Extended Virtues and the Boundaries of Persons

ABSTRACT: *What sorts of things undergird a person's character and what does the answer tell us about the person's relationship to her body, her environment, and the people who surround her? For the purpose of ascribing virtue, vice, and character, what are the boundaries of the person? Traditionally, philosophers have accepted mentalism, the view that only mental features are relevant to character. Even philosophers inclined to say that character is partially constituted by nonmental features would be inclined to accept skindividualism: a person's virtue must involve a disposition that is wholly grounded by features inside the person's skin. Drawing from arguments in the philosophy of mind and data from social psychology, I argue that skindividualism is wrong. Virtues can be extended, in the sense that the grounds for a person's virtues might not be inside that individual's skin. If this is the case, persons are not skindividuals.*

KEYWORDS: ethics, moral psychology, philosophy of mind, philosophy of psychology

Philosophical discussions of character and virtue tend to focus on what it takes for a person to possess them. The traditional account is roughly Aristotelian: for persons to have a virtue is for them to have a stable state that grounds a reliable disposition to feel and act appropriately (Annas 2011). To say the state is stable is to say that the agent has it across a range of times and situations, and the state grounds the dispositions it does even in the face of adverse conditions. For example, if Cate has a disposition to tell the truth but only when it would be pleasing to the hearer, she's not really honest. To call a disposition reliable is to say that it can be counted on. If Cate has a disposition to tell the truth only three quarters of the time, her disposition might well be stable, but it doesn't issue in truth telling reliably enough to count as the virtue of honesty. A virtuous state grounds a disposition to reason, feel, and behave in appropriate ways and do so for the right reasons. The action part is obvious: if someone is a generous person, he or she must be disposed to give to others when appropriate, and presumably because it is appropriate. According to the Aristotelian, the generous person also feels appropriately. If Cate feels selfish pangs every time she gives to others, she isn't really generous but is in some sense merely going through the motions.

Though it is clearly important to analyze what it takes for a person to have a virtue, focusing on that question tends to eclipse even more important questions about what persons must be if they are to have virtues, vices, and characters. What sorts of things can legitimately undergird a person's character? And what does the answer tell us about the person's relationship to her body, her environment, and the people who surround her? For the purpose of ascribing virtue, vice, and character—and thus certain forms of praise and blame—what are the boundaries of the person?

One traditional answer involves a narrowly circumscribed notion of a person. On this narrow view a person's character traits are wholly determined by her psychological states. One motivation for this view might be the thought that anything outside my mind is just inherited circumstance. I didn't choose it, so I can't be held responsible for it. Of course, this isn't a particularly persuasive argument. Most of my psychological traits are a result of inheritance or circumstance, and therefore the fact that a trait isn't explicitly chosen had better not be a reason to exclude it as the ground for a virtue. A better reason for the above view is the thought that purely bodily states are simply too dumb to harbor virtues. It doesn't make sense for something like honesty to reside in my liver, even in part. Bodily states don't inform agency in the way virtues do. But there is mounting evidence that the body isn't as dumb as all that. Bodily states are implicated in reasoning in surprising ways. According to Damasio's somatic marker hypothesis, for example, parts of the body intertwined with the feeling and expression of emotion can subtly guide and bias reasoning. Individuals with damaged or altered capacities to receive and encode bodily signals associated with emotions can be more likely to engage in risky behavior or less likely to respond to trends in their environment (Damasio 1994). But we needn't go to the lab to see that this is plausible. Whether someone is cowardly is likely to depend a great deal on whether he or she is physically strong. Whether someone is apt to be moderate in the use of certain drugs is likely to depend on his or her physical tolerance to those drugs. Of course, it is a matter of luck, good or bad, whether one has these bodily dispositions, but this doesn't mean the people who have them are not thereby courageous or moderate.

To accept that bodily states can undergird virtues is not to say that they can do so alone, without the subject's having any particular psychological states at all. Facts about one's build are unlikely to determine facts about one's virtuousness or viciousness without a certain set of beliefs, desires, and psychological traits. Nevertheless, the bodily states help ground virtuous dispositions.

Even philosophers who accept this much, however, are likely to think that the bounds of the person—insofar as the person is relevant to virtue—don't extend beyond the body. That is, most philosophers assume *skindividualism*: the person does not extend beyond her or his skin. In particular, when it comes to virtue, skindividualism would maintain that a person's virtue must involve a disposition that is wholly grounded by features inside the person's skin.

I argue that skindividualism is wrong. Virtues, vices, and other character traits can be extended, in the sense that the grounds for these traits might not be within that individual's skin. If this is the case, persons are likewise extended in that

a person can have constituents that are not within the skin. Persons are not skindividuals.

This point is not merely verbal. When we are discussing the nature of virtues, we aren't simply stipulating what something must be if we are to use the word 'virtue' to describe it. Virtue and vice are normative notions, and what counts as a virtue or a vice is guided by normative considerations. It is a condition for something being a virtue that it is the type of thing a person can be praised for having. Our discussion should therefore be led by a consideration of what sorts of things do and do not make someone praiseworthy in this way.

As in the case of virtue, it is important to ask what is guiding our analysis when we ask about the boundaries of the person. Is it just a matter of arbitrary conceptual choice whether we draw the boundaries of the person around the mind, the body, or the skindividual? No, because like virtue, personhood—at least as I am using it—is a normative concept. What we include as part of the person is determined by what persons must be if they are to bear the normative weight we give them and in particular if people are to be praised or blamed for the virtues and vices we ascribe to them.

Since the argument of this paper touches on a number of active debates, it might help to outline the argument at the beginning so as to be clear about the argumentative burden. I will start by considering the argument for extended minds, not because I wish to endorse that thesis but because it suggests an analogous argument for extended persons. If minds are extended, of course, virtues are likely to be as well, and that is interesting in its own right. But since the individuation criteria for persons are apt to be quite different from those for minds, the debates are different. The analogous argument for extended persons will open up a metaphysical possibility that skindividualism is false. I argue that the results of experiments in situationist psychology suggest that if we are to have virtues, they must in fact be extended; situationism speaks against skindividualism. The claim is not that extended virtues will answer all the situationist worries. The claim is rather that some of the prominent results of situationist psychology support a picture on which our virtues are undergirded by systems beyond our skins. Combining the extended virtues argument with the results of situationist psychology, then, gives us a new and surprising picture of character and virtue that deserves to be a serious contender in moral psychology.

1. From Extended Minds to Extended Virtues

In their seminal article 'The Extended Mind', Andy Clark and David Chalmers argue that our minds extend beyond the boundaries of our skin (Clark and Chalmers 1998). Not only do cognitive processes involve external props, such as computer screens and pencils, mental states themselves can be realized in external artifacts such as iPhones and notebooks. The argument essentially turns on the premise that what makes a mental state what it is (a memory versus a desire versus no mental state at all) is not the stuff of which it is made. What makes something a mental state is the functional role it plays. This alone is not surprising, nor is

it new. The novel part is that Clark and Chalmers realize 'there is nothing sacred about skull and skin. What makes something a belief is the role it plays, and there is no reason why the relevant role can be played only from inside the body' (Clark and Chalmers 1998: 14).

Consider an example. Jay and I both know Matt's phone number. At one point we asked him for it, and he told us. If you ask either of us, we can tell you because we can both recall it on a moment's notice. There is a difference, though, in how we store it. Jay stores the information in his brain. He doesn't know how he does it, but after hearing the number, he repeats it to himself a few times, and it gets stored somehow among the clouds of neurons that are his brain. I don't like doing that. Instead, I store Matt's phone number in my phone. When I hear the number, I enter it, press 'save', and it gets stored somehow in that little device. Clark and Chalmers's point is that as long as the stored information plays the same functional role, it doesn't matter where it's stored or how it's encoded.

There are differences, of course, and they aren't limited to the composition and location of the phone. I can leave my phone in a taxi, but Jay's brain state is unlikely to be left behind. If we become marooned on a desert island, my memories will disappear when my phone runs out of charge while Jay's memories will remain intact. These are important differences, and they may provide a reason to say that these are really different sorts of states: one is a memory state, and perhaps the other is a shmemory state. But this doesn't settle the question of the extended mind unless there is some argument for the claim that only memories (and not shmemories) are mental states. This doesn't seem motivated. There are plenty of cases in which brain damage prevents minds from working in a typical fashion so that mental states lose some of their characteristic function. In anterograde amnesia memories don't become long term but become irretrievable after a short span of time. They are still memories. As far as I know, there are people who lose their memories whenever they bump their heads while exiting a taxi, losing their memories just as I lose my iPhone. Do they not have memories? I imagine we would say that they do have them, only oddly fragile ones. We are pretty flexible in our willingness to apply mental concepts to degraded or nonstandard cases because in such cases the essential function is performed well enough. It seems arbitrary, then, to exclude extended minds because they differ from natural minds in some less essential functional respect.

If states like memories can be extended, then it would seem that in principle states like virtues can be extended as well. Virtues, remember, are simply stable states of the person that ground reliable dispositions to act, think, and feel appropriately. If mental states in general need not be located within one's skin, then the states that ground the dispositions relevant to virtue need not be so located either. The fact that the ground for the disposition is not located inside my skin doesn't preclude that virtue from being ascribable to me. What is important is that the state grounds the relevant disposition, not that the state is located in my skin.

The entailment from the possibility of extended minds (EM) to the possibility of extended virtues (EV) is not in fact that straightforward. Even if some psychological states can be extended and some virtues are states that ground psychological dispositions, it doesn't follow that the types of state that can be extended are the

same as the types of state that can ground the psychological dispositions relevant to virtue. Even ardent defenders of the extended mind will allow that there are some psychological states that might not be able to be extended. Clark and Chalmers, for example, think that it is not particularly likely that conscious states can be extended (Clark and Chalmers 1998: 11). Perhaps virtues cannot be extended for the same or similar reasons.

One reason a certain type of state would be unlikely to be extendable is if it were not functionalizable—that is, if what it takes to be in such a state is not simply a matter of playing a role. Conscious states, for example, do not appear to be functionalizable in this way; they appear to have phenomenal essences or ways they feel that go beyond any role they might play. This is, of course, part of what gives rise to the hard problem of consciousness (see Chalmers 1996, for example), and I expect it is part of the reason Clark and Chalmers are reluctant to endorse the idea of extended conscious states.

There doesn't seem to be this concern about virtues. Virtues themselves need not be conscious states, for example. They might ground a disposition for an individual to enter a certain conscious state, but that doesn't make them conscious states, and it doesn't prevent them from being functionalizable. It's hard to see any other reason why they wouldn't be functionalizable, especially since virtues are themselves spelled out in terms of dispositions, and dispositions are paradigmatic cases of functionally defined states. If that's right, and functional states can be extended (at least in principle), then states of virtue can (at least in principle) be extended.

Of course, it doesn't follow from the fact that a type of state can in principle be extended that such states actually are extended. There might be certain functional roles that have to be played by organic body or brain states simply because nothing else has the appropriate causal powers. Again, though, it's hard to see why virtues would require anything organic or internal.

Consider, for example, the 'virtue' of punctuality. If I am punctual, this is likely to owe quite a bit to the fact that I have a watch. The watch is located outside of me, but if I am punctual it is in part because of the thing on my wrist. Of course, there are things located inside me that are crucial to my punctuality, such as the desire to be on time. But the virtue of punctuality requires more than that. It requires that I have the disposition to realize that desire. This disposition is partly located in my watch, which is outside of the skin but always with me.

2. Extended Virtues without Extended Minds

In the previous section I've suggested that the extended mind hypothesis has some plausibility and that if minds can be extended, virtues can be, too. In other words, I've argued that if EM is the case, EV is likely to be the case. This is important because it provides a certain normative significance to EM that helps explain why the thesis is important. Those who are inclined to think, for example, that the EM debate is basically a verbal dispute might be more reticent if the truth of EM entails that people have virtues they wouldn't otherwise have.

However, EM is far from universally accepted, and it faces substantial criticism on a number of fronts. For example, Adams and Aizawa (2010) argue that nonbrain devices such as iPhones lack 'intrinsic content' and have only derived content. According to them, this is a stark difference that prevents things such as iMemories from counting as memories. In other work, such as Adams and Aizawa (2008), they argue that the range of possible extensions is too motley to be the object of any single science, which suggests brains and their extensions cannot be considered part of the same system. Rupert (2009) argues that extended minds fail to be sufficiently durable, and he agrees that once one abandons the way that cognitive science individuates cognitive systems, one faces an unsolvable problem of demarcating the bounds of cognition. Weiskopf (2008) argues that at least at present extended devices are not sufficiently inferentially integrated to count as part of the mind.

Although I suspect many of these criticisms can be met, I can't hope to make good on that here (for the general strategy I might take in responding, see Fisher 2008). In any case, it would be better if the case for EV didn't depend on EM. Fortunately, it doesn't. The general reason is that minds are not persons, and persons, not minds, are the bearers of virtue. Arguments against extended minds, then, aren't necessarily arguments against extended persons and extended virtues. This is not to deny that virtues and character traits are 'anchored in the specific dispositions to form relevant beliefs and desires' (see Miller 2013: 10–11) or in the disposition to have some conscious feelings under certain circumstances. What I'm arguing is that the grounds of these dispositions need not themselves be mental states.

The dispositions that anchor character traits could be constituted by states of the person that—for many of the reasons specified by Rupert, Adams and Aizawa, and Weiskopf—are not themselves mental states. They might, for example, be bodily states. Jeff, for example, might be particularly unaggressive, sexually moderate, and inclined toward depression because of abnormally low levels of testosterone. These low levels can be due to physiological problems ranging from testicular cancer to the mumps. Jeff has the traits of being unaggressive, depressive, and sexually temperate, and these are grounded in features of his physiology.

Now consider Cal, who also has the traits associated with low testosterone. But instead of that being caused by some internal issue, Cal is a new father of twins. Recent studies have confirmed that fatherhood decreases testosterone and that, in fact, the more time fathers spend caring for their young, the more their testosterone levels decreases (Gettler et. al. 2011). We can imagine that Cal is a stay-at-home father and that while his friends are not surprised at his dedication to his kids, they are surprised at the fact that he has become less aggressive and has had much less of a wandering eye even when he is out with the guys. Fatherhood has changed him, no less than the cancer has changed Jeff. The grounds for Cal's dispositions lie in something social in this case—his role as a father.

Taken together, the cases of Jeff and Cal show that the argument for extended virtues and character traits does not depend on the argument for the extended mind. Jeff's case shows that the ground for the relevant disposition needn't be mental at all, and in Cal's case the same role is played by something that isn't wholly inside his skin. In neither case are the grounds mental, and therefore reasons why mental

states can't be extended would not be reasons that these grounds can't be outside the skin.

It's important to realize the general reason the cases for EM and EV are distinct. People, not minds, have virtues, and criteria for being part of the same mind are not criteria for being part of the same person. Perhaps it is plausible, for example, that all mental states must somehow employ intrinsic representations, but it is not plausible that all parts of a person must be representational. Similarly, not all parts of a person need be inferentially integrated for the simple reason that not all parts of people are the sorts of things involved in inferences. Persons include more than minds. They include bodies, but if I am right we should recognize that they include things and even other people outside the body as well. What makes something a part of a person will not necessarily be a matter of informational integration or anything of the sort. More relevant is whether that something bears appropriately on our evaluation of the person. That is, the question we should ask is whether or not skindividualism is adequate to bear the load of our treatment of persons and our conception of their traits and virtues. There is good reason to believe that it isn't and that if we want to retain our commitment to the existence of moral character, we should reject skindividualism.

3. The Situationist Critique

The traditional view of character and virtue has received significant criticism from social psychology. Attributions of virtue and character should have predictive value (Harman 1999; Doris 2002). Part of what it means to ascribe a disposition to something is to say that the thing will behave in a certain manner in relevant contexts, and virtues are supposed to be particularly stable and characteristic dispositions of a person. According to the situationist critique of virtue, if it turns out that people in general do not demonstrate the constancy of behavior predicted by attributions of virtues and character traits, we have to conclude that there are no such things. Moral character will turn out to be a fiction; it is just a part of the story we tell about ourselves that doesn't really withstand empirical scrutiny.

According to many, this is what we find when we look at how people behave. According to situationists, people do act somewhat consistently, but this is because of the consistency of the situations in which they are acting, not due to the robustness of their characters. If one is looking for a predictor of behavior, one does much better to look at the context of action than at the dispositions of the agent. If this is correct, our ascriptions of virtue and character are in trouble. Virtues are supposed to be robust in that they manifest behaviors across contexts, not only when the situation is favorable. If it turns out that we fail to have robust traits like this, we seem little more than leaves in the wind.

Probably the most well known—and most troubling—experiments cited by the situationist are Milgram's experiments in the early 1960s. In those experiments, the subjects were led to believe that they were part of a test about memory. Some subjects would take a test while wired with electrodes, while others would sit in a different room with the experimenter, administering electric shocks to the

'learners' when they answered questions incorrectly. In fact, the 'learners' and the 'experimenter' were confederates (playing a role for the real experimenter), and the purpose of the experiment was to see how severe a shock subjects would be willing to give. As the wrong answers came, the subjects increased the shocks, and the 'learner' expressed increasing pain. The horrifying result of the experiment was that throughout the many trials, approximately 65 percent of participants were willing to continue administering shocks after a very dangerous 345 volts, counting the learner's silence (after his pleas that he had heart trouble and wanted to quit) as a wrong answer. Though 35 percent were disobedient, almost all subjects were willing to administer shocks past 150 volts, a level at which the learner demanded release (description from Doris 2002: 39–51).

Milgram's results cast a poor light upon the traditional picture of character. According to Doris, 'personality research has failed to find a convincing explanation of the Milgram results that references individual differences. Accordingly, Milgram gives us reason to doubt the robustness of dispositions implicated in compassion-relevant moral behavior; his experiments are powerful evidence for situationalism' (Doris 2002: 39).

A second famous case is provided by the Stanford prison experiment conducted by Zimbardo and colleagues in the early 1970s. The experimenters chose 21 subjects to replicate the conditions of a prison in the basement of the psychology building. The participants were randomly divided into 'prisoners', who would be kept in cells and 'guards' who were to maintain order and distribute nonviolent punishment, such as solitary confinement and menial labor. The prison ran 24 hours a day until the experiment was shut down. The result? Those selected to be 'guards' began to act like guards, reveling in their power and using it to punish the 'prisoners' in severe ways. Those selected to be 'prisoners', meanwhile, acted like prisoners. They experienced severe stress, to the point of near breakdown. Some devised an uprising (quelled by the 'guards' with fire extinguishers), and some turned against the other prisoners to curry favor with the guards (Haney, Banks, and Zimbardo 1973).

We needn't even turn to experiments, however, to find the most searing cases of monstrosity. In Nazi Germany, a disproportionate amount of people were willing to commit atrocities—sometimes at the behest of their superiors, but at other times voluntarily. In Rwanda in the 1990s, the Hutus massacred the Tutsis in indescribable ways. Though both the Hutus and the Nazis acted monstrously, we have to ask which is more plausible—that in these places at these times people of especially low character wielded the Mauser and the machete or that the situation led ordinary people to do terrible things? The situationists say the latter, and they seem to have a strong case (Doris 2002).

Although there are a number of more conservative responses to this debate, some philosophers have arrived at rather startling conclusions. Harman (2000) is probably the most extreme, essentially adopting eliminativism about character traits. In his later work, Harman (2009) is much more cautious, but he remains convinced that at best character traits describe 'actual world regularities' rather than deep dispositions that have counterfactual implications (Harman 2009: 241).

Doris (2002) is slightly more sanguine than Harman (1999, 2000). Doris is willing to admit that 'local traits' exist in relatively circumscribed conditions:

> The catch is that the 'trait-relevant eliciting conditions' for local traits are specified quite narrowly. This means that local traits are not robust; they are not reliably expressed across diverse situations with highly variable degrees of trait conduciveness. (Doris 2002:65–66)

In other words, Doris admits that there are traits and dispositions that allow some prediction of behavior but fall far short of traditional virtues (Upton [2005 and 2009] develops a 'contextual' account of virtues and character traits that is similar to the local-traits view).

4. Skindividualism and Situationism

Doris and Harman have responded to the situationist data by eliminating or narrowly circumscribing traits and virtues. But these conclusions are warranted only on the assumption of skindividualism. It is assumed that whatever dispositions one has outside the experiment one also has within the experiment, and only on this assumption can behavior within the experiment indicate that the disposition was not robust. However, if the extended virtue thesis is correct, virtues do not supervene on what is inside a person's skin. This opens up the possibility that in some situations the subjects' dispositions are not present because the subjects are missing an important part of themselves. It is not that the subjects have dispositions with narrowly circumscribed eliciting conditions, as the local virtue approach has it. It is rather that the subjects have dispositions with broad eliciting conditions that are grounded in states that depend on more than the skindividual. We can still say of the person as a whole, extensions included, that this person has robust dispositions even if dispositions are not manifested when their extensions are sheared away.

Compare a case of an individual who has been on antidepressants for years. He takes them regularly, and they are effective. One of their effects is that he is much more compassionate. He is more likely to help others, less likely to be brusque or insulting, and likely to be far less aloof. He is compassionate. If he goes on a trip and forgets to bring his medications, he is unlikely to be an enjoyable travel companion, however. Is that a reason to say he was never a compassionate person? No. All of our psychological dispositions are the result of delicate chemical balances that could be upset artificially. If believing in virtues means we must believe in traits that exist no matter what happens to the chemistry of the mind or the constitution of the person, the case was hopeless from the start. In general, we don't hesitate to praise or blame people for traits that are partly a result of circumstance or that can be disrupted due to a change in circumstances.

The extended virtue theorist will insist that the individual with extended virtues is like the individual on antidepressants. He or she has virtues when the extensions are in place even though he or she will not necessarily act virtuously when the

extensions are gone. In the experimental situations that fail to elicit the expected virtuous behavior, either important parts of the extended person are not present or they are disabled. Most crucially, our virtues are in part constituted by our social environment. Gallagher (2013a) expresses a similar view, but his argument and focus emphasize the extension of cognition and presuppose an enactivist framework (see also Gallagher 2013b and 2013c). In my argument above, the social environment can be construed narrowly, as when the presence of one's spouse keeps one's behavior in check, or broadly, when social mores help regulate behavior. Take these things away or construct environments where they are not present, and the person's virtues will be disabled. This is akin to taking someone off antidepressants—the resulting behavior does not reflect the weakness of previously existing dispositions; rather, it reflects the fact that the ground of the moral dispositions is missing due to a change in circumstance.

Consider another, slightly more fanciful, analogy. Suppose there were a room pervaded by a peculiar magnetic field that had the effect of decreasing activity in the limbic system. Among other effects, people who enter the 'limbic room' have an extremely dim emotional life so that they no longer experience empathetic responses or feel moral outrage when contemplating causing harm. Under such circumstances, it would hardly be surprising if the subjects in the room were willing to cause a great deal of pain to others for trivial reasons. Their behavior in the room would likely be terrible. But would this indicate that they lacked virtue or that their virtue depended on the neutralized brain region? Clearly the latter. The peculiar nature of the room would be considered an excusing condition, lessening the moral responsibility of the agents since in some sense they are not themselves.

Compare the Milgram experiment. In these cases, people are moved into relative isolation where the rules of the game are unfamiliar. No one is present to help them navigate the situation; there is only an 'experimenter' who is not only failing to ground extended virtues, but is arguably neutralizing them with the insistence that this is an experiment that must continue. Instead of being in the 'limbic room', these subjects are in the 'science room' in which typical social supports are absent and many interpersonal rules are suspended. Evidence suggests that the individuals would not act that way if these neutralizing forces weren't present and if the typical social factors were present. Kilham and Mann (1974) replaced the experimenter with another confederate who was assumed to be another participant in the study. In this case 25 percent of participants were fully obedient as opposed to the 65 percent in Milgram's experiment and other more standard replications. In another trial, a second 'experimenter' was added to contradict the first experimenter at 150 volts, saying the experiment ought not continue (Sabini and Silver 2005). No subject continued to shock after that point. In yet another version, when two peer subjects refused to continue, full obedience dropped to 10 percent (Milgram 1974). Given the data, we face the same choice we did in the case of the limbic room. Should we say that the virtues are absent or that the virtues are actually dependent on systems neutralized by the room? The extended virtues model says the latter and so preserves our intuition that people can be virtuous, but the model also recognizes that other people and environmental factors can play a role similar to that of parts of our brains.

The Zimbardo case yields a similar analysis. Individuals were placed in an alien circumstance where all the standard societal supports were replaced by other individuals playing a role. Being asked to play this role is akin to being asked to drop one's typical extended virtues—if not to suspend one's virtues altogether. (I largely agree with a critique of Zimbardo by Gray [2013] and Banuazizi and Movahedi [1975] that the experiment suffers from the existence of numerous demand characteristics that essentially tell the subjects how the experimenter expects they will behave.) According to the extended virtue hypothesis, Zimbardo's experiment does not test whether a single enduring disposition is elicited in the experimental conditions; rather, it smothers the disposition by replacing it with a new one that is corrupted by experimenter suggestions and expectations.

Of course, not only virtues are extended. Vices can be extended as well. Cases of group effects abound, often reflecting an increased complacency when it comes to moral situations. Latané and Darley (1968) describe an experiment of releasing acrid smoke from a vent into a room where undergraduate subjects were filling out forms. When only one subject was in the room, 75 percent of those reported the smoke. When two silent confederates were introduced, reporting dropped to 10 percent, and when three subjects were in the room only 38 percent of the time did anyone in the group report the smoke. In general, experiments have borne out the fact that likelihood to intervene in a 'troublesome' situation is inversely proportional to the number of people in the surroundings (Latané and Nida 1981).

While there are various interpretations of such group effects (see Latané and Darley 1970, for example) the extended virtue hypothesis offers a plausible picture of what is going on. When we are around other people, they become part of how we detect relevant features of our environment and even part of our moral characters. When no one is around, the skindividual is on his own. These external moral grounds can lead us in the right direction, as when a peer in the Milgram experiment insists that the shocks should stop. However, external factors can also mislead us and prevent us from seeing the morally salient parts of a situation.

The situationist challenge thus offers us an argument for extended virtues and against skindividualism. We are deeply wedded to the claim that people can be evaluated based upon their character traits. Some people are virtuous; others are vicious. What social psychology shows us is that these character traits do not exist in skindividuals. But this shouldn't lead us to abandon talk of character traits; it should lead us to recognize that people are not skindividuals. This suggests the following *modus ponens*:

1. Persons have traits and virtues.
2. Persons only have traits and virtues if virtues are extended.
   Therefore, virtues are extended.

If virtues are extended and people have virtues, then people are not skindividuals.

Of course, where there is a modus ponens, there is a corresponding modus tollens that some people will find much more tempting:

1. People can have traits and virtues only if there are extended virtues.
2. There are no extended virtues.
   Therefore, there are no traits and virtues.

According to the modus tollens, people are skindividuals, and their traits and virtues—if they have any—have to be grounded in them. Anything outside the skin is part of the situation, not part of the person.

How to decide between the modus tollens and the modus ponens? We must reflect on what we hold most dear in our concepts of character and personhood. It would be surprising, I think, if a metaphysical view about the boundaries of the person were crucial. It doesn't seem to be an essential part of the concept of a person, for example, that a person is a skindividual. After all, the traditional Cartesian picture is not skindividualist since it involves a soul that doesn't exist in space and thus cannot be within the boundaries of the skin. What goes to make up a person seems less important than the fact that a person is the sort of thing that can have virtues and vices and can be praised or blamed for having them. If we are deeply committed to the idea that there are persons—and it seems we are—shouldn't we be agnostic about what makes them up? Suppose a philosopher insisted that persons were souls, then became convinced that there are no souls. Would we think it more reasonable for the philosopher to conclude that there are no persons or for the philosopher to admit a mistake about the metaphysics of persons? Clearly the latter. Remaining wedded to some metaphysical picture about persons, to the extent that one is willing to deny that there are such things, seems stubborn and unnecessary. Our grasp of persons as loci of responsibility seems to be indifferent to our picture of their metaphysical constitution. This would seem to speak against the modus tollens.[1]

5. Coupling, Constitution, and the Boundaries of the Person

By now a common objection is screaming to be heard. The extended virtue theorist seems to be committing a particularly egregious version of what Adams and Aizawa call the 'coupling fallacy' (Adams and Aizawa 2008, 2010). Just because *A* is frequently coupled with *B* doesn't mean that *A* constitutes *B* or vice versa. It is incontestable that external things and other people significantly influence a person's virtues. But constitution doesn't follow from this causal connection.

It is true that causal coupling is not the same as constitution. But this point alone gets us nowhere. Parts of the mind are so not simply because of spatial location; it is the causal and functional integration that makes them part of the same 'cognitive system'. In some cases of integration we get parthood, in some not. What determines which cases are which will depend on what is important in the

---

1 Alfano (2013: 106) recognizes a social element in the nature of virtue, but his view considers virtue 'factitious' while mine does not. Merritt (2000) holds the position closest to mine, emphasizing 'the sustaining social contribution to character' (Merrit 2000: 374). This leads her to reject the Aristotelian model of virtue while the extended persons proposal retains it.

individuation of cognitive systems.[2] Similarly, parts of my mind and body are part of me as a person because of how they relate to my disposition to behave and feel in certain ways. Why must something be within the skin to play this role?

There is little to no reason to think persons as systems are governed by cognitive scientific individuation criteria, and there is reason to believe that they are governed by the norms governing persons and the ascriptions of personality. We can take the situationist critique as essentially making the point that the skindividualist way of cutting things up does not jibe with personality and trait talk. The extended persons approach does much better on this score. This suggests that in the case of the extensions of persons, we have reason to speak of constitution instead of mere coupling.

The skindividualist has a comeback, however. All of these supposed extensions can be screened off in various ways, and the individual still has the relevant character traits. What seems to matter for whether or not an individual has a certain set of virtues isn't the individual's environment but the way she represents that environment. Suppose Jean moderates her drinking in part because she believes her partner thinks she drinks too much. Perhaps she interprets various gestures and expressions as indications that the limit has been reached. But suppose she is wrong. In fact, Jean's partner wishes she would loosen up and is surprised but pleased when she orders the third drink. In a situation like this, the partner's actual beliefs are playing no role whatsoever in maintaining a trait. Jean's beliefs, internal to Jean, are playing that role based on her interpretation of her environment. This gives us reason to doubt extension in general.

This objection is serious, but at the very least it moves too quickly at the last step. That Jean's virtue isn't grounded in extensions in cases like this doesn't mean that people's virtues aren't grounded in extensions when they are in tune with their environment. Few would find the following argument persuasive: a brain lesion can result in screening off the effects of the amygdala on the rest of the mind. Regardless of the signals the amygdala sends, the area near the lesion determines the person's behavior. A fear response might be unreasonably generated, for example, because though the amygdala is silent, the area around the lesion is active. In this case the activity of the amygdala doesn't determine the individual's fear response; it is screened off. Since the effect of the amygdala can be screened off, the amygdala is not a part of the mind. It is perhaps causally coupled with the mind but not part of it.

This argument doesn't stand because the mere fact that a part of the brain can be screened off from the workings of the mind doesn't mean that it doesn't constitute part of the mind in normal circumstances. Similarly, just because in infelicitous cases the externalities that might constitute part of the ground of one's virtues are screened off doesn't mean that they don't constitute the ground of one's virtues in the felicitous cases.

2 What the individuation criteria for parts of cognitive systems are by no means a settled matter. For a positive proposal that would oppose extended cognition see Rupert (2009). For a push for more liberal criteria see Gallagher (2013a) and Wilson (2004).

Even if the coupling fallacy and the screening objection don't succeed as objections to the extended person's hypothesis, they do point to a task the defender of extended persons should take on. What, exactly, are the criteria for extension? Under what circumstances do external things, people, and relationships constitute part of a person or part of the grounds of that person's virtues and character traits?

These are fair questions, but answering them is not a minor project. This shouldn't surprise us since answering the question of what makes something part of a cognitive system is similarly tricky even on the skindividualist model. If anything, we should suspect that things will be even more difficult in this case since we are dealing with normative notions like personhood, blameworthiness, and ultimately agency. What we want to say about the parts of a person should track what we want to say in these related debates, and things in those areas are obviously less than settled. Nevertheless, we should be able to point the way forward even if we can't hope for necessary and sufficient conditions.

What we are looking for is a set of criteria in virtue of which a ground of a behavioral disposition can be considered part of the person. We can be guided in this rather abstract quest by the more intuitive question of when a behavioral disposition should be called a part of a person's character. Miller (2013: 12–13) suggests some elements of the 'functional role' of character traits, and we can use a few of those to guide us. One role traits play is explanatory—we can appeal to them to explain why people do what they do. Another is predictive—we can predict with some reliability what people will do based on their character traits. A third is evaluative—we use character traits to normatively evaluate a person. If a disposition grounded by something outside of the skin doesn't play these roles to a substantial degree, it isn't plausibly a trait. On the other hand, if a disposition can play this sort of role, it doesn't seem that its being grounded in features outside of a person's skin should prevent it from being a character trait.

These general features are helpful, but it is in application that their implications will become clearest. Consider an unsuccessful case of extension.

Tony Soprano is not a compassionate person even though he might appear to be so when at home with his wife Camilla. It would be implausible to say he's compassionate, but because he loses part of himself every time he leaves the house, he can't be blamed for butchering his adversaries. Camilla is not an extension. Why? A suggestion from the extended mind debate might be that Camilla is not around Tony enough to be a part of him. (The iPhone that I leave at home every day can't plausibly be an extension of my mind since it isn't reliably available; see Clark [2008: 79].) But the problem here isn't that Camilla isn't with Tony all the time. The problem is, among other things, the fact that Tony can so quickly duck out of the sight of his family and commit murder. Even if he were never to do so, the ease with which he could do so is a problem.

To see what's wrong with Tony, let's consider the simplistic analogue of a solar-powered traffic light. There are many ways to build such a light, not all of them equally good. A particularly bad design would be a solar traffic light that was only powered by the sun's rays during the equinox. Even if it worked well in that rare perpendicular light, it would not be a particularly good traffic light. One way to

think of Tony is like this—he is dependent on a particular sort of source in the way good people are not. Compassionate people find themselves powered, as it were, by more than their intimates. When they leave home, the presence of other people and their behaviors connect them with the source of compassion. One way to put this is that given an environment subject to change, the ground of virtue must be fungible. A solar traffic light could be bad in other ways, however. If its dependence on the sun were so strict that it failed to work on a cloudy day, it would also be a poor light. In typical environments there would need to be a power reserve that would not dry up the minute the external grounds were gone. Well-designed solar-powered traffic lights have to depend on the sun, but not in so strict a fashion that they go out when the sun is down. This is another problem with Tony—his dependence on Camilla is too strict given his environment. In both of these cases making Tony virtuous requires an internal change, but this doesn't make Tony a skindividual. Again, compare the solar-powered traffic light. There will be an internal difference that allows the light to be powered by the solstitial as well as the equinoctial rays. There will also be an internal difference that allows the light to work when the sun does not shine. Does the necessity of these internal changes mean we can say that all that really matters is internal to the traffic light? No. The traffic light does what it does only in virtue of its dependence upon an outside source. Only with an understanding that it is a solar-powered traffic light can we make predictions about when it will work, explain why it is and isn't working, and tell whether or not it is a good traffic light. To some extent, the degree to which virtues can be extended is itself a function of the environment. On a twin earth where the sun never sets, a solar-powered stoplight could be good without an internal battery. Similarly, if a virtuous disposition's external grounds are stable enough (and nonaccidentally so), there is less need for internal sources of moral power. This is perhaps the case for those individuals who might be virtuous because of the presence of civilization. That environment is stable, and nonaccidentally so, and if it suddenly were missing, it's quite possible many people who are now virtuous would no longer be so.

Most of us have character traits that are sustained by a diversity of things and people around us, and for most of us those characteristics don't simply evaporate when we are alone. Nevertheless, those things and people are part of what sustains our virtue and are in effect part of who we are. Partners and colleagues create an atmosphere that can sustain virtuous or vicious traits, and even though they are part of the ground of those traits, we are the ones who are virtuous or vicious. Other, nonsentient devices are apt to play increasing roles in supporting our characters. As phones become smarter and apps proliferate, it is likely that many virtues and vices might begin to be extended. Selinger and Seager (2012) document a number of 'Digital Jiminy Crickets' that keep people on the straight and narrow.

> In many cases, good-behavior technology gets the job done by bolstering resolve with digital willpower. By tweaking our responses with alluring and repulsive information, while also shielding us from

distracting and demoralizing data, digital willpower helps us better control and redirect destructive urges.

In the future, the grounds of our virtue or our vices might increasingly be located in hi-tech devices sitting in our pockets.

6. Conclusion

It has been a largely unquestioned assumption that what makes an individual the person he or she is must reside within the individual's skin. The most influential tradition maintains that a person is what he or she is because of his or her mind, and all factors outside of those narrow confines are external causes that affect the person but do not constitute him or her. When we limit our focus to conscious cognition, this can seem plausible, but a person is not made by conscious cognition alone. States of a person's body are clearly involved in whether that individual is virtuous, but if I am right, we must also look beyond a person's skin to find the dispositions that make that individual the person he or she is. A person is in part constituted by the social environment, which includes other people as well as surrounding technologies.[3] All of these features go toward making the individual vicious or virtuous and excluding them because they are outside of the skin risks arbitrariness as well as the very existence of character traits. All of this points in one direction. Persons are not skindividuals, but they are extended across their environment.

Many will remain uneasy about extended virtues. Among other things, there is the enduring sense that someone whose virtues are extended is less than fully virtuous. Not everyone who entered the Milgram room acted horrendously. Some defected. Aren't those people better in some moral sense? Don't those people, whose virtue appears to be more skindividualistic, deserve more credit for their virtues even in non-Milgram circumstances when everyone is acting well? I'm inclined to think that this is so, but this concession is consistent with the extended virtues thesis. Aristotle recognized that there is a difference between someone who has natural virtue and someone who has a more reflective understanding of his virtue. Nothing prevents us from recognizing the same distinction—very possibly many whose virtue is extended lack such a reflective, full virtue. Similarly, nothing prevents us from claiming that those whose virtue is less extended tend to have a different degree of virtue from people whose character is more constituted by their social environment. Recognizing this is akin to recognizing that extended memories have some less than desirable features or that a person on antidepressants should count as a less than perfectly happy or compassionate person because of features of his dependence. A complete description of virtue cannot neglect these shades of

---

3 Russell (2012) claims that not only our bodies but also our loved ones are part of the self (see 2012: 94–103 and chapter 9). Because he means something different by 'the self' (96), he is arguing for the fact that other people, etc., constitute part of our self-conception. I am arguing something stronger, namely, that other people and the environment constitute part of a subject's moral character.

character. Nor, however, can it continue to neglect the fact that for many of us what virtue we have is at least in part extended.

ROBERT J. HOWELL
SOUTHERN METHODIST UNIVERSITY
rhowell@smu.edu

References

Adams, Frederick R., and Kenneth Aizawa. (2010) 'Defending the Bounds of Cognition'. In *The Extended Mind*. (Cambridge, MA: MIT Press), 67–80.

Adams, Fred, and Kenneth Aizawa. (2009) 'Why the Mind Is Still in the Head'. In Philip Robbins and Murat Aydede (eds.), *Cambridge Handbook of Situated Cognition*. (New York: Cambridge University Press), 78–95.

Adams, Frederick R., and Kenneth Aizawa. (2008) *The Bounds of Cognition*. Hoboken, NJ: Blackwell Publishing.

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

Annas, Julia. (2011) *Intelligent Virtue*. Oxford: Oxford University Press.

Banuazizi, A., and S. Movahedi. (1975) 'Interpersonal Dynamics in a Simulated Prison: A Methodological Analysis'. *American Psychologist,* 30, 152–160.

Chalmers, David. (1996) *The Conscious Mind*. New York: Oxford University Press.

Clark, Andy, and David J. Chalmers. (1998) 'The Extended Mind'. *Analysis*, 58, 7–19.

Clark, Andy. (2008) *Supersizing the Mind*. Oxford: Oxford University Press.

Damasio, Antonio R. (1994) *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam.

Doris, John M. (2002) *Lack of Character: Personality and Moral Behavior*. New York: Cambridge University Press.

Fisher, Justin. (2008) 'Critical Notice of *The Bounds of Cognition*'. *Journal of Mind and Behavior*, 29, 345–57.

Gallagher, Shaun. (2013a) 'The Socially Extended Mind'. *Cognitive Systems Research*, 25–26, 4–12.

Gallagher, Shaun. (2013b) 'A Pattern Theory of Self'. *Frontiers in Human Neuroscience*, 7, 1–7.

Gallagher, Shaun. (2013c) 'Ambiguity in the Sense of Agency'. In A. Clark, J. Kiverstein, and T. Vierkant (eds.), *Decomposing the Will* (Oxford: Oxford University Press), 118–35.

Gettler, Lee T., Thomas W. McDade, Alan B. Feranil, and Christopher W. Kuzawa. (2011) 'Longitudinal Evidence that Fatherhood Decreases Testosterone in Human Males'. *Proceedings of the National Academy of Sciences*, 108, 16194–99. doi:10.1073/pnas.1105403108.

Gray, Peter. (2013) 'Why Zimbardo's Prison Experiment Isn't in My Textbook'. Available at: http://www.psychologytoday.com/blog/freedom-learn/201310/why-zimbardo-s-prison-experiment-isn-t-in-my-textbook.

Haney, C., W. C. Banks, and P. G. Zimbardo. (1973). 'A Study of Prisoners and Guards in a Simulated Prison'. *Naval Research Review*, 30, 4–17.

Harman, Gilbert. (1999) 'Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error'. *Proceedings of the Aristotelian Society*, 99, 315–31.

Harman, Gilbert. (2000) 'The Nonexistence of Character Traits'. *Proceedings of the Aristotelian Society*, 100, 223–26.

Harman, Gilbert. (2009) 'Skepticism about Character Traits'. *Journal of Ethics*, 13, 235–42.

Kilham, W., and L. Mann. (1974). 'Level of Destructive Obedience as a Junction of Transmitter and Executant Roles in the Milgram Obedience Paradigm'. *Journal of Personality and Social Psychology*, 29, 696–702.

Latané, B., and J. M. Darley. (1968) 'Group Inhibition of Bystander Intervention in Emergencies'. *Journal of Personality and Social Psychology*, 10, 215–21.

Latané, B., and J. M. Darley. (1970) *The Unresponsive Bystander: Why Doesn't He Help?* New York: Appleton-Century-Crofts.

Latané, Bibb, and Steve Nida. (1981) 'Ten Years of Research on Group Size and Helping'. *Psychological Bulletin*, 89, 308–24. doi:10.1037/0033-2909.89.2.308.

Merritt, Maria. (2000) 'Virtue Ethics and Situationist Personality Psychology'. *Ethical Theory and Moral Practice*, 3, 365–83.

Milgram, S. (1974) *Obedience to Authority: An Experimental View*. New York: Harpercollins.

Miller, Christian B. (2013) *Moral Character: An Empirical Theory*. Oxford: Oxford University Press.

Rupert, Robert D. (2009) *Cognitive Systems and the Extended Mind*. New York: Oxford University Press.

Russell, Daniel C. (2012) *Happiness for Humans*. Oxford: Oxford University Press.

Sabini, John, and Maury Silver. (2005) 'Lack of Character? Situationism Critiqued'. *Ethics*, 115, 535–62.

Selinger, Evan, and Thomas Seager. (2012) 'Digital Jiminy Crickets'. *Slate*, July 13, 2012. Available at: http://www.slate.com/articles/technology/future_tense/2012/07/ethical_decision_making_apps_damage_our_ablity_to_make_moral_choices_.html.

Upton, Candace L. (2005) 'A Contextual Account of Character Traits'. *Philosophical Studies*, 122, 133–51.

Upton, Candace L. (2009) *Situational Traits of Character: Dispositional Foundations and Implications for Moral Psychology and Friendship*. Lanham, MD: Lexington Books.

Weiskopf, Daniel A. (2008) 'Patrolling the Mind's Boundaries'. *Erkenntnis*, 68, 265–76.

Wilson, Robert A. (2004) *Boundaries of the Mind: The Individual in the Fragile Sciences: Cognition*. New York: Cambridge University Press.