

# Character in Kant's Moral Psychology: Responding to the Situationist Challenge

DRAFT

In recent years, “situationist” findings in psychology have been used by philosophers such as John Doris, Gilbert Harman, Mark Alfano, and others to criticize character-based ethical theories: “The experimental record suggests that situational factors are often better predictors of behavior than personal factors . . . To put it crudely, people typically lack character”(Doris 2002:2). Experiments have found that trappings of authority or artificial and arbitrary identifications with social groups can lead people to perform horribly, while finding a dime or feeling un-rushed can lead to particularly helpful behavior.<sup>1</sup> Because “virtue ethics presupposes that there are character traits of the relevant sort, that people differ in what character traits they have, and these traits help to explain differences in the way people behave,” virtue ethics is empirically false (Harman 2000:168). Thus, “[r]ather than striving to develop characters that will determine our behavior in ways substantially independent of circumstance, we should invest more of our energies attending to the features of our environment that influence behavioral outcomes” (Doris 2002: 146). As Alfano has helpfully summarized, “skeptics . . . argue that situational influences swamp dispositional ones, rendering them predictively and explanatorily impotent. It’s but a simple step from such impotence to the dustbin” (Alfano 2013:2).<sup>2</sup>

---

<sup>1</sup> See, for example, Darley and Batson 1973, Isen and Levin 1972, Haney et. al. 1973, Milgram 1963, Zimbardo 2007, and discussion of these and related studies in Doris 2002, Alfano 2013, and Miller 2014.

<sup>2</sup> While it eventually vindicates talk of character as a “moral fiction” useful instrumentally, Alfano 2013 develops earlier critiques into a book-length argument against traditional virtue ethics and virtue epistemology. A more complicated appropriation of recent empirical work on character can be found in Miller 2013. Miller’s helpful survey of the relevant literature provides an excellent albeit inadvertent foundation for my Kantian claim here that what the evidence shows is that human beings are “evil” rather than that notions of morality need to be amended, and Miller’s appeal to “mixed traits” as the best explanations for most human actions fits extremely well with Kant’s descriptions of human “impurity” (see §2.3 below).

Many have responded to these situationist challenges;<sup>3</sup> in this essay, I approach situationism from a specifically Kantian perspective. After first showing how the experimental data might seem to raise problems for Kant's moral philosophy, I show how Kant not only adequately incorporates similar sorts of data into his moral theory but also develops a framework for better interpreting and dealing with such data. In my conclusion, I briefly suggest what a Kantian empirical-psychological research program in moral psychology might look like.

## 1. Situationism as a challenge to Kant's moral philosophy

Generally, situationist criticisms of "virtue ethics" are directed towards broadly Aristotelian approaches, and one might think that a Kantian ethics focused on the permissibility of maxims of action would escape such criticisms. But Kant's ethical theory is *prima facie* as susceptible to situationist critique as Aristotelian virtue ethics. To see why, we must start with Kant's conception of "character." Character is the "distinguishing mark of a creature endowed with freedom" that "shows what the human being is prepared to make of himself" (7:285). It is defined as "that property of the will by which the subject has tied himself to ... firm principles (not shifting hither and yon like a swarm of gnats)" (7:292; see too 6:651-2; 25:1386).

Character in this sense is essential for a good will for three reasons. First, the good will acts on *maxims*, or practical principles, that conform to the moral law. Second, the categorical imperative itself is a *principle* to which one with a good will must be committed. One who does nice things for others merely from goodheartedness rather than a commitment to following the categorical imperative fails to have a good will. Finally, one's commitment to the moral law must have "purity and stability" (6:63). One might think that one could act out of principle on one occasion and not on another, or that one could have "fragmentary" virtues (see Doris 2002; Vranas 2005) or "mixed

---

3 See, e.g., Annas 2005; Badhwar 2009; Frierson 2010b; Kamtekar 2003; Kristjánsson 2008; Kleingeld 2015; Kupperman 2001; Sabini and Silver 2005; Sarkissian 2010; Upton 2009; Webber 2006.

traits” that partake of both virtue and vice (Miller 2013, 2014). But Kant rejects all such forms of moral “latitudinarianism” (6:24-5). Because the moral law is a law of *universality*, one who truly acts in accordance with it commits herself to *unconditional*, principled action. Any transgression shows that the moral law, which is supposed to be universal, is applied in some cases but not in others, so that something other than the moral law itself must be governing one’s motivation. It follows that in order to have a good will, the stable disposition to act on the categorical imperative, one must have *character*, that is, a stable disposition to act on a fixed principle. Thus, whereas many situationists see virtuous *conduct* as “the closest we can come to actual virtue” (Alfano 2013:102; cf. 185), for Kant, character as such is actually closer to virtue than virtuous conduct without principle (see especially 25:631 and Frierson 2006).

Kantian character is not the sort of settled disposition of feeling and action on which many virtue ethicists focus; instead, it is a self-conscious and enduring commitment to practical principles. Nonetheless, this character is as essential for Kantian virtue as any “character” in contemporary virtue theory, which arguably makes Kant’s moral philosophy even *more* vulnerable to empirical research that shows the situational variability of human actions. Even if one might allow for complex and situation-dependent dispositions within a traditional virtue ethic, there seems to be good empirical evidence that people lack the established commitments to principles that *Kant* requires for the good will.

## 2. The Kantian response to situationism

While Kant *seems* vulnerable to situationist critique, this vulnerability is merely apparent. Recent empirical research, properly understood and interpreted, actually supports Kant, properly understood and interpreted. Five key features of Kant’s view of human beings provide a philosophical framework that is an important alternative to recent philosophical appropriations of

situationist research. First, he admits (or, better, anticipates) the empirical reality of situation-dependence in human action. Second, in sharp contrast to philosophers like Doris, Alfano, and even Miller, Kant articulates a conception of “ought implies can” that vindicates his moral theory in the face of evidence of the rarity of character. Third, he provides an interpretive framework for the situation-dependence of most human motivation in terms of what he calls humans’ “propensity to evil” (6:37). Fourth, he provides a framework for highlighting empirical bases for moral hope, which sees lack of character as something human beings can overcome. Finally, he offers a framework for developing what Alfano has called “moral technologies” and what Kant calls “moral anthropology.” In Kant, these technologies are oriented not merely towards good *behavioral* outcomes (as in Doris and Alfano) but rather towards the cultivation of good character itself.

## 2.1. The rarity of character

Situationist critics of character-based ethical theories tend to portray such theories as relying on an empirical psychology according to which action motivated by general, stable, relatively-fixed character traits is the norm, where the function of ethics is merely to pick out which character traits constitute virtuous ones (e.g. Harman 2000:168). But this is not the only way in which an ethics might be “character-based,” and Kant’s moral philosophy provides an excellent example of a character-based ethical theory that does *not* rely on the assumption that all or even many people are motivated by character in his sense. As we have seen, Kant insists that generality and stability are central to what it means to have “character,” and character is essential to having a good will. For Kant, however, *both* character *and* good principles are “acquired” rather than natural endowments, and both are “rare” (7:294, 292; cf. 25:1174).<sup>4</sup> Kant thus often refers to one “who has no character” (25:822; 25:1170) and even outlines groups that are particularly unlikely to have character, such as those under forty (25:1172; 7:294), those with “sanguine temperament” (7:287-8), priests and artists

---

<sup>4</sup> Alfano 2013:23 recognizes this feature of Kant’s view.

(25:1172), and – infamously – “the fair sex” (2:232; 7:308). By emphasizing a form of character that is rare, Kant is closer to situationists like Doris and Alfano than to some recent attempts to rehabilitate character through appeal to “mixed traits” or other alternatives (see Miller 2013, 2014; Bahwar 2009; Kleingeld 2015; Sabini and Silver 2005).

Partly because of its rarity, Kant argues that character in the strict sense “has something precious and admirable to it” and “has an inner worth which is above all price” (7:292; cf. 6:651-52). This claim is supported by situationist research, which shows not only that people lack character but also that this lack of character is precisely the “shifting hither and yon like a swarm of gnats” (7:292) that Kant rails against as a serious problem for both morals and individual happiness. The sense that we can be brought to do just about anything by an authority figure in a lab coat (Milgram 1974) or by being made to feel good in trivial ways (Isen and Levin 1972) reinforces one of the central motifs of Kant’s anthropology, that in order to truly take charge of our own lives we need to govern ourselves by fixed and stable principles.

Given that character in the strict sense is rare, Kant lays out a complex psychological account of how human beings *are* typically motivated. In his published *Anthropology*, he contrasts those with character, “from whom one knows what to expect . . . from his will” with those “from whom one knows what to expect . . . from his instinct” (7:285, see too 25:822). This contrast seems to imply a strict dichotomy between willed, character-based action and merely instinctual behavior, but elsewhere Kant develops the motivational possibilities more richly. He explains that for “a human being without character . . ., with each opportunity he is another human being” (25:1170; cf. 25:1176), implying that one might exhibit something like character, where this “character” is insufficiently stable for one to count as the same person from one moment to the next. In one of his best descriptions of those without true character, he explains,

Those who have no character at all take on a semblance of character. They pretend to have rules and principles, and imitate something character-like, because they do not have any character in themselves. Such people make general rules for themselves, for example, not to lend anything to anyone or not to believe anyone who promises them something. They do not have a bad character, they just have no principles, because they are often deceived by their good heart, and cannot rely on it at all ... [One] who has no character at all, is not at all in accord with himself; today he gives something to everyone, he is very liberal, but tomorrow he becomes very frugal and stingy, because he sees that he comes up too short, if he does it this way every day. Hence he quickly catches hold of another rule, and thus with him nothing is dependable. Such people are like soft<sup>5</sup> wax; every instant they catch hold of another rule. (25: 630-31)

The extended account here clarifies how instincts and changing situations affect deliberation. It is not generally the case that one acts *purely* from instinct or inclination, even when one lacks a commitment to firm principles. Instead, those without character “every instant catch hold of another rule.” Rather than applying consistently-held maxims *to* particular situations, those without character allow fluctuating situations and passing inclinations to determine which maxims will hold sway. This can range from merely unconscious motivational processes (mere inclination) with at best principled after-the-fact confabulations to (subtly) inclination-influenced deliberations based on principles.<sup>6</sup> And these claims, far from being in conflict with situationist research, fit extremely well with research that shows how situational factors prompt alterations in decision-making (e.g. Nisbett and Wilson 1977; Wilson 1998).

---

5 The Academy Edition has “hard [*zähes*] wax” here, but a variation uses “soft [*weiches*],” which makes more sense in the context.

6 For more detailed discussion of these forms of volitional failure, see Frierson 2014:215-58.

## 2.2. The priority of moral theory over empirical research

Given Kant's recognition that character is rare, one might expect him to share Doris's view that the "moral psychology ... better suited to effecting the practical aims of ethical reflection" will be one that recognizes the role of situation by "invest[ing] more of our energies attending to the features of our environment that influence behavioral outcomes ... [r]ather than striving to develop characters that will determine our behavior in ways substantially independent of circumstance" (Doris 2002:146, order reversed). Even if one does not want to focus exclusively on behavior, Kant's own principle that ought implies can (see 5:30) seems to better support Doris and Alfano's efforts to seek moral standards that conform to what is generally discovered about human psychology than Kant's insistence that people ought to have a character that psychology shows not to be present.

To see how Kant's approach to the principle that ought implies can fundamentally differs from that of contemporary situationists, it is important to highlight an important feature of situationist arguments against the moral requirement to have character in a strict sense. Typically, such arguments take data about how people *actually behave* as a guide to what is *possible* for them. Thus, for instance, Alfano rejects as "obviously implausible" a strong ethical "egalitarianism" according to which "anyone could reach the normative ideal of sainthood" on the grounds that we don't see many people (if any) *actually achieve* this ideal (see Alfano 2013:32-33). Similarly, in support of the claim that "virtue ethics is descriptively inadequate," he argues "that not enough people do *or could* possess the sorts of traits virtue ethicists care about" (Alfano 2013:82, emphasis added), where the evidence for the "could" claim is given by a description of what traits people *do* possess. Throughout, arguments proceed from what people in fact typically do, to what we should consider people capable of, and thenceforth to what constitute reasonable moral demands.

Situationists' criticisms are often directed against virtue ethical approaches that purport to be

predictively powerful, but their arguments' structure seemingly threatens Kantian approaches. Sabini and Silver helpfully summarize this structure:

Ethics is otiose if it prescribes behavior that people cannot perform or, more generally, if it urges people to be the sorts of people they cannot become. But if empirical research discovers that no one does (or is) what virtue ethics say they should do (or be), then that is, at least, *prima facie* evidence that people cannot do (or be) what virtue ethics requires. (Sabini and Silver 2005:537-8)

The basic argument has three steps.

- (1) Statistically few people exhibit cross-situationally consistent (general) character (traits).
- (2) Thus, few people are capable of cross-situationally consistent character.
- (3) Thus, moral theory should not require cross-situationally consistent character of all people.

The move from (2) to (3) is grounded in the principle that ought implies can. If ought implies can, then can't implies needn't. The move from (1) to (2) is a non sequitur, and is probably best understood in the more mitigated way that Doris and Stich present it, as a burden-shifting move rather than a straightforward and fallacious inference from "don't" to "can't." That is, given that people don't *actually* exhibit character, "the burden of argument has importantly shifted: The advocate of virtue ethics can no longer simply assume that virtue [or character] is psychologically possible" (Doris and Stich 2005:121).

In this section, I start with the move from (2) to (3), arguing that the situationist move from cannot to need-not gets the import of Kant's principle precisely backwards; for Kant, the principle should be used to argue *from* ought *to* can, not vice versa. I then turn to consider the burden shifting move from (1) to (2), showing in what sense – for Kant – evidence about what *is* the case can justify claims about possibility, but ultimately arguing that Kant – through his transcendental idealism but more basically in a methodological priority of moral theory – undermines this move in a way that

vindicates the emphasis on character in his moral theory, regardless of whether psychological evidence would support the view that such character is psychologically impossible. In §2.4, I argue that given the actual state of empirical research, one can defend the possibility of character even on empirical grounds.

First, then, a bit on the principle that ought implies can. Kant articulates this principle after a pair of hypothetical scenarios presented to one who “asserts of his lustful inclination that ... it is quite irresistible.” In the first, “a gallows [is] erected ... and he w[ill] be hanged on it immediately after gratifying his lust” and in the second, “his prince demand[s], on pain of the same immediate execution, that he give false testimony against an honorable man” (5:30). With respect to the latter case, Kant says,

He would perhaps not venture to assert whether he would do it or not, but he must admit without hesitation that it would be possible for him. *He judges ... that he can do something because he is aware that he ought to do it.* (5:30, emphasis added)

Importantly, Kant does *not* suppose any empirical evidence that one can do the right thing even when faced with death. He nonetheless claims to judge justifiably that it is possible on the grounds that it is morally required. Kant reasons *from* ought *to* can. That one recognizes a duty shows that there is a way, however psychologically difficult, of fulfilling that duty (see too 6:29, 32). Morally speaking, we are entitled – and in fact required – to assume that character is possible for all human beings.

This moral defense of the possibility of character might seem to contradict psychological data that implies the impossibility of character. To fully respond to situationists, Kant cannot simply reverse from the argument from (2) to (3). He must undermine or mitigate the argument *for* (2). One key way in which Kant aims to mitigate that argument is to reject the apparent dilemma according to which we must either modify our moral theory to account for our best psychological theories about

what's possible or we modify our psychology to allow for possibilities that (for the sake of argument we assume) never occur. For Kant, we should do neither, but instead maintain two senses of possibility, while according priority to the moral over the non-moral one.

Kant's defense of the compatibility between moral and empirical-psychological ways of conceiving of human action draws on his transcendental idealism, a full discussion of which is far beyond the scope of the present paper.<sup>7</sup> Any of several different interpretations of that idealism would support the argument of the present paper (see, e.g., Allison 1990; Kohl 2015; Korsgaard 1996; O'Neill 1989). On more metaphysical interpretations of this idealism (e.g. Allais 2015, Ameriks 1982, McCarty 2009, Watkins 2006), the basic idea is that even if all human actions as appearances in the world are governed in accordance with causal laws, those actions can at the same time be ascribed to human freedom. Thus the situation-dependence of actions is itself a *consequence* of humans' (noumenal) freedom, rather than an *ultimate* explanation. Less metaphysically-loaded approaches to Kant's idealism highlight instead the different standpoints – practical-moral and empirical-scientific – one can and must take on human actions. On these approaches, the practical-moral standpoint must ultimately be seen as more fundamental and more complete, because, as Onora O'Neill has put it, “all naturalistic explanations—even the most impressive explanations of some future neuroscience—are conditional explanations ... for they can never explain that any natural law should take the form that it does” (O'Neill 1989:68). On either account, the key point is that however much human behavior conforms to empirically accessible laws, one cannot show the necessity of the ultimate ground of those laws, so that it remains open to ascribe human agents responsibility for the lack of character that they demonstrate. With an independent argument that lack of character is morally evil, the prevalence of that lack shows only the prevalence of evil, not any reason to reject character-based moral theory.

---

<sup>7</sup> I present my reading of Kant's transcendental idealism as it applies to the issues in this paper in Frierson 2003, 2010a, and 2013.

Even if one rejects transcendental idealism altogether, one can still accept Kant's fundamental methodological point about the relationship between empirical psychological theory and moral theory.<sup>8</sup> For one thing, even without a robust transcendental idealism, one can legitimately distinguish moral from psychological possibility. For the purpose of empirical prediction and explanation, scientists assume that phenomena that consistently occur in particular ways do so because of underlying laws. In seeking, finding, and articulating those laws, they make claims about what is empirically possible based on what actually occurs. In that sense, the inference from (1) to (2) is consistent with scientific practice. What it means for something to be "possible" in this context is for it to conform to laws describing what things actually do.<sup>9</sup> This methodological principle of natural science, however, leaves room for another sense of possibility. Moral philosophy can and must develop its own account of what is possible, based on moral arguments for moral obligations. Just as it is illegitimate to infer that phenomena will take a certain form merely because they ought to, so it is illegitimate to infer that they cannot take a certain form (in the morally relevant sense of "cannot") merely because, in our experience, they do not in fact take that form. Because the possibility established by natural science and the possibility required by moral theory are different, one cannot infer from empirical or psychological impossibility to moral conclusions.

While transcendental idealist claims are particularly plausible in the context of moral psychology, and the way transcendental idealism supports the priority of moral over empirical reasoning in this area provides a strong practical reason for supporting it, all that Kant needs

---

8 For a relevant objection to the two-standpoint approach to transcendental idealism, see Nelkin 2000. Ultimately, however, I think that transcendental idealism can be vindicated (see e.g. {author paper}).

9 For Kant, the distinctive status of psychology makes this analysis all the more apt. Kant sees physics as having a priori laws that condition any metaphysically possible experience of material nature. Psychology, by contrast, articulates "laws" only in the sense of empirical generalizations, a "historical systematic natural doctrine" (4:471). Moreover, even in the merely empirical realm, the difficulties of empirical psychological research into human motivation preclude laws as determinate as those of, say, chemistry. While Allen Wood goes too far when he claims that this impossibility of "satisfactory empirical knowledge of the mind" makes possible Kant's "conjectures about noumenal freedom" (2003:50), he rightly sees how, even without a *comprehensive* transcendental idealism, the specific problems of psychology make an emphasis on moral freedom plausible for human actions.

(whether from transcendental idealism or elsewhere) for responding to situationism is a methodological priority of moral theory over psychological research *for the purposes of moral theory*. That is, insofar as moral theorists are able to show – as Kant purports to show<sup>10</sup> – that morality requires character (or any other condition), the moral function of empirical psychology would be to determine whether, to what extent, and how human beings meet (or fail to meet) that requirement. Whether or not one accepts transcendental idealism, there is no good reason to reject an otherwise sound moral theory merely because, given the data one has about human behavior, it implies widespread human evil. The sort of can implied by ought is not the sort of can that would be justified *or undermined* by observations of what is.

### 2.3. Kantian moral pessimism

In §1, I showed how, on the basis of a priori moral theory, Kant argues that the good will requires not merely good deeds but consistent character that stably acts on the basis of moral principles. But situationist research seems to show that all human motives are fragile, susceptible to influence by features of their situation that can override or undermine moral principles. §§2.1-2.2 showed how Kant acknowledges the facts to which situationists have drawn attention while preserving his moral theory through arguing that the possibility required for a moral obligation to have character is not precluded by the empirical evidence against the occurrence of character in human beings. However, even if this result insulates Kant's moral theory from situationist critique, it does little to vindicate that moral theory as a helpful framework for integrating empirical psychology and moral philosophy. In this section, I begin a shift from the merely negative task of defending Kant against situationist critiques to the positive task of laying out a Kantian framework for interpreting and responding to the data. The first step, in this section, is to show how Kant's

---

<sup>10</sup> Establishing the soundness of Kant's moral theory would be as large a project as defending his transcendental idealism. My point here depends only upon there being some method of moral reasoning that does not rely on empirical psychological theories.

discussion of human evil can not only accommodate situationist research but can provide a framework for better interpretation of that research and even for more focused future experimentation. In §2.4, I show how Kant provides a basis for conducting and interpreting experiments towards the justification of moral hope. And in §3, I turn to the most important role of empirical psychology within a Kantian framework, the development of moral technologies for the cultivation of character.

While philosophers like Alfano, Doris, and Miller see situationist research as the basis for shifting moral theory towards focusing on situation and behavior rather than character as such, on the grounds that “[i]f it’s the closest we can come to actual virtue, it would be strange to pooh-pooh it” (Alfano 2013:102), Kant sees the empirical data as evidence that human beings are “radical[ly] evil” (6:18). For Kant, observation and experiment show that evil corrupts the most fundamental basis of our choices; in this general context, lack of character is one – perhaps the most typical – expression of this evil. Evidence of the situation-dependence of human actions thus indicts *human actors*, not the character-dependent moral standards we fail to live up to.

Kant’s argument for humans’ radical evil involves precisely the points made in this paper thus far: moral goodness requires consistency and stability in willing (§§1, 2.2) and people lack consistency and stability in willing (§2.1).<sup>11</sup> In *Religion*, where he develops his account of human “evil,” Kant’s key move is his affirmation of “moral rigorism,” the view discussed in §1 that “a human being can[not] be morally good in some parts, and at the same time evil in others” (6:24). He elucidates the claim that “the human being is *evil*” by explaining that this “cannot mean anything else than that he is conscious of the moral law and yet has incorporated into his maxim the (occasional) deviation from it” (6:31). Kant’s identification of evil with “occasional” deviation highlights how

---

<sup>11</sup> The nature of Kant’s argument for humans’ radical evil is hotly contested. Anderson-Gold (2001) and Wood (1999) argue for broadly empirical arguments, while Allison (1990; 2001) and Grenberg (2005) defend a priori arguments. For a concise articulation and defense of my view, see Frierson 2013:72-81.

*inconsistency* – or the lack of character in the strict sense – is a paradigmatic form of human evil.

Given this moral rigorism, he establishes the fact of human evil by appeal to “the multitude of woeful examples that the experience of human *deeds* parades before us” (6:32-3), including a classic phenomenon within contemporary situationist psychology: “to us someone already counts as good when *his evil is common to a class*” (6:33, cf. e.g. Zimbardo 2007). In the end, human evil is summed up in the maxim that “every man has his price, for which he sells himself” (6:38). As Christian Miller helpfully put it, “No one thinks that the possession of *deep* honesty or compassion or any other folk moral virtue is common, and we do not need psychological studies in order to reject such a view. It is obvious from our experience of the world around us” (Miller 2014:96).

Kant does not *merely* diagnose human beings as evil. He offers a subtle taxonomy of different kinds of evil, each of which has corresponding ways in which it will sometimes – even often – show up in good behavior. For Kant, the universal human “propensity to evil” takes three main forms, which he labels “frailty,” “impurity,” and “depravity” (6:37-8). Each manifests susceptibility to situational variables. The first, frailty, has the most obvious situational susceptibility. As I have argued elsewhere (Frierson 2014:232-48), frailty has a range of degrees, from being overwhelmed by affects to complex rationalizations of doing what one has resolved not to do. In every case, one’s lower faculty of desire – our sub-rational and non-deliberative volitional tendencies – exerts influence on action. In the case of affects such as shock or sudden rage, reflection becomes impossible and one hardly can be said to “choose” at all; one’s “faculty of choice,” in Kant’s technical terminology, is simply suspended or bypassed. But in other cases, sub-rational volitional tendencies – what Kant calls the “lower faculty of desire” – can provoke and redirect rational reflection through various forms of “[self-]deception” (7:151). With this broad conception of frailty, it is easy to see how much situation-dependence of action can be seen as falling under it. In the Princeton seminary case, for example, a general resolution to help others in distress may be

overwhelmed by the urgency of making it to one's next appointment, and this overwhelming-by-urgency can range from affect-induced literal blindness to one in distress to the more complicated rationalizations involved in saying things such as "Well of course this person should be helped, but there are plenty of others nearby with less pressing business than mine."

Kant's clear provision for the influence of what he calls the "lower faculty of desire" on human action provides a natural way to account for the systematic role of "situational non-reasons" (Alfano 2013:43ff.) such as moods or ambient smells. Unlike Christian Miller – who has to make the far-fetched claim that "if ... we ... adopt the interpretive perspective of the person actually faced with these situations, then we might realize that this one 'slight' difference actually matters a lot to him at some level of conscious or subconscious psychological processing" (Miller 2014:55) – Kant can allow that there are situational factors that don't *matter* to a person qua rational agent, but that affect his lower faculties of desire and thereby influence his decision-making. This could be seen as a sort of subconscious mattering, but that move risks collapsing "the perspective of the person" and the explanatory perspective of the scientist (ibid.). Instead, Kant can say that, for example, a scent doesn't *matter* to me, but because it has a causal effect on my inclinations, and because I – in my frailty – allow inclinations to govern my conduct in non-principled ways, such a scent can causally influence my behavior.

Frailty is classically situation-dependent, but Kant's other kinds of evil – impurity and depravity – also provide nuanced interpretations of many cases of situation-dependent behavior. Impurity of will occurs when one's "maxim is good with respect to its object" but "has not, as it should ..., adopted the law *alone* as its *sufficient* incentive but, on the contrary, often (and perhaps always) needs still other incentives" (6:30). If we think about how such an impure will would appear to observers or empirical psychologists, it fits extremely well with situationists' accounts of human

beings.<sup>12</sup> People with impure wills often do what morality requires, but in extreme or unusual cases, where the moral law speaks clearly but customary supplementary incentives are absent, they falter. Alternatively, in cases where new supplementary incentives are provided, or where customary incentives are highlighted, they tend to exhibit more morally praiseworthy behavior. In Milgram's experiments, for instance, fear of embarrassment (see Sabini and Silver 2005) or desire for approval from accepted authority figures, two sorts of love of honor that generally support the moral law, conflict with the demands of the moral law. One with an impure will – who in a well-ordered society generally acts well and is well-regarded – suddenly loses an important support to upright behavior. Having never cultivated a character that makes duty sufficient for action, she falls short in this case. And for an equally impure will, unexpected good fortune – even if only a dime (see Isen and Levin 1972) – can provide the emotional boost needed to motivate acts of beneficence, especially beneficence that is relatively easy to perform and rewarded with social esteem.

Even Kant's most extreme category of evil, the "depraved will" that consistently and deliberately "subordinate[s] the incentives of the moral law to others (not moral ones)," will nonetheless "still [perform] legally good actions" (6:30) because the world in which we live is one within which non-moral incentives often promote the same ends as moral ones. Depraved wills are more likely than frail or impure ones to find themselves amongst what we might call the "immoral minorities," the 33% of low-hurry seminarians or the 12% of dime-finders who didn't help the person in need (see Darley and Batson 1973; Isen and Levin 1972). But they are still highly influenced by situation because the details of situations affect not only how one can best satisfy

---

12 In fact, Kant's concept of impurity is similar in many ways to Miller's notion of "mixed traits." The moral ambiguity that Miller identifies in these traits makes them, for Kant, fundamentally different *in kind* from the good will, and hence (given rigorism) essentially *evil*. Perhaps even more disturbingly, Kant's account of the impure will is echoed by Frimer's description (in Frimer et. al. 2011) of "moral giants," who impress with their moral excellence but who achieve this through effective integration of self-interest with other-regard. While there is a good way to effect this integration, Frimer's description suggests that in many cases, moral heroes are really just people with impure wills and generally good inclinations.

one's non-moral ends, but even what non-moral incentives one takes oneself to have; sudden joy, even merely at finding a dime, might help generate an immediate and hence non-moral inclination to spread that joy.

Seeing susceptibility to situation in terms of radical evil provides Kant with a substantive explanatory framework for the empirical data. If good character is rare, and if human evil can manifest in good conduct in various situations, then one would expect there to be a high situation-dependence to human actions (cf. Sabini and Silver 2005:542-4). Moreover, Kant's framework helps categorize important distinctions between different ways that people are situationally susceptible. It may turn out, of course, that Kant's three-fold classification of human evil is too simplistic<sup>13</sup> or simply misguided, but *prima facie* it provides for more subtle, interesting, and differently-oriented hypotheses than those of Doris or Harman (or even Milgram, Darley, or Bateson) about the variety of psychological factors at play in the situation-dependence of human action, and it is at least as plausible as Miller's "mixed trait" account of the data. In particular, Kant's approach highlights explanations that get at *morally important* (in his sense) features of motivation, rather than merely at factors useful for prediction and control. And insofar as these morally-important factors trace back to one or more forms of a propensity to evil, Kant does not compromise his fundamental moral principle that character is essential for having a truly good will.

## 2.4. Moral Hope

For Kant, data showing that human behavior is highly situation-dependent provide evidence of human evil. The principle that ought implies can, combined with a subordination of empirical possibility to moral possibility (whether justified by transcendental idealism or not), shows that we

---

<sup>13</sup> Kant himself would see the present account as too simplistic. For one thing, several forms of situation-dependence are due not to human evil but to the fact that many duties are imperfect, requiring that one adopt and advance various ends (such as the happiness of others) without specifying precisely how, when, or to what degree one must work towards this end. One with a firm commitment to being generous need not be generous to *everyone all the time*, so there is room for situational variables to play a role in governing the implementation of this commitment to duty.

“can” (in the morally relevant sense) have the sort of character that is morally required. But one might still worry that this philosophical victory rings hollow without some plausible way of seeing *how* humans can overcome the situation-dependence implicit in radical evil. Kant himself worries about the practical force of morality for those who recognize themselves as the sorts of beings that, albeit through their own fault, are now character-less and thus incapable of moral virtue. When he insists, “the command that we *ought* to become better human beings still resounds unabated in our souls; consequently, we must also be capable of it” (6:45), he means more than merely that we ought to act in ways that we know we will not act. He means that there must be a way, for each individual given where they are right now, to cultivate and act from a morally good character.

Kant defends this stronger possibility of character in three ways. The first, an appeal to a supernatural influence that can support our own efforts towards virtue, is not likely to be seen as a particularly helpful part of a Kantian approach to empirical psychology.<sup>14</sup> A second – “moral anthropology” – is the topic of §3. A third element of defending what he calls “hope” in moral reform is Kant’s emphasis on the persistence of our “moral personality” – our sense of duty – even in the light of our failure to follow it. The “hope that good will develop in us” is “awakened by our belief in our original moral predisposition to good” (7:43); the very fact that we still recognize the authority of the moral law implies that we have psychological resources for acting in accordance with it. Kant develops a philosophy of history within which he highlights empirical evidence of this still-present moral disposition. Most famously, he notes the “disinterested sympathy” of spectators with the revolutionaries in the French Revolution as evidence for “a character of the human race at large ... at least in its predisposition” that offers “hope for progress towards the better” (7:85). For Kant, in other words, empirical evidence is not *uniformly* negative; even in the midst of much evidence of evil, one also finds – and should specifically look for – evidence of the possibility of

---

14 I discuss this component in Frierson 2003:114-22; Frierson 2007. See too Mariña 1997; Michaelson 1990; Pasternak 2013.

moral improvement.

Recent psychological research supports this Kantian hope. In most studies where subjects behave immorally (most notably Milgram), such behavior is accompanied by significant moral distress, vindicating Kant's emphasis on the persisting possibility of motivation by moral principles. And several studies have shown the importance of recognizing personal responsibility and/or the authority of moral norms for supporting persistent adherence to those norms (see Casey and Burton 1982; Dweck 2000, 2006; Kohlberg and Candee 1984; Mazar et. al. 2008). But situationist research, not to mention Kant's own concerns about lack of character, shows the limits of this response. Kant's evidence for the predisposition to the good is sufficient to justify the possibility of a "moral" motive in particular cases, but not to justify the possibility of the character necessary for such motives to be *enduring* and thus *genuinely* moral. Insofar as the threat to virtue is lack of character, the persistence of our recognition of the authority of the moral law does not provide an adequate basis for Kantian hope.

Fortunately, even within the studies frequently cited by situationists, there is evidence for the possibility of stable character commitments. Often overlooked by philosophical appropriations of this research, the experimental data show that the situation-dependence of human actions is *not* universal but *varies* considerably between different people. Even in the most extreme experiments, we only *very* rarely find 100% of test subjects' behavior to be governed by situational variables. For Kant, that most subjects in the Princeton experiment neglected to offer aid to a person in need when they felt rushed is a symptom of human evil (lack of character). But the fact that 10% of these rushed subjects *did* stop provides empirical support for the transcendental claim that all *could have* stopped.<sup>15</sup> And in the Milgram case, not only did many subjects end their participation in the

---

<sup>15</sup> For a concise response to exaggerated claims of the impossibility of character, see Peterson 2006:93.

experiment (see Milgram 1973), but there is clear anecdotal<sup>16</sup> and statistical<sup>17</sup> evidence of traits of character shared by those who stopped as opposed to those who did not. Thus, for example, in one version of Milgram's experiment, only 30% of subjects refused to continue the experiment to its conclusion, but when later experimenters interviewed the subjects and classified them in terms of general moral conscientiousness, they found that *no* non-conscientious person quit and nearly 90% of morally conscientious people quit (Kohlberg and Candee 1984:69; see too Kilham and Mann 1974).<sup>18</sup> Moreover, there is at least some evidence that these or similar traits of character can be cultivated (e.g. Kohlberg 1981; Dweck 2000, 2006; Gollwitzer and Bargh 1996). Situationists typically focus on how one can modify environmental conditions to improve good *behavior*, relinquishing any focus on character. But there is evidence that some people do have character, or at least the pre-conditions of character. This provides *prima facie* grounds for hope that character can be cultivated in people in general.

For Kant, the best research on the situation-dependence of human action would *both* highlight the problem of lack of character *and* provide the evidence for hope in moral improvement; extant data confirm both Kantian claims. Unfortunately, however, many studies of situation-dependence are conducted in ways that *only* show lack of character in the aggregate and make it impossible to see the extent of divergence amongst individual participants. Milgram did little in the way of follow-up and characterization of those who did not fit his typical pattern; Darley and Bateson did none. Most egregiously, many studies focused on the situation-dependence of cheating on tests (e.g. Mazar et. al. 2008) look only at aggregate scores and compare average scores of groups under different conditions to show that some groups cheated more than others, but this does

---

16 E.g. Milgram 1974 and the insightful analysis by Roberts and Wood (2007:316-19) of a subject who refused to continue with the experiment.

17 See Kohlberg and Candee 1984.

18 "Morally conscientious" is my term to describe what Kohlberg and Candee call a "type B" person, who is "oriented to fairness rather than to rules or pragmatics" and "who ... in her ... 'conscience' perceives the central values or obligations ... and uses these intuitions to generate a judgment of responsibility" (Kohlberg and Candee 1984:63).

nothing to show *how many* people cheated; it could be that a few cheated very egregiously, or that all cheated a little, or that most but not all cheated a little.<sup>19</sup> Because such facts would be relevant for justifying not only human evil but also moral hope, Kant would rightly call on experimenters to gather more and better data.

Ultimately, Kant's approach to the study of character requires moving beyond psychological studies of lack of character to studies of how far we might expect character to be cultivatable. And this requires not only cataloging cases of divergence but also, and more importantly, investigating specific methods for cultivating character. Rather than seeing lack of character as a psychologically-sound reason to change moral theory, Kant rightly sees it as a practical moral *problem*, one that a renewed and refocused empirical psychology can help solve. In §3, I focus on this important direction for Kant-inspired research on human character.

## 2.5. Kant's Dodge?

Before turning to this new Kantian approach to empirical psychological research, it is worth briefly addressing one final objection to Kant's overall response to situationism. In many respects, his response is not unique; like certain neo-Aristotelian approaches (e.g. Kamtekar 2003; Miller 2003; Swanton 2003), it represents what Mark Alfano calls the "dodge" strategy:

Advocates of the dodge point out that for the ancients, full virtue was a rare last fruit of a lifelong project. That most people are not virtuous is therefore no surprise; in fact, it may be a prediction of a suitably elitist virtue ethics. (Alfano 2013:63)

Alfano objects to such a strategy on two main grounds. First, it "appears to cede the *egalitarianism*, *explanatory power*, and *predictive power* conjuncts in the hard core of virtue ethics" (ibid. 63). Alfano here

---

<sup>19</sup> Where standard deviations are provided for these averages – as for instance in Shu 2011 – they are often quite large; even in these cases, insufficient information is provided to determine whether any systematically varying factors explain who is affected by situation is who is not.

argues against a conjunct of views that Kant explicitly rejects. Unlike some contemporary virtue ethics, the basic concepts of Kantian moral theory need not have explanatory or predictive value for human actions. Kant insists, “it is absolutely impossible by means of experience to make out with complete certainty a single case in which the maxim of an action ... rested simply on moral grounds” (4:407), and his concept of human evil provides for prediction in terms of *failure* to adhere to moral ideals. That said, Kant does think that virtue concepts have *some* explanatory and predictive value; from one with character, “one knows what to expect, not from his instinct ... but from his will” (7:285). But he is not interested in using virtue or even character to explain the behavior of most people. Moreover, Alfano provides no reason, other than textual support from some virtue ethicists, why we *should* aim for a moral theory that has these features, so giving up this conjunct is no loss for Kant. If egalitarianism *itself* were threatened by Kant’s dodge, then it would cede an important part of Kant’s theory, since he holds that all people ought to – and thus can – have good wills. As we saw in §2.2, however, Kant’s concession that few if any people *have* good character does not imply that few if any people *can* have character, and it is only egalitarianism regarding the *possibility* of virtue that Kant’s (or any sensible) moral theory requires.

Alfano’s second objection is that the dodge simply fails to work for certain cases. Here again, he focuses on uses of the dodge within virtue ethics, which “refer ... to the distinctions among virtue, continence, incontinence, and vice” (Alfano 2013: 63) to explain situation-dependent inconstancy in human behaviors. Alfano objects,

this response constricts the range of situational factors that it can handle. Temptations and situational demand characteristics – what I called bad reasons in the previous chapter – are naturally handled by this response: people do the right thing but not wholeheartedly because they’re tempted, or they do the wrong thing because of situational demand characteristics. But situational non-reasons such as ambient sensibilia and mood effects have nothing to do with

continence and incontinence. Thus, the dodge at best responds to a sub-class of situational influences, and a sub-class that is intuitively less worrisome anyway. (Alfano 2013: 63)

Again, Kant's set of moral categories provides a much better response. Impurity and depravity of will, where one deliberately prioritizes non-moral incentives over non-moral ones, are limited in that they can only explain bad reasons (temptations). Insofar as one is committed both to moral reasons and to some bad reasons for action, one may act irregularly depending upon which is more forceful in a particular situation. But Kant's category of frailty introduces a whole other set of considerations. Here the lower faculty of desire influences the higher not by providing candidate reasons that are taken up deliberately as reasons for action, but through bypassing or manipulating deliberation directly. The fact that the lower faculty of desire responds to "stimuli" rather than reasoned "motives" (28:254) provides a natural way of making sense of situational non-reasons. And frailty as a disposition is ascribable to freedom; it is because one does not sufficiently care about character as such that one is susceptible to these sorts of influences from the lower faculty. Rachana Kamtekar has made a similar point in defense of (ancient) virtue ethics:

It may require a strong interest (in the consequences of deceiving or not, or in the activity of deceiving or not) to lead one to extend one's strategies (for deception or non-deception) across situations . . . [T]he absence of a strong enough interest . . . may help to explain cross-situational inconsistency. (Kamtekar 2003: 269-70)

Given his three-fold account of human evil, Kant could go further, cataloging the different sorts of factors – including some that are not conscious interests at all – that human beings allow to affect their actions even when opposed to the moral law. Situationists tend to emphasize, in particular cases, that "the deeds in question do not require heroic commitment or sacrifice" (Doris 2002: 31). But Kant can rightly point out that the development of the sort of character that would act consistently across situations does require a sort of sacrifice, one that is heroic not because it is

impossible or requires extreme moral fortitude but because the ordinary but categorical commitment to virtue required in order to cultivate and preserve character is so rare among human beings.

Kant's moral theory is committed to an egalitarianism of moral possibility that empirical evidence cannot contravene and to distinguishing moral demands from explanatory and predictive sciences. And his account of human evil includes a conception of frailty that recognizes the moral importance of constancy in the face of sub-rational volitional influences. In these respects, his character-based ethics has resources to dodge situationism more effectively, and with more internal coherence, than contemporary neo-Aristotelian approaches to virtue.

### 3. The Kantian alternative: moral anthropology

Without giving up on the moral importance of character, Kantians can incorporate contemporary situationist research as evidence of humans' radical evil and also – given the mixed nature of that situationist research – as empirical support for moral hope. The framework of human evil and efforts towards looking for grounds of hope can also redirect empirical research programs in psychology to focus on the different ways and degrees to which individuals are susceptible to situational influence. A further component of Kant's response to evil addresses the problem of lack of character more practically. Kant proposes a shift in emphasis from the identification of perfectly good wills (hence perfectly consistent characters) towards the task (both individual and social) of *cultivating* good wills (and hence cultivating character). In this respect, he anticipates Mark Alfano's emphasis on what Alfano calls "moral technology" (see Alfano 2013) and Kant "moral anthropology" (6:217). However, whereas Alfano seeks to use moral technology to *replace* genuine, agential, character-based virtue with more frequent instances of virtuous *behavior*, Kant advocates moral anthropology to strengthen human character itself. For Alfano, empirical research should inform the content of moral theories; insofar as human beings typically do not have character but

can be made to act reliably well, we should focus energies on helping them act reliably well. For Kant, by contrast, moral theory must precede empirical research in the sense that empirical research should focus on figuring out how best to bring about what moral theory shows to be the ideal. As Kant puts it,

moral anthropology ... would deal only with the subjective conditions in human nature that hinder people or help them in fulfilling the laws of a metaphysics of morals. It would deal with the development, spreading, and strengthening of moral principles (in education in schools and in popular instruction) and with other similar teachings and precepts based in experience.

(6:217)

Given the arguments laid out in §1 that show why character is essential to the good will, moral anthropology should focus, among other things, on investigating the empirical conditions for the cultivation of character.

Throughout Kant's anthropology, character cultivation is a central concern. He emphasizes that "[t]he acquisition of good character with people happens through education" (25:1172; cf. 7:294) and offers specific pedagogical recommendations, such as avoiding imitation (5:154; 7:325; 9:487-8) and using uplifting moral stories and examples (5:156; 8:286). While good education is the most important influence on character, Kant also mentions the role that stable and just political regimes, peace, polite society, and progress in the arts and sciences can have on character (8:375).<sup>20</sup> Beyond external influences, he offers practical guidance for how to cultivate character in oneself through basic maxims of self-care: "a) Not to speak an untruth intentionally ... b) not to dissemble ... c) not to break one's legitimate promise ... d) not to join the company of evil-minded people ... e) not to pay attention to slander" (7: 294). All these factors support the development of character as such.

---

<sup>20</sup> For discussion of these and other influences on character development, see Cohen 2009; Frierson 2005, 2014; Moran 2012; Munzel 1999, 2012.

Kant may, of course, have been wrong about these claims. His justification for the detrimental effects of imitation in education or the salutary effects of polite society are not, unlike his moral theory itself, grounded in a priori claims about what is essential to the good will as such. They are empirical claims about what forces strengthen the ability to act on principles in a consistent way. They are based on careful observation, plausible psychological reconstruction, and some armchair speculation, rather than rigorous experimental study. Still, they offer conjectures that could be tested in a Kantian psychological research program, and more importantly, they model the *kinds* of psychological theories that should be part of such a research program, theories not about what is possible for human beings as one finds them, but about what methods are most effective for helping human beings come to have the qualities – such as character – that are morally required.

Kant might even have been wrong about the importance of character for moral virtue. Importantly, if Kant *is* right, the sort of character that is essential to virtue is quite different from the largely unreflective and habitual “character” emphasized in neoAristotelean theories (including, e.g., Miller 2013) and even the move towards “character” in contemporary “positive psychology” (see Peterson and Seligman 2004, especially p. 10; see too Merritt et. al. 2010). Kantian character is grounded in self-reflective, deliberate commitments to general practical principles. The arguments for the importance of character in this sense are briefly mentioned in §1 and are echoed in much recent neoKantian work in moral theory and even in the philosophy of action more broadly (see, e.g., Korsgaard 1996, 2009; Shapiro 1999, 2009; and even Bratman 1987, 2007). But even if Kant can be challenged with respect to his moral theory, this challenge cannot *simply* be based on observations of how human beings actually behave.

Arguably the most important insight of Kant’s approach to moral anthropology is his insistence that psychologists must get clear on what is normatively required *prior to* analyzing actual human behavior with an eye to moral cultivation. If morality is fundamentally about maximizing

good deeds or overall welfare, then situation-driven good deeds may be a step towards an ideal condition. But if morality is fundamentally about a good will that is the only thing good without qualification, and if this good will *essentially* involves character, then increasing the frequency of situation-driven good deeds may well be a step *away* from the moral ideal. And while Kant (and contemporary Kantians) provide serious normative *arguments* for their moral claims, situationists generally simply *assume* a sort of consequentialism according to which maximizing good results or good actions is the end, so that virtues would be at best reliable ways of achieving those goals. Perhaps situationists just think that it's obvious that "the flourishing of people and societies" is the most important good, and that such "flourishing" is relatively straightforward to define (Alfano 2013:203). But Kant's arguments at least shift the burden onto situationists to show why we should take as worth promoting a sort of "flourishing" that does not involve the only unconditional good, the good will.

Consider, in this context, Mark Alfano's recent suggestion that rather than assuming "that the bearers of virtues are individual agents ... [i]t might ... be better to think of a virtue as a triadic relation among an agent, a social milieu, and an environment" (Alfano 2013:106). Such a suggestion prompts the question of what virtue concepts *are* and are *for*. For Kant, the concept of virtue picks out a particular way in which the good will shows up in human agents who must struggle against temptations to violate the moral law. It is empirically-situated, but its normative core is grounded in his account of the good will. Virtue is necessary for human wills to manifest that goodness that is "without qualification." And importantly, this goodness is one that is necessarily a property of individual agents; only agents are the sorts of things that *can* be good without qualification. Moreover, Kant's virtue concept makes sense of many important features of virtue, such as that, as Christian Miller puts it, a virtue "is a ... trait for which a person who possesses it is (at least to some degree) normatively responsible" (2014:11). All of these features of virtues are lost in the triadic

relation view, which diffuses responsibility and separates virtue from the agential goods that essentially define it. There *might* be a reason to change the definition of virtue in this way. Since people are generally interested in promoting virtue, if what *really* matters are, say, instances of conduct conducive to welfare, then perhaps we should change the concept so that what we seek to promote more accurately tracks the promotion of such conduct. But then we need a fuller defense for why welfare or good-conduct-instances are the ultimate goods worth promoting, rather than agential excellence, and the fact that the latter is harder to achieve – or even psychologically impossible to achieve – doesn't show that the former should displace it.

The stakes of getting clear on the relationship between empirical psychology and ethics are high. One of the main ways in which humans' tendency to self-deception works to foster and cultivate radical evil in human nature is precisely through watering down the demands of morality and/or weakening our responsibility for our own actions (and characters). Recent appropriations of situationism lead to precisely those moral theories of "lenient justice" (27:331) that have the effect of making human beings more like what they describe us as being, just as in Kant's own example of those "practical men (for whom morals is mere theory)" who "pretend to see in advance, from the nature of the human being, that *he is never going to will* what is required" and thereby "*make improvement impossible*" (8:371, 373). In the face of situationist excuses for bad behavior, especially conjoined with an increased emphasis on good results rather than good characters, human beings will tend to be even *more* situation-driven, gradually moving from frail or impure wills that are situation-dependent despite endorsing the importance of genuine virtue towards *depraved* wills that settle for mere legality rather than genuine morality, a life of doing good things, but without a good *will*. This point is confirmed in recent studies showing that one of the main factors that makes people susceptible to sub-rational influences is the belief that they lack control or responsibility for their behavior or attributes (e.g. Kohlberg and Candee 1984; Dweck 2000, 2006). In that context, it

is no surprise that Milgram's final exhortation to test subjects – "You have no choice; you must go on" – involves specifically seeking to undermine their sense that they have control over their own decisions.

Ultimately, bearing fruit in improved human lives – lives that are more and more governed by character in Kant's sense – depends upon developing a successful research program that can be applied in education, parenting, social organization, and a host of other social settings where we influence each other's development. Part of this research program involves identifying the precise scope of situation-dependence. This sort of research would resemble Kohlberg's analysis of Milgram's experiment, showing systematic differences in people's susceptibility to influence, or Carol Dweck's studies on the role of "mindset" in reducing or even eliminating the impact of various situational variables, such as her response to Steele and Aronson's famous work on the framing effects of racial labels (Steele and Aronson 1995; cf. Dweck 2006:75-7). By articulating clear even if only approximate criteria for distinguishing between those who seemingly have Kantian character and those who do not, a Kantian research program could specifically investigate whether and to what extent such character is resistant to various situational pressures. Insofar as it is – and Kohlberg, Dweck, and others give reasons to think that it might be – this would show (problematic) situation-dependence to be a contingent and preventable trait, rather than a fixed aspect of human nature.

Another, and even more important, area for study would be the investigation of what can be done – either directly or indirectly – to cultivate Kantian character. So-called "positive psychology" and the character studies associated with it are still sufficiently young subfields within psychology that the development of real Kantian alternatives to the relatively shallow, consequentialist-Aristotelian models currently being employed in positive psychology could have real impact. There is a need for providing better (more Kantian) foci even for existing "results-oriented" and "fact-

based” studies of how to cultivate character. Given widespread public appeal to “psychological facts” to justify practical programs from social policy to personal child-rearing and self-improvement, there is a demand for psychological research that can improve human lives. But in the absence of good philosophical frameworks, psychologists and thereby the general public will base their lives and social programs on bad moral philosophy. Kantians, in cooperation with sympathetic psychologists, need to offer better psychology, backed by better philosophy.<sup>21</sup>

For Kant, avoiding the *misuse* of empirical psychology and opening the way to morally healthy uses of psychology require clarifying the right relationship between moral philosophy and empirical psychology and getting each of those elements of the theory correct. As an empirically-rooted discipline investigating the conditions for strengthening moral principles, Kantian moral anthropology essentially depends upon a *prior* understanding of what the relevant moral principles are. But to properly diagnose the problem of human evil and develop appropriate means for cultivating virtue in oneself and others, empirical research can be invaluable.<sup>22</sup>

## Bibliography

- Alfano, M. (2013) *Character as Moral Fiction*. Cambridge University Press.
- Allais, L. (2015) *Manifest Reality: Kant's Idealism and his Realism*. Oxford University Press.
- Allison, H. (1990) *Kant's Theory of Freedom*. Cambridge University Press.
- Allison, H. (2001) “Ethics, Evil, and Anthropology in Kant: Remarks on Allen Wood's Kants Ethical Thought,” *Ethics* 111(3):594-613.
- Ameriks, Karl, 1982, “Recent Work on Kant's Theoretical Philosophy,” in: *American Philosophical Quarterly*, 19, pp. 1-24.
- Anderson-Gold, S. (2001) *Unnecessary Evil*. SUNY Press.
- Annas, J. (2005) “Comments on John Doris's *Lack of Character*,” *Philosophy and Phenomenological Research* 71:636-42.

---

21 For further discussion of what such a Kantian research program in positive psychology might involve, including serious obstacles it would face, see Frierson (forthcoming).

22 Many thanks go to Karl Ameriks, Pauline Kleingeld, Katharina Kraus, Katheryn Frierson, and the two anonymous reviewers for *Archiv*. Their encouragement to work on this project, substantive criticisms, and help bringing it to fruition were invaluable. I also thank the participants in the Maxims and MRI's conference in Toronto in 2014 and the Kant and the Sciences conference in Dortmund for helpful comments on earlier drafts of the paper.

- Badhwar, N. (2009) "The Milgram Experiments, Learned Helplessness, and Character Traits." *Journal of Ethics* 13:257-89.
- Bratman, M. (1987) *Intention, Plans, and Practical Reason*, Harvard University Press.
- Bratman, M. (2007) *Structures of Agency: Essays*, Oxford University Press.
- Casey, W. M. and Burton R. V. (1982) "Training Children to Be Consistently Honest through Verbal Self-Instructions" *Child Development* 53:911-19.
- Cohen, A. (2009) *Kant and the Human Sciences: Biology, Anthropology, and History*, Houndmills, UK: Palgrave Macmillan .
- Darley, J. M. and Batson, C. D. (1973) "From Jerusalem to Jericho: A study of situational and dispositional variables in helping behavior," *Journal of Personality and Social Psychology* 27:100-108.
- Doris, J. (2002) *Lack of Character: Personality and Moral Behavior*. Cambridge University Press.
- Doris, J. and Stich, S. (2005) "As a Matter of Fact: Empirical perspectives on ethics," in F. Jackson & M. Smith (eds.), *The Oxford Handbook of Contemporary Philosophy*. Oxford University Press.
- Dweck, C. (2000) *Self-Theories: Their role in motivation, personality, and development*. New York: Psychology Press.
- Dweck, C. (2006) *Mindset: The new psychology of success*. New York: Ballantine Books.
- Frierson, P. (forthcoming) "Towards a Research Program in Kantian Positive Psychology," *Studies in the History and Philosophy of Science*.
- Frierson, P. (2014) *Kant's Empirical Psychology*. Cambridge University Press.
- Frierson, P. (2013) *Kant's Questions: What is the Human Being?* Routledge.
- Frierson, P. (2010a) "Two Standpoints and the Problem of Moral Anthropology," in *Kant's Moral Metaphysics* (ed. James Krueger and Benjamin Lipscomb), Berlin: Walter deGruyter Press, 2010, pp. 83-110.
- Frierson, P. (2010b) "Kantian Moral Pessimism" in *Kant's Anatomy of Evil*, ed. Sharon Anderson-Gold and Pablo Muchnik, Cambridge University Press, pp. 33-56.
- Frierson, P. (2007) "Providence and Divine Mercy in Kant's Ethical Cosmopolitanism," *Faith and Philosophy* 24:143-63.
- Frierson, P. (2006) "Character and Evil in Kant's Moral Anthropology," *Journal of the History of Philosophy* 44: 623-34.
- Frierson, P. (2005) "The Moral Importance of Politeness in Kant's Anthropology," *Kantian Review* 9: 105-27
- Frierson, P. (2003) *Freedom and Anthropology in Kant's Moral Philosophy*. Cambridge University Press.
- Frimer, J., et. al. (2011) "The Integration of Agency and Communion in Moral Personality: Evidence of Enlightened Self-Interest," *Journal of Personality and Social Psychology* 101:149-63.
- Grenberg, J. (2005) *Kant and the Ethics of Humility*. Cambridge University Press.
- Gollwitzer, P. and Bargh, J., Eds. (1996) *The Psychology of Action: Linking cognition and motivation to behavior*. New York:

Guilford

- Haney, C. et. al. (1973) "Interpersonal Dynamics in a Simulated Prison," *International Journal of Criminology and Penology* 1:69-97.
- Harman, G. (2000) "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error" in *Explaining Value and other essays in moral philosophy*, Oxford University Press.
- Isen, A. and Levin, P. (1972) "Effect of Feeling Good on Helping: Cookies and Kindness," *Journal of Personality and Social Psychology*, 21(3):384-8.
- Kamtekar, R. (2003) "Situationism and Virtue Ethics on the Content of Our Character," *Ethics* 114: 458-91.
- Kleingeld, P. (2014) "Debunking Confabulation: Emotions and the Significance of Empirical Psychology for Kantian Ethics" in *Kant on Emotions and Value* (ed. A. Cohen) Basingstoke: Palgrave MacMillan.
- Kleingeld, P. (2015) "Consistent Egoists and situation managers: two problems for situationism," *Philosophical Explorations*, preprint at <http://dx.doi.org/10.1080/13869795.2015.1038293>.
- Kohlberg, L. (1981) *The Philosophy of Moral Development*. San Francisco: Harper and Row.
- Kohlberg, L. and Candee, D. (1984) "The Relationship of Moral Judgment to Moral Action," in *Morality, Moral Behavior, and Moral Development* (ed. W. M. Kurtines and J. L. Gewirtz) New York: Wiley-Interscience.
- Korsgaard, C. (2009) *Self-Constitution: Agency, Identity, and Integrity*, Oxford: Oxford University Press.
- Korsgaard, C. (1996) *The Sources of Normativity*, Cambridge: Cambridge University Press.
- Kilham, W., & Mann, L. (1974) "Level of destructive obedience as a function of transmitter and executant roles in the Milgram obedience paradigm," *Journal of Personality and Social Psychology*, 29(5):692-702.
- Kristjánsson, K. (2008) "An Aristotelian Critique of Situationism," *Philosophy* 83:55-76.
- Kupperman, Joel J. 2001. "The Indispensability of Character," *Philosophy* 76: 239–250.
- Mariña, J. (1997) "Kant on Grace: A Reply to his Critics." *Religious Studies* 33: 379-400.
- Mazar, N., O. Amir, and D. Ariely (2008) "The Dishonesty of Honest People: A theory of self-concept maintenance," *Journal of Marketing Research* 45:633-44.
- McCarty, R. (2009) *Kant's Theory of Action*. Oxford University Press.
- Merritt, M. and J. Doris and G. Harman (2010) "Character," in *Moral Psychology Handbook*, ed. J. Doris and the Moral Psychology Research Group, Oxford University Press, pp. 355-401.
- Michaelson Jr, G. E. (1990) *Fallen Freedom: Kant on Radical Evil and Moral Regeneration*. Cambridge University Press.
- Milgram, S. (1963) "Behavioral Study of Obedience," *Journal of Abnormal and Social Psychology* 67:371-8.

- Milgram, S. (1974) *Obedience to Authority: An experimental view*. New York: Harper and Row.
- Miller, C. (2013) *Moral Character: An empirical theory*. Oxford University Press.
- Miller, C. (2014) *Character and Moral Psychology*. Oxford University Press.
- Moran, K. (2012) *Community and Progress in Kant's Moral Philosophy*. Washington, D.C.: The Catholic University of America Press.
- Munzel, G.F. (2012) *Kant's Conception of Pedagogy: Toward Education for Freedom*. Evanston, IL: Northwestern University Press.
- Munzel, G. Felicitas (1999) *Kant's Conception of Moral Character: The "Critical" Link of Morality, Anthropology, and Reflective Judgment*. Chicago, The University of Chicago Press.
- Nelkin, D. (2000) "Two Standpoints and the Belief in Freedom," *The Journal of Philosophy* 157:564-76.
- Nisbett, R.E. and Wilson, T.D. (1977) "Telling more than we can know: Verbal reports on mental processes" in *Psychological Review* 84:231-259.
- O'Neill, O. (1989) *Constructions of Reason: Explorations of Kant's Practical Philosophy*. Cambridge University Press.
- Pasternak, L. (2013) *Routledge Philosophy Guidebook to Kant on Religion within the Boundaries of Mere Reason*, London: Routledge.
- Peterson, C. (2006) *A Primer in Positive Psychology*, Oxford University Press.
- Roberts. R. and Wood, W. J. (2007) *Intellectual Virtues: An essay in regulative epistemology*. Oxford University Press.
- Sabini, J. and M. Silver (2005) "Lack of Character? Situationism Critiqued" *Ethics* 115:535-62.
- Sarkissian, H. (2010) "Minor Tweaks, Major Payoffs: The Problems and Promise of Situationism in Moral Philosophy," *Philosopher's Imprint* 10(9):1-15.
- Shu, L., F. Gino, and M. Bazerman (2011) "Dishonest Deed, Clear Conscience: When cheating leads to moral disengagement and motivated forgetting," *Personality and Social Psychology Bulletin* 37:330-49.
- Steele, C. and Aronson, J. (1995) "Stereotype Threat and Intellectual Test Performance of African-Americans," *Journal of Personality and Social Psychology* 68:797-811.
- Swanton, C. (2003) *Virtue Ethics: A Pluralistic View*. Oxford University Press.
- Upton, C. L. (2009a). "Virtue Ethics and Moral Psychology: The Situationism Debate," *Journal of Ethics* 13:103-115.
- Vranas, P. (2005) "The Indeterminacy Paradox: Character Evaluations and Human Psychology," *Noûs* 39:1-42.
- Watkins, E. (2006) *Kant and the Metaphysics of Causality*, Cambridge: Cambridge University Press.
- Webber, J. (2006) "Virtue, Character and Situation," *Journal of Moral Philosophy* 3:193-213.
- Wilson, T. (1998) *Strangers to Ourselves: Discovering the adaptive unconscious*. Harvard University Press.

Wood, A. (1999) *Kant's Ethical Thought*. Cambridge University Press.

Zimbardo, P. (2007) *The Lucifer Effect*, New York: Random House.