 The Empirical Adequacy of Nonreductivism and Recent Models of Virtue

We promised at the outset to show that character realism can meet reasonable conditions on empirical adequacy without committing to reductivist metaphysics. Now it is time to make good on that promise. To do so we first review what we call the destabilizing data—evidence from psychological studies that theorists like Miller, Fleeson, and Jayawickreme have been keen to accommodate. Second, we examine the strategies deployed by the reductivist realists to handle the destabilizing data. Third, we show that the reductivists' main strategy for reconciling character

---

[8] Discussions of moral worth and acting for right reasons are replete with such cases; see for instance, Arpaly (2001), Markovits (2010), Howard (2021).

[9] Mrs. Jellyby, on a standard Aristotelian analysis, cannot be generous because she lacks practical wisdom. Practical wisdom is one candidate for the psychological feature that unifies not just dispositions that compose a trait, but also character traits themselves (Darnell et al. 2019). We are grateful to an anonymous reviewer for pressing us to make this point.

realism with the data relies not on *reductivism*, but rather on a rejection of the idea that traditional virtues and vices—or at least high degrees of these traits—are widespread. Since nonreductivists can make the very same move, nonreductivism can in principle meet the conditions for empirical adequacy. We further support this by arguing that several recent psychological models of virtue are compatible both with the data and with hylomorphic nonreductivism.

4.1 The Destabilizing Data

According to Miller, two conceptual commitments of character realism seem threatened by psychological studies of character and behavior. The first is:

> Stability Over Time: To the degree that a person has a character trait, she will exhibit characteristic behavior in multiple cases of the same situation over time. (Miller 2013, 92)

The second commitment is to consistency of behaviors produced by character:

> Cross-situational Consistency: If a person has a character trait to a high degree, she will manifest that trait in behavior across different kinds of situations. (Miller 2013, 92).

If Krista really possesses the trait of patience, then we would anticipate that she would be patient at work and at home. And if she exhibits highly patient behavior at work on Mondays, but by Fridays she displays low levels of patience, we are unlikely to call her highly patient. That is because we assume that, in order to have a character trait, a person has to exhibit a certain sort of stability over time. Stability and Cross-situational Consistency are *conceptual* claims. They state what something has to look like to count as a character trait *at all*. So let us treat them as fixed.

What psychological research threatens to destabilize is the metaphysical claim of character *realism*, namely, that the concept of character traits here described is actually *instantiated* in the world by ordinary people. Moreover, character realists cannot simply dismiss psychological studies on the grounds that virtue is rare (Kamtekar 2004). For character realism to support normative ethical theories for which the possession of virtue is important, it must support the idea that the psychological mechanisms for habit formation are readily available, that people in fact develop character traits and act from them. That is, it must meet what Owen Flanagan calls the "Principle of Minimal Psychological Realism" (Flanagan 1991, 32). In other words, virtue ethics is beholden to the best evidence of what we are like and what we can be (Stichter 2018). If there are validated psychological studies that cast doubt on the idea that most of us can acquire and behave on the basis of character traits generally, that would be bad news for character realism and virtue ethics since virtues are a type of character trait.

What psychological data casts the shadow of doubt on character realism? Fleeson et al. in categorize it into three types of studies (Fleeson et al., 2015, 44-45):

> Type (1) studies challenge cross-situational consistent traits, where, for instance, honest behavior varies widely across situations (Hartshorne and May 1928). Subsequent studies have found little within-person consistency in behavior we would describe as honest across different situations (e.g., Vernon 1964, Mischel 1968, Kouchaki 2014).
>
> Type (2) studies question the role of character traits in people's manifested behaviors by showing situational features to be better predictors of behavior than traits (Miller 2013, Darley and Batson 1973, Aquino et al. 2009). For instance, being in a negative or positive mood or the absence or presence of bystanders robustly predicts behavior associated with helping and compassion.

Type (3) studies exhibit errors in attribution of character traits to people (Jones and Harris, 1967; Ross 1977). When situational factors seem to clearly robustly influence people's behavior, people tend to disregard situational factors and attribute behavior to character.

Together, this body of empirical work suggests that what we think of as character traits might exist mainly as figments of our imagination. Moreover, the "fundamental attribution error," (so named by Ross (1977)) raises the troubling possibility that character attribution (including self-attribution) might be little more than post-hoc rationalization of behavior that was really induced by situational factors. If behavior is better predicted by situational factors, and if we have good evidence that we misattribute character traits to people, it looks plausible that our folk assumptions about the reality of character traits may be incorrect.

4.2 Two Reductivist Strategies

WTT and MTT purport to show that the results of these types of studies are compatible with the view that character traits are possessed, widespread, and account for behavior. If a particular way of conceptualizing character traits can accommodate the data in type 1, 2, and 3 studies, then we can hold onto realism without eschewing evidence from such studies (Miller 2014). We mentioned that Miller sees reductivism as a key part of the character realist's response to the situationist challenge. But on closer look, in fact, it is another part of Miller's strategy (and that of Fleeson and Jayawickreme) that does the main work.

To see that this is so we first need to distinguish between two conclusions that might be drawn from the body of psychological literature on types 1, 2, and 3 studies. Here is the first conclusion:

Negative Claim: The large body of the experiments referenced above is incompatible with the widespread possession of traditional virtues and vices.

Here is the second:

Positive Claim: At best, our manifest behavior is a product of situational forces, with personality making only modest (if any) contributions.

Miller's strategy is to accept Negative Claim and provide an error theory—one consistent with character realism—for why Positive Claim might appear true. People do not have the virtues and vices, but they have other sorts of traits that do explain their behavior, namely the mixed traits—character traits that, morally speaking, are a mixed bag. These clusters of dispositions to produce behaviors are instantiations of the concept of character traits, given that they exhibit Stability Over Time and Cross-situational Consistency. Miller argues that if we look closely at type 1 and 2 studies, we discover that behavior is in fact patterned; the pattern is simply not the pattern given by traditional virtues and vices.

The way to countenance the Destabilizing Data and hold onto character realism, then, is to deny the assumption that links Negative Claim and Positive Claim:

Traditional Traits: If people really possess and act from character traits, then those traits are the traditional virtues and vices.

By rejecting the idea that the pattern we are looking for is the traditional one, Miller is able to discern *other* patterns of behavior. Those other patterns justify the inference to the existence of a different set of mental state dispositions.

For instance, in studies about cheating and honesty, studies do show that behavior exhibits reliable patterns: most people will cheat to avoid personal failure or embarrassment; most people

will not cheat when they fear punishment. We can infer from these that people have dispositions to form beliefs and desires concerned with cheating in order to avoid personal failure, embarrassment, and so forth; not cheating in order to avoid getting caught and punished; achieving various goals or forms of enjoyment (for oneself and others) by cheating; benefits (for oneself and others) of following moral norms against cheating (Miller 2014, 73). These fine-grained dispositions, according to Miller, compose the Mixed Cheating Trait. Psychological studies in types 1 and 2 evidentially support the claim that the Mixed Cheating Trait is widely possessed.[10]

Fleeson and Jayawickreme explain the Destabilizing Data by maintaining that situational variability is compatible with the possession of traditional character traits like courage or honesty but chalk up variability to moving the degree to which someone has the traditional traits. People can display more and less honest behavior at different times, but their behaviors and enacted social-cognitive mechanisms form a recognizable pattern that helps us distinguish between persons, even with much in-person variability. For the density distribution's center for each person indicates a degree of a traditional virtue like honesty as an aggregate of personality states (Fleeson and Jayawickreme 2017).

Moreover, these aggregated distributions have serious predictive power when it comes to the high and low end of the trait-relevant moral behavior; using iterated simulations, Fleeson and Jayawickreme illustrate that people with high levels of moral traits like compassion are extraordinarily unlikely to commit morally heinous acts (ibid.). Even if someone's compassionate

[10] For example, in an experimental setup by Gino and colleagues, people given an opportunity to cheat on a 20-question test without getting caught self-reported scoring answering on average 5 more questions correctly than did people not given the opportunity to cheat (Gino et al. 2009: 397). In another setup, Gino and Galinsky observed that people were much more likely to cheat in a condition where, before the test, they found out they had the same birth month and year as fellow test-taker, and then the fellow test-taker stood up and declared they solved all the problems one minute into the test (Gino and Galinsky 2012: 22, discussed in Miller 2021). Cheating also decreased in a similar experiment under a condition in which participants had to follow the test by writing down "their past and current duties, obligations and responsibilities," (Gino and Margolis 2011:151).

behavior varies between time points, if she has a high degree of compassion it is nearly certain that she will not perform an action that is extremely opposed to compassion, on their model. Fleeson and Jayawickreme thus find reason to reject both the Positive and the Negative Claim derivable from studies of types 1 and 2.

Whole Trait Theorists accommodate the Destabilizing Data with a slight revision to the concept of character trait: they deny that it is part of the concept of character traits that it is highly cross-situationally consistent in its expression. Rather, having a character trait is compatible with a wide range in one's expression of that trait in relevant situations so long as the distribution of one's personality states has a stable *center* over time. Additionally, Whole Trait Theorists deny Traditional Traits: rather, people have varying degrees of those traits—that they can possess them but with the center of their density distribution being at a low end of the trait. This degreed-notion of trait possession allows them to accommodate especially the empirical data from studies of one-off scenarios.

4.3 Strategies for Nonreductivists

These strategies for accommodating the Destabilizing Data certainly allow us to hold onto character realism in light of our evidential situation. But similar strategies can be used without adopting reductivism, and so without running into the dilemma we posed above.

First, nonreductivists can reject Traditional Traits, too. If Traditional Traits is false, evidence that people do not possess traditional virtues and vices is not evidence against the reality of character traits. Perhaps our conceptions of the virtues in the west are wrong, and some other list of virtues better aligns with the kinds of traits we do find. Admittedly, this is not a very attractive strategy because it would require so much additional heavy lifting for the nonreductivist.

Alternatively, it is possible to deny both Negative and Positive Claim by finding flaws in the type 1-3 psychological studies without claiming, like the advocate of the Rarity Thesis, that normative theories involving character realism are not beholden to empirical data. This strategy, too, is compatible with nonreductivism. For example, one might pick out a problem with the studies' assumption that participants are in the "same situations." When psychologists talk of stability, they use the language of "same situation." That phrase admits of two interpretations, however, and the difference between them turns out to matter. One reading is the *nominal situation* reading: if the characteristics of situations observable by third parties are the same, such as the location, time, or generic features of the activity, then the situations are the same. But as psychologist Walter Mischel puts it, "people behave in ways that are consistent with the meanings that particular situations have for them," and possibly "different situations acquire different meanings for the same person," (Mischel 2009, 284). The interpretation that takes into account such subjective meanings is the *psychologically salient situation* reading: if the features that are psychologically salient to the subject are the same, then the situations are the same. The import of the ambiguity in "same situation" is this: *what qualifies as good evidence* that someone does or does not have a trait that issues in stable behavior will depend on which reading we take. On the second reading, for a person to have a character trait requires only that their behavior is stable in the same *psychologically salient* situations. If Krista is in situations of conflict at work on Mondays and Fridays, but views the Monday conflicts as trivial and Friday conflicts as involving important matters of injustice– perhaps they are in meetings about policies she takes to be discriminatory against women and people of color in her workplace– they are not the same situation psychologically.

Insofar as studies code behavior based on nominal rather than psychological construals, they will fail to measure behavior across "the same" situations for the population studied. This is particularly important in studying character traits, since it is widely agreed in psychology that motivation is important for character and virtue assessment. If a study ignores what motivates the actions, goals, or moral identities (Elliot & Sheldon 1999, Darnell et al. 2019), it is likely to miss important distinctions between trait-relevant and non-trait-relevant behaviors (Schnitker et al. 2019; Wright et al. 2020; Fowers et al. 2021).

What about type 2 and 3 studies? Here the nonreductivist might argue that it is actually quite difficult to determine which situations *call for* certain behaviors without a deep understanding of trait-level *normative standards*. As with the Hermit case, without reference to the distal goal of temperance, we cannot evaluate Hermit's behaviors correctly. Yet type 2 (indeed most psychological studies) either measure goals or values but not behaviors, *or* measure behaviors but not goals.

Nonreductivism would suggest a need for a mixed-measure: a measure that captures the behavior and the goal the person has in mind, along with psychologically salient construals of situations in which behavior arises. If hylomorphic nonreductivism is the right way to think about character traits, then whether a behavior manifests a trait depends a great deal on the way the behavior relates to the subject's ultimate goals or identity (McAdams and Pals 2006, Darnell et al. 2019). So, nonreductivist hylomorphic realism can maintain empirical adequacy without running into the dilemma we posed for Miller's nonreductivist monism. The nonreductivist can either reject Traditional Traits or can lodge complaints about study design that show what is being measured is not character on her conception of it (Sreenivasan 2002).

A proponent of hylomorphic nonreductivist realism might have extra resources to explain the data. She will appeal to trait-level goals to explain why certain dispositions belong to a trait as parts, and others don't. So she might be able to explain why an otherwise honest person would misreport test results upon finding out someone her exact age scored perfectly (perhaps she doubts herself or thinks it more likely she made a mistake in scoring herself than that her peer's score so widely diverged from hers). The hylomorphic nonreductivist can accept that certain dispositions may be activated in some situations but not others, as when Krista engages in conflicts on Fridays but not Mondays, due to the grounding trait of justice or peace.

The reductivist monist cannot escape counterexamples easily, since she thinks the trait just is those dispositions she committed to in her reductive analysis. By contrast, the hylomorphist takes a functionalist approach: in different circumstances, different desires and beliefs and subsequent behaviors function to serve the goal of the character trait, like temperance. If Hermit has temperance, then dispositions which, when activated, serve the goal of being of sound body and of controlling the appetites in a given moment— whatever these dispositions might be— are compatible with the trait and Hermit will not be a counterexample to the analysis of temperance in terms of that ultimate moral goal. The hylomorphist nonreductivist is able to read the Hermit in a way Cassian intends it to be read: as an opportunity to update her understanding of temperance.

4.4 New Views on Character and Nonreductivism

To shore up the case for nonreductivism's empirical adequacy, consider four views of character currently on offer in contemporary psychology. All of these are compatible with nonreductivism and are empirically respectable.[11]

---

[11] Although not really an account of character, the "New Big Five" framework proposed by McAdams and Pals (2006) is also compatible with nonreductivism. An ambitious proposal, the New Big Five seeks to integrate varying

First, Fowers, Carroll, Leonhardt, and Cokelet defend the STRIVE-4 model of virtues (Fowers et al. 2021). Here, positive moral character traits, or virtues, have five features: they are "scalar traits that are role sensitive, interact with situational influences, help the individual to attain valued ends, and promote a flourishing life" (2021, 119). The proponents of STRIVE-4 do not discuss the metaphysics of virtues, but we can gain insight into how their model might fit within our taxonomy based on what they say about the explanation for within-person variability and the relation between virtue and neurophysiology.

The STRIVE-4 Model shares with WTT the idea that a trait's expression over time is expected to vary within persons: "virtues can be seen as individual differences that also vary over time for a specific person," (ibid., 127). Defenders of STRIVE-4 and WTT agree that countenancing within-person variability is crucial to empirical adequacy (ibid., 132). But while WTT explains variability by reference to disparate components of a trait – personality states and the mechanisms at each time point that compose them—proponents of STRIVE-4 offer a different explanation. They hold that the level of expression of a trait will probably fluctuate between different situations and over time within a person, but propose that these fluctuations are in part due to variations in the relevance of a trait to different situations: "no trait is appropriate in every situation" (ibid., 130). Another source of variation within persons on the STRIVE-4 model is different social roles a person might inhabit and move between over time, like the personal role of being a parent, the professional role of being a teacher, the civic role of being a neighbor or council-

- and seemingly conflicting - approaches to personality into one comprehensive framework that would allow psychologists to understand a person as a whole. On this account, "personality is an individual's unique variation on the general evolutionary design for human nature, expressed as a developing pattern of dispositional traits, characteristic adaptations, and integrative life stories complexly and differentially situated in culture" (ibid., 212). McAdams and Pals do not discuss their commitments when it comes to metaphysics of character, but this is probably because character traits form only one aspect of their integrative conceptual framework. However, as an account that is designed to accommodate distinct individual frameworks of personality, it has room for nonreductivist character realism.

person. Someone might manifest less generosity in giving time to charity, for instance, while inhabiting the parent role because she takes herself to owe that time to her small children more than to a stranger, but while at work she might spend much of her time in a service capacity by voluntarily mentoring or tutoring struggling students.

This sounds a lot like what a hylomorphist nonreductivist would say; the trait does the governing, determining when it is going to be expressed and in what circumstances or roles it will be triggered. This provides a way of getting around the seeming lack of consistency in behavior across situations without conceding that fine-grained dispositions attuned to those situations are grounding the trait. Moreover, the authors indicate that the morally good trait, or virtue, patterns fine-grained dispositions, rather than the other way around: "virtue traits describe the ways these components are patterned" (2020, 129).

Second, Schnitker, King, and Houltberg write:

> "In relation to Aristotelian conceptions of virtue development, prosocial behavior alone does not constitute virtue; instead, moral intentionality is required (McAdams, 2014). Thus, characteristic adaptations are better suited for structuring virtue as they involve agency, motivation, and at least some choice. Whereas other dispositional units of personality might lead people to adopt particular recurring patterns of prosocial behaviors (i.e., socio-emotional performance trends captured in conscientiousness or agreeableness), virtues require intentionality and motivation that are dependent upon the cognitive and emotional infrastructure provided by characteristic adaptations" (2019, 279).

While Schnitker and colleagues are also not committal about their metaphysics, note the emphasis placed on intention and motivation "as provided by" character on their view. This makes it sound like the character trait supplies the more specific intentions and motives that drive behaviors. If so,

the trait is grounding the finer-grained dispositions to form specific beliefs, desires, and emotions in particular contexts, making the model nonreductivist.

Darnell and colleagues (2019) put forward a view that is not only nonreductivist about traits, but which also suggests that the virtues are themselves unified and given shape from the top down by practical wisdom, or phronesis. Phronesis bridges the knowledge-action gap, on their account, by integrating with one's moral identity various cognitive, perceptual, and regulatory capacities such that one perceives situations aright, feels fitting emotions in response to situationally specific features, and responds appropriately. The authors admit that it is empirically unrealistic to think that a person who possesses practical wisdom has all of the virtues to the same degree, and question that some virtues "hunt in packs" so to speak. But they maintain that practical wisdom may be both degreed and its ability to govern particular domains may vary with the extent to which the person with practical wisdom has relevant experience in that domain (ibid. 115, Turiel and Nucci 2018).

Finally, Cole Wright and colleagues (2020) also argue that virtues must be guided by practical wisdom or phronesis. Traditionally, practical wisdom serves an architectonic role, coordinating which virtues are activated in which situations. Practical wisdom on their view seems to play just this role. And moral virtues coordinated by practical wisdom "activate" particular patterns of thought and feeling, and, subsequently, behavior. It does not seem to us that dispositions to form particular thoughts and feelings would have to form the basis of the trait or cause the existence of the trait on this view. Instead, more in keeping with the Aristotelianism of their model, it would make sense for the virtues to give rise to dispositions form particular thoughts and feelings, and for that reason it would be perspicuous to say that virtues "activate" those thoughts and feelings.

6 Conclusion

In this essay we have argued that realists about character traits ought not to be reductivists because of the formidable theoretical dilemma realist reductivism faces. We also showed that there are alternative realist views, notably hylomorphic nonreductivism, that can accommodate the empirical data about character as well as going reductivist realist views. We will close by offering up a few very general practical suggestions for the impact nonreductivism might have on the way we identify and measure virtues.

First, suppose we take seriously the idea that having a character trait involves having certain normative goals that govern behavior and ground more fine-grained mental-state dispositions. To test for the existence of a character trait would require a mixed measure with a complex observational component to measure a subject's behavior and a self-report component measuring goals to capture motivation of the subject. The nonreductivist view predicts that some people will exhibit behaviors that look like generous, or temperate, or honest behaviors but lack the trait-level motivational goals. Such people will not qualify as generous, temperate, or honest. It also predicts that certain behaviors like giving money to someone in need, which seem to belong to a character trait like generosity, may ot be called for in certain situations given that, in those situations, it would not accomplish the normative goal of generosity (as in Sergio's case). In their recent article on virtue measurement, Ng and Tay argue that tools for virtue measurement need to take into account the variability of what is required to exhibit virtue in different contexts and for different social roles (Ng & Tay 2021). The hylomorphic model would support this argument.

Further, if the nonreductivist hylomorphic model of character is preferable to the reductivist one, then testing for a cluster of fine-grained mental-state dispositions with no normative content will be unlikely to yield informative results. To detect character traits

accurately, the psychological measures and coding will need to be infused with normative content based on an appreciation of the goals of the trait. Modifying items on an observational measure to accommodate this normative goal will require an appreciation of the trait's overarching goal. It will also require an ability to identify when the goal is or is not being served by a behavior, or could reasonably be expected to be served by a behavior. Finally, this will need to be triangulated against the subject's motives. Measuring motives after the fact, of course, raises issues with post-hoc rationalization and confabulation; but there are ways to mitigate these potential issues.

We can hope that psychological studies that employ mixed measures and openly normative content will allow us to more seriously test for the existence of character traits as conceptualized by nonreductivist theories. Only once we have studies using such measures in place can we return to the question of whether traditional virtues and vices, or character traits like continence, or mixed traits like Miller posits are widespread.

References

Alfano, M. (2013). Character as Moral Fiction. Cambridge: Cambridge University Press.

Aquino, K., Freeman, D., Reed, A. I., Lim, V. K. G., & Felps, W. (2009). Testing a social-cognitive model of moral behavior: The interactive influence of situations and moral identity centrality. Journal of Personality and Social Psychology, 97, 123–141.

Arpaly. (2002). Moral Worth. The Journal of Philosophy 99(5): 223-245.

Beary, A. (2019). A Thomistic principle of virtue individuation (Publication No. 10977626) [Doctoral dissertation, Baylor University]. ProQuest Dissertations & Thesis Global.

Cole Wright, J., Warren, M. T., & Snow, N. (2020). Understanding virtue: Theory and Measurement. New York: Oxford University Press.

Darnell, C., Gulliford, L., Kristjánsson, K., Paris, P. (2019). Phronesis and the Knowledge-Action Gap in Moral Psychology and Moral Education: A New Synthesis? Human Development 62(3): 101-129.

Doris, J. (2002). Lack of Character. Oxford: Oxford University Press.

Elliot, A. J., Sheldon, K. M., & Church, M. A. (1997). Avoidance personal goals and subjective well-being. Personality and Social Psychology Bulletin, 23(9), 915-927

Farrell, A. D., Bettencourt, A. F., & Mehari, K. R. (2019). Beliefs About Fighting and Their Relations to Urban Adolescents' Frequency of Aggression and Victimization: Evaluation of the Beliefs About Fighting Scale. The Journal of Early Adolescence, 39(6), 785–813. https://doi.org/10.1177/0272431618791297

Farrell, A.D., Erwin, E.H., Bettencourt, A., Mays, S., Vulin-Reynolds, M., Sullivan, T., Allison, K.W., Kliewer, W., & Meyer, A. (2008). Individual Factors Influencing Effective Nonviolent Behavior and Fighting in Peer Situations: A Qualitative Study with Urban African American Adolescents, Journal of Clinical Child & Adolescent Psychology, 37:2, 397-411, https://doi.org/10.1080/15374410801955821

Flanagan, O. (1991). The Varieties of Moral Personality: Ethics and Psychological Realism. Cambridge: Harvard University Press.

Fleeson, W., & Jayawickreme, E. (2015). Whole trait theory. Journal of Research in Personality, 56, 82-92.

Fowers, B. J., Carroll, J. S., Leonhardt, N. D., & Cokelet, B. (2021). The emerging science of virtue. Perspectives on Psychological Science, 16(1), 118-147.

Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel. Psychological Science, *20*(3), 393–398. https://doi.org/10.1111/j.1467-9280.2009.02306.x

Gino, F., & Galinsky, A. D. (2012). Vicarious dishonesty: When psychological closeness creates distance from one's moral compass. Organizational Behavior and Human Decision Processes, 119 (1), 15–26. https://doi.org/10.1016/j.obhdp.2012.03.011

Gino, F., & Margolis, J. D. (2011). Bringing ethics into focus: How regulatory focus and risk preferences influence (un)ethical behavior. Organizational Behavior and Human Decision Processes, 115 (2), 145–156.

Hartshorne, H., & May, M. A. (1928). *Studies in the nature of character, [part] 1: Studies in deceit: book 1, General methods and results; book 2, Statistical methods and results* (pp. 134–158). New York: MacMillan.

Howard, N.R. (2021). The Goals of Moral Worth. In ed. R. Shafer-Landau, Oxford Studies in Metaethics: Volume 16. Oxford: Oxford University Press.

Jayawickreme, E., & Fleeson, W. (2017). Does whole trait theory work for the virtues. *Moral psychology*, *5*, 75-103.

Kamtekar, R. (2004). Situationism and Virtue Ethics on the Content of Our Character. Ethics 114: 458-491.

Kouchaki, M., Gino, F., & Jami, A. (2014). The burden of guilt: Heavy backpacks, light snacks, and enhanced morality. Journal of Experimental Psychology 143(1), 414.

Markovits, J. (2010). Acting for the Right Reasons. Philosophical Review 119(2): 201-243.

McAdams, D., and Pals, J. (2006). A New Big Five. Fundamental Principles for and Integrative Science of Personality. American Psychologist 61(3): 204-217.

Miller, Christian. (2013). Moral Character: An Empirical Theory. Oxford: Oxford University Press.
--- (2014). Character and Moral Psychology. Oxford: Oxford University Press.
----(2018). Generosity: A Preliminary Account of a Surprisingly Neglected Virtue. Metaphilosophy 49(3): 216-245.
——(2020). The Motivation and the Virtue of Honesty: Some Conceptual Requirements and Empirical Results. Ethical Theory and Moral Practice 23: 355-372.
Mischel, W.(1968). Personality and Assessment. New York: Wiley.
----(2009). From Personality and Assessment (1968) to Personality Science, 2009. Journal of Research in Personality 43:282-290.
Ng, V. & Tay, L. (2020). Lost in translation: The construct representation of character virtues. Perspectives on Psychological Science, 15 (2), 309-326
Papish, L. (2017). CAPS Psychology and the Empirical Adequacy of Aristotelian Virtue Ethics. Ethical Theory and Moral Practice 20 (3): 537-54.
Russell, D. (2009). Practical Intelligence and the Virtues. New York: Oxford University Press.
Schnitker, S. A., King, P. E., & Houltberg, B. (2019). Religion, spirituality, and thriving: Transcendent narrative, virtue, and telos. Journal of Research on Adolescence, 29(2), 276-290.
Sreenivasan, G. (2002). Errors about Errors: Virtue Theory and Trait Attribution. Mind 111(441): 47-68. https://doi.org/10.1093/mind/111.441.47
Stohr, K. (2018). Virtuous motivation. The Oxford Handbook of Virtue, 453-469.
Snow, N. (2009). Virtue as social intelligence: An Empirically Grounded Theory. New York: Routledge.
Stichter, M. (2018). The Skillfulness of Virtue. Cambridge: Cambridge University Press.
Turiel, E., & Nucci, L. (2018). Moral development in contexts. In A. Dick & U. Mueller (Eds.), Advancing developmental science: Philosophy, theory and method (pp. 95–109). London, UK: Routledge.
Vernon, P.E. (1964). Personality Assessment: A Critical Survey. New York: Wiley.
West, R. (2016). Anger and the virtues: A critical study in virtue individuation. *Canadian Journal of Philosophy, 46*(6), 877-897. doi:10.1080/00455091.2016.1199232