CHAPTER **5**

Can People Be Virtuous?

Abstract

In his contribution, Mark Alfano lays out a new (to virtue theory) naturalistic way of determining what the virtues are, what it would take for them to be realized, and what it would take for them to be at least possible. This method is derived in large part from David Lewis's development of Frank Ramsey's method of implicit definition. The basic idea is to define a set of terms not individually but in tandem. This is accomplished by assembling all and only the common sense platitudes that involve them (e.g., typically, people want to be virtuous), conjoining those platitudes, and replacing the terms in question by existentially quantified variables. If the resulting sentence is satisfied, then whatever satisfies *are* the virtues. If it isn't satisfied, there are a couple of options. First, one could just admit defeat by saying that people can't be virtuous. More plausibly, one could weaken the conjunction by dropping a small number of the platitudes from it (and potentially adding some others). Alfano suggests that the most attractive way to do this is by dropping the platitudes that deal with cross-situational consistency and replacing them with platitudes that involve social construction: basically, people are virtuous (when they are) at least in part because other people signal their expectations of virtuous conduct, which induces virtuous conduct, which in turn induces further signals of expected virtuous conduct, and so on. In his response, James Montmarquet does not reject Alfano's proposals regarding Ramsification as an analytical device, but does question whether Alfano's own conception of moral character traits—whatever its empirical adequacy from the standpoint of social

science—does justice to our conception of moral responsibility. Because assignments of moral responsibility are so important, he suggests, we would like them and our closely related attributions of moral character to be clear, exact, and amenable to scientific treatment. From both the Tolstoyan and Humean perspectives he distinguishes, however, these assignments and attributions are neither exact nor scientific, but more like an art form.

Ramsifying Virtue Theory

Mark Alfano

Can people be virtuous? This is a hard question, both because of its form and because of its content.

In terms of content, the proposition in question is at once normative and descriptive. Virtue-terms have empirical content. Attributions of virtues figure in the description, prediction, explanation, and control of behavior. If you know that someone is temperate, you can predict with some confidence that he won't go on a bender this weekend. Someone's investigating a mysterious phenomenon can be partly explained by (correctly) attributing curiosity to her. Character witnesses are called in trials to help determine how severely a convicted defendant will be punished. Virtue-terms also have normative content. Attributions of virtues are a manifestation of high regard and admiration; they are intrinsically rewarding to their targets; they're a form of praise. The semantics of purely normative terms is hard enough on its own; the semantics of "thick" terms that have both normative and descriptive content is especially difficult.

Formally, the proposition in question ("people are virtuous") is a generic, which adds a further wrinkle to its evaluation. It is notoriously difficult to give truth conditions for generics (Leslie 2008). A generic entails its existentially quantified counterpart, but is not entailed by it. For instance, tigers are four-legged, so some tigers are four-legged; but even though some deformed tigers are three-legged, it doesn't follow that tigers are three-legged. A generic typically is entailed by its universally quantified counterpart, but does not entail it. Furthermore, a generic neither entails nor is entailed by its counterpart "most" statement. Tigers give live birth, but most tigers do not give live birth; after all, only about half of all tigers are female, and not all of them give birth. Most mosquitoes do not carry West Nile virus, but mosquitoes carry West Nile virus. Given the trickiness of generics, it's helpful to clarify them to the extent possible with more precise non-generic statements.

Moreover, the proposition in question is modally qualified, which redoubles the difficulty of confirming or disconfirming it. What's being asked is not simply whether people *are* virtuous, but whether they *can be* virtuous. It could

turn out that even though no one is virtuous, it's possible for people to become virtuous. This would, however, be extremely surprising. Unlike other unrealized possibilities, virtue is almost universally sought after, so if it isn't widely actualized despite all that seeking, we have fairly strong evidence that it's not there to be had.

In this chapter, I propose a method for adjudicating the question whether people can be virtuous. This method, if sound, would help to resolve what's come to be known as the situationist challenge to virtue theory, which over the last few decades has threatened both virtue ethics (Alfano 2011, 2013a; Doris 2002; Harman 1999) and virtue epistemology (Alfano 2012 2013a, forthcoming b; Olin & Doris 2014). The method is an application of David Lewis's (1966, 1970, 1972) development of Frank Ramsey's (1931) approach to the implicit definition of theoretical terms. The method needs to be tweaked in various ways to handle the difficulties canvassed above, but, when it is, an interesting answer to our question emerges: we face a theoretical tradeoff between, on the one hand, insisting that virtue is a robust property of an individual agent that's rarely attained and perhaps even unattainable and, on the other hand, allowing that one person's virtue might inhere partly in other people, making virtue at once more easily attained and more fragile.

The basic principle underlying the Ramsey-Lewis approach to implicit definition (often referred to as 'Ramsification') can be illustrated with a well-known story:

> And the Lord sent Nathan unto David. And he came unto him, and said unto him, "There were two men in one city; the one rich, and the other poor. The rich man had exceeding many flocks and herds: But the poor man had nothing, save one little ewe lamb, which he had bought and nourished up: and it grew up together with him, and with his children; it did eat of his own meat, and drank of his own cup, and lay in his bosom, and was unto him as a daughter. And there came a traveler unto the rich man, and he spared to take of his own flock and of his own herd, to dress for the wayfaring man that was come unto him; but took the poor man's lamb, and dressed it for the man that was come to him." And David's anger was greatly kindled against the man; and he said to Nathan, "As the Lord liveth, the man that hath done this thing shall surely die: And he shall restore the lamb fourfold, because he did this thing, and because he had no pity." And Nathan said to David, "Thou art the man."

Nathan uses Ramsification to drive home a point. He tells a story about an ordered triple of objects (two people and an animal) that are interrelated in various ways. Some of the first object's properties (e.g., wealth) are monadic; some of the second object's properties (e.g., poverty) are monadic; some of the first object's properties are relational (e.g., he steals the third object from the

second object); some of the second object's properties are relational (e.g., the third object is stolen from him by the first object); and so on. Even though the first object is not explicitly defined as *the X such that . . .* , it is nevertheless implicitly defined as *the first element of the ordered triple such that . . .* The big reveal happens when Nathan announces that the first element of the ordered triple, about whom his interlocutor has already made some pretty serious pronouncements, is the very person he's addressing (the other two, for those unfamiliar with the 2nd *Samuel* 12, are Uriah and Bathsheba[1]).

The story is Biblical, but the method is modern. To implicitly define a set of theoretical terms (henceforth 'T-terms'), one formulates a theory T in those terms and any other terms (henceforth 'O-terms') one already understands or has an independent theory of. Next, one writes T as a single sentence, such as a long conjunction, in which the T-terms $t_1 . . . , t_n$ occur (henceforth '$T[t_1 . . . , t_n]$' or 'the postulate of T'). The T-terms are replaced by unbound variables $x_1 . . . , x_n$, and then existentially quantified over to generate the Ramsey sentence of T, $\exists x_1 . . . , x_n T[x_1, . . . , x_n]$, which states that T is realized, that is, that there are objects $x_1 . . . , x_n$ that satisfy the Ramsey sentence. An ordered n-tuple that satisfies the Ramsey sentence is then said to be a *realizer* of the theory.

Lewis (1966) famously applied this method to folk psychology to argue for the mind-brain identity theory. Somewhat roughly, he argued that folk psychology can be treated as a theory in which mental-state terms are the T-terms. The postulate of folk psychology is identified as the conjunction of all folk-psychological platitudes (commonsense psychological truths that everyone knows, and everyone knows that everyone knows, and everyone knows that everyone knows that everyone knows, and so on). The Ramsey sentence of folk psychology is formed in the usual way, by replacing all mental-state terms (e.g., 'belief,' 'desire,' 'pain,' etc.) with variables and existentially quantifying over those variables. Finally, one goes on to determine what, in the actual world, satisfies the Ramsey sentence; that is, one investigates what, if anything, is a realizer of the Ramsey sentence. If there is a realizer, then that's what the T-terms refer to; if there is no realizer, then the T-terms do not refer. Lewis claims that brain states are such realizers, and hence that mental states are identical with brain states.

Lewis's Ramsification method is attractive for a number of reasons.[2] First, it ensures that we don't simply change the topic when we try to give a philosophical account of some phenomenon. If your account of the mind is wildly inconsistent with the postulate of folk psychology, then—though you may be giving an account of something interesting—you're not doing what you think you're doing. Second, enables us to distinguish between the *meaning* of the T-terms and *whether they refer*. The T-terms *mean* what they *would* refer to, if there were such a thing. Whether they in fact refer is a distinct question. Third, and perhaps most importantly, Ramsification is holistic. The first half of the twentieth century bore witness to the fact that it's impossible to give an

independent account of almost any psychological phenomenon (belief, desire, emotion, perception) because what it means to have one belief is essentially bound up with what it means to have a whole host of other beliefs, as well as (at least potentially) a whole host of desires, emotions, and perceptions. Ramsification gets around this problem by giving an account of *all* of the relevant phenomena *at once*, rather than trying to chip away at them piecemeal.

Virtue theory stands to benefit from the application of Ramsification for all of these reasons. We want an account of virtue, not an account of some other interesting phenomenon (though we might want that too). We want an account that recognizes that talk of virtue is meaningful, even if there aren't virtues. Most importantly, we want an account of virtue that recognizes the complexity of virtue and character—the fact that virtues are interrelated in a whole host of ways with occurrent and dispositional mental states, with other virtues, with character more broadly, and so on.

Whether Lewis is right about brains is irrelevant to our question, but his methodology is crucial. What I want to do now is to show how the same method, suitably modified, can be used to implicitly define virtue-terms, which in turn will help us to answer the question whether people can be virtuous. For reasons that will become clear as we proceed, the T-terms of virtue theory as I construe it here are 'person,' 'virtue,' 'vice,' the names of the various virtues (e.g., 'courage,' 'generosity,' 'curiosity'), the names of their congruent affects (e.g., 'feeling courageous,' 'feeling generous,' 'feeling curious'), the names of the various vices (e.g., 'cowardice,' 'greed,' 'intellectual laziness'), and the names of their congruent affects, (e.g., 'feeling cowardly,' 'feeling greedy,' 'feeling intellectually lazy'). The O-terms are all other terms, importantly including terms that refer to attitudes (e.g., 'belief,' 'desire,' 'anger,' 'resentment,' 'disgust,' 'contempt,' 'respect'), mental processes (e.g., 'deliberation'), perceptions and perceptual sensitivities, behaviors, reasons, situational features (e.g., 'being alone,' 'being in a crowd,' 'being monitored'), and evaluations (e.g., 'praise' and 'blame').

Elsewhere (Alfano 2013a), I have argued for an intuitive distinction between high-fidelity and low-fidelity virtues. High-fidelity virtues, such as honesty, chastity, and loyalty, require near-perfect manifestation in undisrupted conditions. Someone only counts as chaste if he never cheats on his partner when cheating is a temptation. Low-fidelity virtues, such as generosity, tact, and tenacity, are not so demanding. Someone might count as generous if she were more disposed to give than not to give when there was sufficient reason to do so; someone might count as tenacious if she were more disposed to persist than not to persist in the face of adversity. If this is on the right track, the postulate of virtue theory will recognize the distinction. For instance, it seems to me at least that almost everyone would say that helpfulness is a low-fidelity virtue whereas loyalty is a high-fidelity virtue. Here, then, are some families of platitudes about character that are candidates for the postulate of virtue theory:

(A) *The Virtue/Affect Family*

    (a_1) If a person has courage, then she will typically feel courageous when there is sufficient reason to do so.

    (a_2) If a person has generosity, then she will typically feel generous when there is sufficient reason to do so.

    (a_3) If a person has curiosity, then she will typically feel curious when there is sufficient reason to do so.

    $\vdots$

    (a_n) . . . .

(C) *The Virtue/Cognition Family*

    (c_1) If a person has courage, then she will typically want to overcome threats.

    (c_2) If a person has courage, then she will typically deliberate well about how to overcome threats and reliably form beliefs about how to do so.

    $\vdots$

    (c_n) . . . .

(S) *The Virtue/Situation Family*[3]

    (s_1) If a person has courage, then she will typically be unaffected by situational factors that are neither reasons for nor reasons against overcoming a threat.

    (s_2) If a person has generosity, then she will typically be unaffected by situational factors that are neither reasons for nor reasons against giving resources to someone.

    (s_3) If a person has curiosity, then she will typically be unaffected by situational factors that are neither reasons for nor reasons against investigating a problem.

    $\vdots$

    (s_n)

(E) *The Virtue/Evaluation Family*

    (e_1) If a person has courage, then she will typically react to threats in ways that merit praise.

    (e_2) If a person has generosity, then she will typically react to others' needs and wants in ways that merit praise.

    (e_3) If a person has curiosity, then she will typically react to intellectual problems in ways that merit praise.

    $\vdots$

    (e_n)

(B) *The Virtue/Behavior Family*

    (b_1) If a person has courage, then she will typically act so as to overcome threats when there is sufficient reason to do so.

(b_2) If a person has generosity, then she will typically act so as to benefit another person when there is sufficient reason to do so.

(b_3) If a person has curiosity, then she will typically act so as to solve intellectual problems when there is sufficient reason to do so.

$\vdots$

(b_n)

(P) *The Virtue Prevalence Family*

(p_1) Many people commit acts of courage.

(p_2) Many people commit acts of generosity.

(p_3) Many people commit acts of curiosity.

(p_4) Many people are courageous.

(p_5) Many people are generous.

(p_6) Many people are curious.

$\vdots$

(p_n)

(I) *The Cardinality/Integration Family*

(i_1) Typically, a person who has modesty also has humility.

(i_2) Typically, a person who has magnanimity also has generosity.

(i_3) Typically, a person who has curiosity also has open-mindedness.

$\vdots$

(i_n)

(D) *The Desire/Virtue Family*

(d_1) Typically, a person desires to have courage.

(d_2) Typically, a person desires to have generosity.

(d_3) Typically, a person desires to have curiosity.

$\vdots$

(d_n)

(F) *The Fidelity Family*

(f_1) Chastity is high-fidelity.

(f_2) Honesty is high-fidelity.

(f_3) Creativity is low-fidelity.

$\vdots$

(d_n)

Each platitude in each family is meant to be merely illustrative. Presumably they could all be improved somewhat, and there are many more such platitudes. Moreover, each family is itself just an example. There are many further families describing the relations among vice, affect, cognition, situation, evaluation, and behavior, as well as families that make three-way rather than two-way connections (e.g., "If a person is courageous, then she will typically act so as to overcome threats when there is sufficient reason to do so and

because she feels courageous."). For the sake of simplicity, though, let's assume that the families identified above contain all and only the platitudes relevant to the implicit definition of virtues. Ramsification can now be performed in the usual way. First, create a big conjunction

$$a_1 \wedge a_2 \wedge ... \wedge a_n \wedge$$
$$c_1 \wedge c_2 \wedge ... \wedge c_n \wedge$$
$$s_1 \wedge s_2 \wedge ... \wedge s_n \wedge$$
$$e_1 \wedge e_2 \wedge ... \wedge e_n \wedge$$
$$b_1 \wedge b_2 \wedge ... \wedge b_n \wedge$$
$$p_1 \wedge p_2 \wedge ... \wedge p_n \wedge$$
$$i_1 \wedge i_2 \wedge ... \wedge i_n \wedge$$
$$d_1 \wedge d_2 \wedge ... \wedge d_n \wedge$$
$$f_1 \wedge f_2 \wedge ... \wedge f_n$$

(henceforth, simply the 'postulate of virtue theory'). Next, replace each of the T-terms in the postulate of virtue theory with an unbound variable, then existentially quantifies over those variables to generate the Ramsey sentence of virtue theory. Finally, check whether the Ramsey sentence of virtue theory is true and—if it is—what its realizers are.

After this preliminary work has been done, we're in a position to see more clearly the problem raised by the situationist challenge to virtue theory. Situationists argue that there is no realizer of the Ramsey sentence of virtue theory. Moreover, this is not for lack of effort. Indeed, one family of platitudes in the Ramsey sentence specifically states that, typically, people desire to be virtuous; it's not as if no one has yet tried to be or become courageous, generous, or curious.[4] In this chapter, I don't have space to canvass the relevant empirical evidence; interested readers should see my (2013a and 2013b). Nevertheless, the crucial claim—that the Ramsey sentence of virtue theory is not realized—is not an object of serious dispute in the philosophical literature.

One very common response to the situationist challenge from defenders of virtue theory (and virtue ethics in particular) is to claim that virtues are actually quite rare, directly contradicting the statements in the virtue prevalence family. I do not think this is the best response to the problem, as I explain below, but the point remains that all serious disputants agree that the Ramsey sentence is not realized.

As described above, Ramsification looks like a simple, formal exercise. Collect the platitudes, put them into a big conjunction, perform the appropriate substitutions, existentially quantify, and check the truth-value of the resulting Ramsey sentence (and the referents of its bound variables, if any). But there are several opportunities for a critic to object as the exercise unfolds.

One difficulty that arises for some families, such as the desire/virtue family, is that they involve T-terms within the scope of intentional attitude verbs.[5] Since existential quantification into such contexts is blocked by opacity, such families cannot be relied on to define the T-terms, though they can be used to double-check the validity of the implicit definition once the T-terms are defined.[6]

Another difficulty is that this methodology presupposes that we have an adequate understanding of the O-terms, which in this case include terms that refer to attitudes, mental processes, perceptions and perceptual sensitivities, behaviors, reasons, situational features, and evaluations. One might be dubious about this presupposition. I certainly am. However, the fact that philosophy of mind and metaethics are works-in-progress should not be interpreted as a problem specifically for my approach to virtue theory. Any normative theory that relies on other branches of philosophy to figure out what mental states and processes are, and what reasons are, can be criticized in the same way.

A third worry is that the list of platitudes contains gaps (e.g., a virtue acquisition family about how various traits are acquired). Conversely, one might think that it has gluts (e.g., unmotivated commitment to virtue prevalence). To overcome this pair of worries, we need a way of determining what the platitudes are. Perhaps surprisingly, there is no precedent for this in the philosophy of mind, despite the fact that Ramsification is often invoked as a framework there.[7] This may be because it's supposed to be obvious what the platitudes are. Here's Frank Jackson's flippant response to the worry:

> I am sometimes asked—in a tone that suggests that the question is a major objection—why, if conceptual analysis is concerned to elucidate what governs our classificatory practice, don't I advocate doing serious opinion polls on people's responses to various cases? My answer is that I do—when it is necessary. Everyone who presents the Gettier cases to a class of students is doing their own bit of fieldwork, and we all know the answer they get in the vast majority of cases. (1998, pp. 36–37)

After all, according to Lewis, everyone knows the platitudes, and everyone knows that everyone knows them, and everyone knows that everyone knows that everyone knows them, and so on. Sometimes, however, the most obvious things are the hardest to spot. It thus behooves us to at least sketch a method for carrying out the first step of Ramsification: identifying the platitudes. Call this *pre-Ramsification*.

Here's an attempt at spelling out how pre-Ramsification should work: start by listing off a large number of candidate platitudes. These can be all of the statements one would, in a less-responsible, Jacksonian mood, have merely asserted were platitudes. It can also include statements that seem highly likely

but perhaps not quite platitudes. Add to the pool of statements some that seem, intuitively, to be controversial, as well as some that seem obviously false; these serve as anchors in the ensuing investigation. Next, collect people's responses to these statements. Several sorts of responses would be useful, including subjective agreement, social agreement, and reaction time. For instance, prompt people with the statement, "Many people are honest," and ask to what extent they agree and to what extent they think others would agree. Measure their reaction times as they answer both questions. High subjective and social agreement, paired with fast reaction times, is strong but defeasible evidence that a statement is a platitude. This is a bit vague, since I haven't specified what counts as "high" agreement or "fast" reaction times, but there are precedents in psychology for setting these thresholds. Moreover, this kind of pre-Ramsification wouldn't establish dispositively what the platitudes are, but then, dispositive proof only happens in mathematics.

It's far beyond the scope of this short chapter to show that pre-Ramsification works in the way I suggest, or that it verifies all and only the families identified above. For now, let's suppose that it does, that is, that all of the families proposed above were validated by pre-Ramsification. Let's also suppose that we have strong evidence that the Ramsey sentence of virtue theory is not realized (a point that, as I mentioned above, is not seriously contested). How should we then proceed?

Lewis foresaw that, in some cases, the Ramsey sentence for a given field would be unrealized, so he built in a way of fudging things: instead of generating the postulate by taking the conjunction of all of the platitudes, one can generate a weaker postulate by taking the disjunction of each of the conjunctions of *most* of the platitudes. For example, if there were only five platitudes, *p, q, r, s,* and *t,* then instead of the postulate's being $p \wedge q \wedge r \wedge s \wedge t$, it would be $(p \wedge q \wedge r \wedge s) \vee (p \wedge q \wedge r \wedge t) \vee \ldots \vee (q \wedge r \wedge s \wedge t)$. In the case of virtue theory, we could take the disjunction of each of the conjunctions of all but one of the families of platitudes. Alternatively, we could exclude a few of the platitudes from within each family.

Fudging in this way makes it easier for the Ramsey sentence to be realized, since the disjunction of conjunctions of most of the platitudes is logically weaker than the straightforward conjunction of all of them. Fudging may end up making it *too* easy, though, such that there are *multiple* realizers of the Ramsey sentence. When this happens, it's up to the theorist to figure out how to strengthen things back up in such a way that there is a unique realizer.

The various responses to the situationist challenge can be seen as different ways of doing this. Everyone recognizes that the un-fudged Ramsey sentence of virtue theory is unrealized. But a sufficiently fudged Ramsey sentence is bound to be multiply realized. It's a theoretical choice exactly how to play things at this point. More traditional virtue theorists such as Joel Kupperman

(2009) favor a fudged version of the Ramsey sentence wherein the virtue prevalence family has been dropped. John Doris (2002) favors a fudged version wherein the virtue/situation and virtue/integration families have been dropped. I (2013a, 2014a) favor a fudged version wherein the virtue/situation family has been dropped and a virtue/social construction family has been added in its place. The statements in the latter family have to do with the ways in which (signals of) social expectations implicitly and explicitly influence behavior. The main idea is that having a virtue is more like having a title or social role (e.g., you're curious because people signal to you their expectations of curiosity) than like having a basic physical or biological property (e.g., being over six feet tall). Christian Miller (2013, 2014) drops the virtue prevalence family and adds a mixed-trait prevalence family in its place, which states that many people possess traits that are neither virtues nor vices, such as the disposition to help others in order to improve one's mood or avoid sliding into a bad mood.

In this short chapter, I don't have the space to argue against all alternatives to my own proposal. Instead, I want to make two main claims. First, the "virtue is rare" dodge advocated by Kupperman and others who drop the virtue prevalence family has costs associated with it. Second, those costs may be steeper than the costs associated with my own way of responding to the situationist challenge.

Researchers in personality and social psychology have documented for decades the tendency of just about everybody to make spontaneous trait inferences, attributing robust character traits on the basis of scant evidence (Ross 1977; Uleman et al. 1996). This indicates that people think that character traits (virtues, vices, and neutral traits, such as extroversion) are prevalent. Zagzebski (1996) concurs, making the somewhat wide-eyed claim that "many of us have known persons whose goodness shines forth from the depths of their being" (p. 83). Furthermore, in a forthcoming paper (Alfano et al., forthcoming), I show that the vast majority of obituaries attribute multiple virtues to the deceased. Not everyone is eulogized in an obituary, of course, but most are (about 55% of Americans, by my calculations). Not all obituaries are sincere, but presumably many are. Absent reason to think that people about whom obituaries differ greatly from people about whom they are not written, we can treat this as evidence that most people think that the people they know have multiple virtues. But of course, if most relations of most people are virtuous, it follows that most people are virtuous. In other words, the virtue-prevalence family is deeply ingrained in folk psychology and folk morality.

Social psychologists think that people are quick to attribute virtues. My own work on obituaries suggests the same. What do philosophers say? Though there are some (Russell 2009) who claim that virtue is rare or even non-existent

with a shrug, this is not the predominant opinion. Alasdair MacIntyre (1984) claims that "without allusion to the place that justice and injustice, courage and cowardice play in human life very little will be genuinely explicable" (p. 199). Philippa Foot (2001), following Peter Geach (1977), argues that certain generic statements characterize the human form of life, and that from these generic statements we can infer what humans need and hence will typically have. For the sake of comparison, consider what she says about a different life form, the deer. Foot first points out that the deer's form of defense is flight. Next, she claims that a certain normative statement follows, namely, that deer are naturally or by nature swift. This is not to say that every deer is swift; some are slow. Instead, it's a generic statement that characterizes the nature of the deer. Finally, she says that any deer that fails to be swift—that fails to live up to its nature—is "so far forth defective" (p. 34). The same line of reasoning that she here applies to non-human animals is meant to apply to human animals as well. As she puts it,

> Men and women need to be industrious and tenacious of purpose not only so as to be able to house, clothe, and feed themselves, but also to pursue human ends having to do with love and friendship. They need the ability to form family ties, friendships, and special relations with neighbors. They also need codes of conduct. *And how could they have all these things without virtues such as loyalty, fairness, kindness, and in certain circumstances obedience?* (pp. 44–45, emphasis mine)

In light of these sorts of claims, let's consider again the defense offered by some virtue ethicists that virtue is rare, or even impossible to achieve. If virtues are what humans *need*, but the vast majority of people don't have them, one would have thought that our species would have died out long ago. Consider the analogous claim for deer: although deer *need* to be swift, the vast majority of deer are galumphers. Were that the case, presumably they'd be hunted down and devoured like a bunch of tasty venison treats. Or consider another example of Foot's: she agrees with Geach (1977) that people need virtues like honeybees need stingers. Does it make sense for someone with this attitude to say that most people lack virtues? That would be like saying that, even though bees need stingers, most lack stingers. It's certainly odd to claim that the majority—even the vast majority of a species fails to fulfill its own nature. That's not a contradiction, but it is a cost to be borne by anyone who responds to the situationist challenge by dropping the virtue prevalence family.

One might respond on Foot's behalf that human animals are special: unlike the other species, we have natures that are typically unfulfilled. That would be an interesting claim to make, but I am not aware of anyone who has defended

it in print.[8] I conclude, then, that dropping the virtue prevalence family is a significant cost to revising the postulate.

But is it a *more significant* cost than the one imposed on me by replacing the virtue/situation family with a virtue/social construction family? I think it is. This comparative claim is of course hard to adjudicate, so I will rest content merely to emphasize the strength of the virtue/prevalence family.

What would it look like to fudge things in the way I recommend? Essentially, one would end up committed to a version of the hypothesis of extended cognition, a variety of active externalism in the family of the extended mind hypothesis. Clark and Chalmers (1998) argued that the vehicles (not just the contents) of some mental states and processes extend beyond the nervous system and even the skin of the agent whose states they are.[9] If my arguments are on the right track, virtues and vices sometimes extend in the same way: the bearers of someone's moral and intellectual virtues sometimes include asocial aspects of the environment and (more frequently) other people's normative and descriptive expectations. What it takes (among other things) for you to be, for instance, open-minded, on this view is that others think of you as open-minded and signal those thoughts to you. When they do, they prompt you to revise your self-concept, to want to live up to their expectations, to expect them to reward open-mindedness and punish closed-mindedness, to reciprocate displays of open-mindedness, and so on. These are all inducements to conduct yourself in an open-minded way, which they will typically notice. When they do, their initial attribution will be corroborated, leading them to strengthen their commitment to it and perhaps to signal that strengthening to you, which in turn is likely to further induce you to conduct yourself in open-minded ways, which will again corroborate their judgment of you, and so on. Such feedback loops are, on my view, partly constitutive of what it means to have a virtue.[10] The realizer of the fudged Ramsey sentence isn't just what's inside the person who has the virtue but also further things outside that person.

So, can people be virtuous? I hope it isn't too disappointing to answer with, "It depends on what you mean by 'can,' 'people,' and 'virtuous.'" If we're concerned only with abstract possibility, perhaps the answer is affirmative. If we are concerned more with the proximal possibility that figures in people's current deliberations, plans, and hopes, we have reason to worry. If we only care whether more than zero people can be virtuous, the existing, statistical, empirical evidence is pretty much useless. If we instead treat 'people' as a generic referring to human animals (perhaps a majority of them, but at least a substantial plurality), such evidence becomes both important and (again) worrisome. If we insist that being virtuous is something that must inhere entirely within the agent who has the virtue, then evidence from social psychology is damning. If instead we allow for the possibility of external character, there is room for hope.[11]

Ramsify (by all Means)—but Do Not 'Dumb Down' the Moral Virtues

James Montmarquet

The Basic Challenge of Situationist Research

In a number of important publications, Mark Alfano has refined a basically naturalist approach to the theory of moral and epistemic character, attempting to understand these in terms of established research findings in the social sciences.[12] In this general endeavor, of course, he is not alone. The dominant ("pragmatist") strain of philosophy in America has long rejected any strict dichotomy of nature and value—or of ethics and the social and natural sciences. I will turn to Alfano's contributions specifically to this volume in due course. But first it may prove instructive to probe, in somewhat more general terms, the kind of broadly naturalist approach to moral character he would advocate, confronting it, where possible, with what seems a viable alternative—not supernaturalism, mind you, but an approach emphasizing the independence and autonomy of moral philosophy, and insisting that our judgments of moral responsibility are too important to be left to social scientists—or maybe philosophers.

With Plato and Aristotle, let us observe, moral philosophy had begun to display this very autonomy and independence relative both to religion and to natural science—even if in the latter case this was mainly because natural science had hardly begun to break free of philosophy. In fact, we have a remarkable intimation of this breakthrough in Socrates' evidently autobiographical discourse in the *Phaedo* (98b-e). Having been as a young man impressed with the physical sciences, he relates how he later came to learn of the insufficiency of their explanations: Why am I sitting here in prison, he asks, awaiting death? Is it because my muscles, bones, and nerves have arranged themselves in these ways—or is it not because I am committed to certain *values*, for which commitment I freely take responsibility and am willing to die?

In the centuries to follow, challenges to moral philosophy's autonomy have come mainly from the sciences: with each great name in this history (Newton, Darwin, Freud, Einstein), a potential challenge is there, mainly emanating from those enthusiasts, inside or outside of philosophy, who would apply the lessons of the new science to 'old moral thinking.' On the whole, it must be said, moral philosophy has withstood these—though with some pretty good battle scars. But now we come to the latest challenge: many experimental results in social sciences appear to question the very idea—so basic to a broadly Aristotelian approach to ethics and to much of our ordinary moral thinking about ourselves and others—of individuals as possessed of, and having their actions considerably shaped by, *moral character*. According to these

results, our actions are, to an unexpected extent, the products of the "situation" in which we find ourselves, and not of distinctive qualities of "moral character."[13] The findings in question are broadly of two kinds:

I. There are ones tending to show that a given situational variable has unexpectedly powerful and largely uniform effects—that is, effects holding largely irrespective of any differences in "moral character" as might have been supposed to exist within the group exposed to this influence.[14]

II. There are also findings tending to show that as one *varies* a situation—but within the confines of what we would take to be a single trait (e.g., honesty), we do not find very much individual consistency. [15] Thus, cheating in one type of educational setting (say, copying another's answers) turns out to be a surprisingly poor predictor of one's cheating in another (say, working past a time deadline).

Such contentions have, to say the least, sparked a lively debate, in which the defenders of traditional virtue ethics have scored their fair share of points.[16] But let me now put my own cards—some of them, those that do not remain up my sleeve—on the table. The situationist's starting point is certain empirical findings and her creed is largely that of "scientific naturalism": roughly, the idea that the findings and methodology of empirical science should be our best guide not just to the nature of the physical universe but in questions of philosophy (of metaphysics and of morality) as well. My starting point, by contrast, is neither science nor a dogmatic adherence to any particular moral outlook or philosophical persuasion; instead I look, initially at least, to the considered moral judgments we make especially concerning moral responsibility (of praise and even more so of blame). Since these judgments are so important, we would like them to be informed by the best, by the most rigorous findings science has to offer; and since questions of moral character are so deeply implicated in these judgments, we would naturally like this science to apply to them. The problem is that insofar as character can be reined in, domesticated and made scientifically serviceable, its most important moral dimensions still have a way of roaming free.

Aristotle's Mixed Legacy

Having indicated the importance I place on the connection between character and responsibility, I immediately yield pride of place briefly to Aristotle, whose ethical theory links praiseworthy and blameworthy conduct to virtuous or vicious dispositions of the agent (*Nicomachean Ethics*, Bk. II, iv-v). Aristotle, however, is also a convenient target for situationists, when he maintains that the virtuous act must reflect a "firm and unchangeable character" (Bk. II, iv). So, for instance, in his early and rather uncompromising statement of situationism, Harman (1999, p. 317) glosses Aristotle's notion of a character trait as a "relatively long-term stable disposition to act in distinctive ways," then

proceeds to argue, along the situationist lines sketched earlier, that no such disposition exists.

Now, Aristotle's idea, sympathetically stated, would seem to be this. If an act does not express the right kind of disposition (inner state), even if it succeeds in being the right thing to do, this reflects a kind of luck; just the right non-virtuous factors have helped out. If, however, an act *does* reflect a virtuous inner state, this state will be of sufficient power and stability as to be repeatable—not just in situations narrowly of that type, but across a greater variety.

Arguably, however, this stance involves—not entirely (as we shall see) but in some measure—a "false dichotomy": arguably, that is to say, we need to recognize a *three-*, and not merely a two-fold distinction in this connection. There is such (superficially) good conduct as may result (1) from not much inner virtue at all, conduct that rightly nets faint praise and that is no very good candidate to be repeated. There is good conduct resulting, as Aristotle would wish, from (2) the stable possession of a virtuous (inner) trait. Finally, there is also good conduct resulting from (3) what may be an equally potent, but relatively unstable, expression of such traits. By analogy: there shots of pure luck by "duffers" in golf; there are shots displaying an excellent, repeatable swing; and there are shots of equal grace and panache pulled off by somewhat erratic young golfers whose swings are not yet consistently good ("grooved").

Of course, there is a kind of "luck" involved in type (3) cases; but this, it also must be pointed out, is not the luck of mere "good fortune." It is good fortune when the duffer's shot hits a rock and bounds out of the hazard, two inches from the pin; and when the coward's act starts a general retreat that ends up saving the army. An act, however, of unusual heroism—performed by a formerly, or just a sometime, very timid soldier—would have been unpredictable, but is not mere good fortune or normally to be described as "mere luck." In fact, this distinction between type (2) and type (3) supports and larger and more significant difference. It is part of the idea of character that it can improve—or worsen—based on what happens in a given case. When this happens, we can in no way judge the moral quality of that display of character according to the *overall* character we suppose this individual to have (or to have had). In the familiar story of *The Red Badge of Courage,*[17] Henry ends up displaying a quality of courage that would have to be heavily, and counter-intuitively, discounted if it were assessed so as to give equal weight to his past cowardice. People grow; people change. As Aristotle himself must recognize, nothing augments one's courage like courage (greater than one's average level to that point): not just doing unusually courageous deeds but with a new inner strength.

Tolstoy and Hume

Our opening discussion has pointed in the direction of a character-based approach to moral responsibility, but emphasizing—what Aristotle does

not—the distinction between the praiseworthiness or blameworthiness of a given character exemplification (this courage of Henry's) and that agent's conduct measured over time (Henry's typical level of courage). To this opening, we now add two contrasting (but not incompatible) lines of thought:

(1) *The Tolstoyan Idea.* Character is not well revealed most of the time, but is so in certain especially suitable, often quite difficult, situations. I associate this idea with Tolstoy, as his fiction displays this author's remarkable gift for revealing character by the apt selection of situation and response. Such situations are of an especially challenging nature: to name but one, there is the life-threatening blizzard in his story, *Master and Man* and the quite different responses to it of the rich merchant and his servant.[18] From this standpoint, cases in which it is "easy to be honest" (or where the Russian winter is merely "a bit cold") are hardly worth bothering about; for even if an agent should fail in one of these, this is typically a sign not of amazingly bad character so much as that more is going on than meets the eye.

(2) *Humean Intelligibility.* Even though Hume famously said, we must judge another's conduct not by his acts, which are "temporary and perishing" but by something "durable [and] constant" in his nature—that is, traits of moral character[19]—this does not mean that he must deny that assessments of praise and blame will properly take into consideration differences of situation. Thus, suppose that some English gentleman, known to be scrupulously honest in his business dealings, has been exposed as lying, on a given occasion, to a prostitute. We have found out something new and perhaps a little surprising about him. The Humean point, however, is not just to add this new information on—as though to say, seemingly with Doris, that the man has "business honesty" and "prostitute" (or, if we dare to generalize, "sexual") dishonesty, but to begin to form some revised but still unified picture of "who this man is."[20] We must add, however, that achieving this fuller picture—not just of 'who he is in general' but 'who he has shown himself to be in this case'—is strictly an art, not a science. Those with a penetrating insight into the true moral significance of a given human act (what it truly displays of that agent) are more like a Rembrandt or some other great portraitist than a Newton or an Einstein.

Now both the Tolstoyan and the Humean viewpoints, let us observe, remain broadly 'Aristotelian' in two basic regards:

First, in recognizing that the implications of any given character attribution extend beyond the present act, both would recognize, with Aristotle, the dispositional character of character attributions. Both seek some discovery or understanding of morally 'who one is'—not in general but so far as this is revealed in the present action. But any such judgment must carry implications—tentative and inexact as these may be—for future acts. Thomas Hurka (2006) has remarked, with some justice, that if you see someone

kick a dog from an evident desire to hurt the dog just for the pleasure of doing so. Do you say, 'That was a vicious act' or 'That was a vicious act on the condition that it issued from a stable disposition to give similar kicks in similar circumstances'? Surely you say the former. (p. 71)

At a minimum, however, such a display—supposing, with Hurka, that it was of genuine (and not merely apparent) viciousness—would have some implications, however inexact, regarding future displays. We have found out something, something rather unpleasant, about this agent. As long as he is around, we must worry that his viciousness may be around as well. *Pace* Hurka, we do not have to make a prior or an independent discovery of a continuing trait in order to attribute something that would carry implications, to a greater or lesser degree, to that effect. If I see someone lift three hundred pounds, I can call him "strong" without further ado; but again this has implications.

Second, neither point of view rejects the Aristotelian notion that a virtuous act must display *some* suitable connection to one's past character displays. In fact, both will say this of any act for which one has moral responsibility. As in the previous example of the Victorian gentleman, the Humean portrait employs shades often drawn from a past understanding of this agent. It is a composite of past and present. For his part, once the challenge of the situation has been met (or not met) Tolstoy will equally want us to understand 'who this person has shown himself to be'—not in abstraction from the past, but in the present case, using the past to help us understand its meaning. Thus, in another story of his ("Father Sergius"[21]), a proud man passes through a whole series of failures—cadet of the Imperial Guard, suitor, monk, hermit, vagabond—until finally discovering true pride from one possessed of none of his advantages. Little in what happens could have been predicted, but all of it needs to be comprehended under the aspect of his pride and his past.

For both points of view, then, one needs to recognize links to the past and implications regarding the future—neither of which, it is important to emphasize at this point, are altogether satisfactory from a rigorously scientific standpoint. We draw on the past, in a selective, hopefully insightful way, to paint—in what is becoming my central metaphor—a picture of the present, a picture whose implications are in some ways profound, but still profoundly inexact. Thus, to vary our mode of artistic appeal, the reader of such a fine biography as Ron Chernow's *Alexander Hamilton*[22] will assuredly come to have much better understanding of this man than can be gleaned from superficial historical facts concerning him; but I would not suppose that it would help one to predict, or retrodict even, Hamilton's actions in a difficult case. Character is not a terribly good predictor—not because it is ineffectual, but because it is effected (and affected) in such complicated ways. What is the "predictive value" of a Rembrandt?

One last point in this connection, it is helpful to remember that character terms bear much in common with such impolite and obviously unscientific terms of opprobrium as "schmuck" or "bastard." There will be occasions when these are quite the correct terms to characterize what someone has done—as in "You say he did that? What a bastard!"—so it is not as though just 'anything goes' with their use in moral contexts. Still, it must be said that their application is more an art than any part of a science.

Hanna's Question

So far, our preoccupation has been with the loose, the unscientific—but not for that the unimportant—nature of character attributions and their connection to past and future. At this point, however, the situationist may raise a useful objection. 'I could almost agree with what you say concerning the independence, and the limited dependence, of situational evaluation on how well or badly an individual has done in other situations. I do not agree, however, if this is your position, that how well *other* agents would do in a given type of situation is irrelevant to the evaluation of a given agent. Learning how well most people function in Milgram experiments can and arguably should affect our judgments of blameworthiness.'

I feel the force of this concern, and add a reference of my own. As an admirer in general of contemporary cinema and of Kate Winslet in particular, I am reminded of the former concentration camp guard, Hanna's, question in the film, *The Reader*: confronted with her own crime of not unlocking the inmates caught in the burning church, she asks her main accuser, the presiding judge: "What would you have done?" There is no answer.

Again, following our usual practice, we will distinguish two different views on the relevance of what others have done or would do, in a given situation:

In a spirit of Tolstoyan—now become Kantian—boldness, it might be urged that we are free, guided only by reason and a sense of what is possible, to set the bar as high as we like, whatever empirical evidences indicates regarding 'normal performance.' (Even in golf, par is not set by the average number of strokes on a hole.) Thus, the judge might have informed Hanna: 'I would hope that I wouldn't have acted as you did, but if I did act in that way, then I must be judged guilty.'

The Humean is able to craft a more nuanced position. If we are talking somewhat superficially about kinds of action, I will agree that one's estimation of how bad (or good) a given kind is, would take into account what the norm would be regarding its performance. But that is not the same thing as the question of how a given agent would properly be evaluated—more fully, in more detail—if he were to do a given act. In the case of keeping the prisoners locked up, if the judge were to have done it, this would have quite a different moral significance than Hanna's act. Basically, the more we keep of the judge's

present moral make-up, knowledge, and sophistication—the worse becomes his offense, the more it becomes an expression of sheer cowardice. As to the Milgram case, the Humean might say this: the behavior of others will to a point help us to understand what a given agent does, but so far as excusing, or even just diminishing the moral badness of, that behavior is concerned—this is neither necessary nor sufficient. The more deeply we understand, the more deeply we penetrate into the motives and moral personality of a given agent, the less we should need to rely on the inevitably different story concerning others.

Finally, by way of comparison, we turn to a situationism-influenced ethicist who has worked out in some detail a conception not only of moral character but of moral responsibility: John Doris (2002). Helpfully, Doris begins his account of responsibility by distinguishing the "intensity" and "frequency" aspects of a possible excusing condition, noting that even if we can isolate a certain situational variable as causally salient, this does not mean that it possesses the kind of intensity—the quality of being "difficult to resist"—which ought to excuse (p. 135). Still, he points to the fact that situationist research may uncover certain "unobtrusive high-intensity stimuli" that seemingly do exert a decisive effect on behavior and thus bid fair to be counted as "undetected excusing conditions" (p. 136). Doris addresses this possibility as, first, a worry concerning "normative competence": our capacity to recognize, reason, and act on morally relevant features of one's situation. Ultimately, though, it becomes a problem concerning "identification": insofar as these variables are hidden, we do not identify with them, thus, seemingly, cannot be held responsible for acting upon them (p. 140).

This, if it had been Doris's conclusion, would have been clear, illuminating, and fairly radical in the extent to which it upsets our common sense, folk psychological notions of responsibility. Doris, however, elects not to go this route. Instead, he argues that even if one is unaware of a situational factor, it may be possible to work up a "narrative" (pp. 142–143) underlying one's action with which one plausibly would identify. Thus, Doris maintains that, in the Darley and Batson, "Good Samaritan"—in which those who were late tended not to stop to help a man apparently in distress—"even where a person fails to identify with the callousness that resulted from haste, he might yet embrace having the sort of packed schedule that induces haste" (p. 144). We may, in other words, find grounds for holding him responsible (morally blameworthy) after all.

I find it most interesting, then, that Doris's bold naturalism when the topic is "virtue ethics" turns into a cautious conservatism when it comes to offering widespread exemption from moral responsibility. (This is not an unusual combination in the history of moral philosophy: attacking the received grounds for moral responsibility, but feeling pressed in the end to supply replacement grounds of one's own.) Doris has in any case worked his way around—as he

may choose—to either the Kantian or the Humean position. One may find in the story of Hanna's life times enough of a humanitarian narrative for her to identify with—so that she may be held to quite a high moral standard. There will also, however, be competing narratives, running closer to the motivations actually at work in this case, whose effect would be exculpatory. In the end, the choice simply lies with us (including the judge) as moral observers. We would *like* assistance from social science—or religion, or philosophy, or something—but it is not clear that anything altogether helpful is available.

An Irenic Approach, Initially

In his contribution to the present volume, Mark Alfano has offered what seems initially a peaceable approach to these situationist wars, an analytical framework with respect to which various approaches (including but by no means limited to his own) can be represented. On this view, we characterize 'moral character traits' as a kind of theoretical entity able to satisfy various types, and combinations of types, of open sentences. These sentences will collect some of the more important truths, or putative truths ("platitudes") concerning a given trait of moral character. Interestingly, it will treat not only character terms ('courage,' 'moral character,' etc.), but "persons" as theoretical terms, implicitly defined by the theory.

Now, as a device, this seems a fair and neutral way of representing certain possible views of moral character; but of course controversial claims may be, as Alfano points out, built into it one or another selection of platitudes. So, for instance, the claim that "if a person has courage, then she will typically be unaffected by situational factors that are neither reasons for nor reasons against overcoming a threat"; coupled with the claim that "many people have courage" would be, for some situationists, "fighting words." But let me focus on two aspects of particular relevance to the foregoing discussion.

First, the notion of "sufficient reasons" figures in a very prominent way in these partial characterizations. Thus, it is said that a courageous person will act in certain ways, will feel certain things, will be unaffected by situational factors—all when there is "sufficient reason" for this. But apply this back to the case of Hanna: Is the situation that one is a concentration camp guard a sufficient reason to omit what would otherwise count as simple kindness and benevolence, etc.? If our notion of sufficiency is sharply Kantian ("of *course* she had sufficient reason to help"), many, even most, people will ultimately prove unkind and not very benevolent. If it is not, we will get a different verdict. In short, the notion of 'sufficient reason,' as it figures in these collections, threatens but to recapitulate the most troubling aspects of cases like Hanna's (or the Milgram subjects').

Second, there seems one notable difference between the Lewis and Alfano projects. In Lewis's case, the interesting development is that, for all the different

things believed about these inner states, it turns out (arguably) that they are brain states. What, comparably, may we discover through Ramsification concerning what moral character turns out to be? Nothing equally surprising or exciting, some may think—but Alfano, I think, would disagree. According to his particular conception of moral character, at least, it turns out that what realizes the relevant Ramsey sentence includes things not going on solely "inside the person who has the virtue" (p. 225).

This, it could be supposed, is both surprising and exciting, but I remain a bit skeptical. Here we must distinguish the full reference of our theory of character, which may include all sorts of things (persons, states of persons, features of the environment, situations, and so forth), and what is referred to by such discourse as "my courage" or "her patience"—or what Alfano himself, in the passage just quoted, speaks of as the person "who has" a given virtue. I am not sure that Alfano is altogether clear in his way of putting this matter. He will say, at the very end (p. 226), that according to some very good social psychology, "being virtuous" does not "inhere entirely within the agent" (again "the agent that has that virtue"). This strikes me as either a truism (that the full range of what must be the case in order to "have a virtue" will include much that is outside of oneself)—or what hardly seems a truism, or true at all, that when Jones is praised for her courage (the courage she "has"), she is being praised for something which is located all over the place (including wherever those people or situations that have helped to encourage this behavior are located). Perhaps Alfano will say that we cannot properly make this distinction (between the full reference of the theory and what things like "Jones's courage" pick out); he might say that this is like asking what the relation is between the reference of Lewis's Ramsification (brain states) and what we meant to refer to in our old folk psychological talk. I will in any case return later to this important topic: the relation between moral praise for courage and what social science may teach us about such qualities.

Virtue Prevalence and Social Construction

I turn, first, to another important concern of Alfano's, which has been to emphasize the price to be paid for giving up, as some character theorists wish to do, "virtue prevalence"—that is, the thesis that virtues are quite widespread throughout the moral community, and are not merely the possession of a few moral heroes. Along these lines, Alfano seizes on a point of Philippa Foot, who makes much of the virtues as being qualities that we "need" as much as the deer needs speed or the bee its stinger. He observes, ironically, that if the virtues were both this necessary and as rare as some suggest, "one would have thought that our species would have died out long ago" (p. 224).

I hold no particular brief for Foot, but would simply draw an analogy here between situationist findings concerning the non-prevalence of virtue,

and discouraging test results concerning American students in science and mathematics. Of course, the latter hardly show that the skills measured by such tests are not important. Likewise, it could be argued, virtue ethicists are entitled to hold that the virtues (and especially the vices) remain important despite the discouraging test results of social science. An ethics of virtue does not hold that we will "test well" for these, so much as that if we do not, this is important—and disturbing. Thus, concern over Milgram's results, like concern with standardized test results, shows how important we take the virtues and vices to be. Moreover, to extend the analogy, we need not deny that these qualities—be they moral, mathematical, or scientific— are broadly necessary for "survival"—but of course we cannot maintain in either case that, at their current, admittedly low levels, we are faced with imminent extinction. Luckily, speed in mathematics is not quite as important for us as speed of foot is for the deer. For that matter, "obedience to authority" may have, or have had, (*pace* Milgram) greater survival value than "moral independence."

To continue the previous thought, it seems evident that certain kinds and levels of virtue are required for—call it, "social solidarity" but that these may not agree with qualities morally sensitive people praise and claim to value. Again, it may be that getting to one's assigned place on time has greater survival value than helping strangers (who may be planning an ambush). It may be that the limitations social scientists have uncovered in our moral characters are actually evolutionary virtues of sorts.

"Fair enough," Alfano may say, "I have just the conception you need for your 'social solidarity' and it is my 'social construction' conception you took swipe at earlier." In fact, that previous discussion notwithstanding, I will agree—partly. It may well be that what survival demands are more or less the very qualities good social science picks out as highly prevalent. Now, his social construction view Alfano describes in these terms:

> What it takes (among other things) for you to be, for instance, open-minded . . . is that others think of you as open-minded and signal those thoughts to you. When they do, they prompt you to revise your self-concept, to want to live up to their expectations, to expect them to reward open-mindedness and punish closed-mindedness, and so on. These are all inducements to think in an open-minded way, which they will typically notice. When they do, their initial attribution will be corroborated, leading them to strengthen their commitment to it, and perhaps to signal that strengthening to you . . . Such feedback loops are, on my view, partly constitutive of what it means to have a virtue.

I would like to probe two areas of possible difficulty for such a view, the second of which will take us back to the main theme of my discussion.

First, there are important questions concerning the *vices*. Are these, too, socially induced by the same kind of feedback loops? Or do they mark, to the contrary, merely failed (positive) loops? It is, let me say in passing, characteristic of the literature on situationism that questions of the virtues, their consistency, and prevalence entirely dominate questions of vice (their consistency and prevalence)—which in a way is strange, inasmuch as questions of blame, not praise, dominate whole other areas of moral philosophy. Now, for his part, Alfano refers to the fact that "social scientists think that people are quick to attribute virtues"; but this tendency would surely suggest that we are quick, perhaps even quicker, to attribute vices. If that is correct, we have some fairly substantial grounds for thinking of the vices as, in great part, acquired via the same social influences as generate, and partly constitute, the virtues. The danger for Alfano, I submit, is now the very distinction between virtue and vice, good and bad, looks to be entirely external to the theory—almost as though it were an afterthought. Of course, this problem does not arise—not in such an acute way—if we think of the vices according to the other suggestion (as failed virtues); but now we seem to be going against the grain of what social science is telling us concerning the acquisition of all qualities (good and bad); moreover, there is also the question of whether, or how, the mere failure of a "feedback loop" would be morally blameworthy.

Second, and most importantly, there is the issue with which we began: Aristotle's formative claim that we are to be praised and blamed primarily for our virtues and vices. The issue becomes this. If we follow social science, we will arrive at a variety of conceptions, one of which is Alfano's 'social construction' view. To extend my earlier comparison, these conceptions would be like attempts to redefine ("dumb down") mathematical and scientific skills in such a way that it turns out that most children *do* have these to a satisfactory degree after all (much relief all around!). However, just as rethinking mathematical ability does not make anyone better at solving mathematics problems, redefinitions of 'virtue' do not actually make any one morally better—just ask someone who is lying along the road, waiting for a Good Samaritan to come along.

But I want to be fair to the social sciences: let us allow that when I refer to courage, I may in some sense be referring to a trait whose nature is such as to be characteristically induced by certain types of social cues, expectations, and feedback loops of the sort Alfano describes. Again, though, the fundamental limitation of this conception is that there is nothing terribly praiseworthy about responding to such social indicators—not as such. Rather, one must suppose that the praiseworthiness of courage lies in the fact that sometimes it is very *difficult* to respond to such indicators (think of the soldiers comprising "The Charge of the Light Brigade"). If you like, sometimes it takes real courage to exhibit this kind of "social courage." Likewise, real generosity and

helpfulness would begin where socially cued generosity and helpfulness leave off. By the same token, the former would be rare; the latter, prevalent.

So, in the end, we can have our cake and eat it too—just so long as we do not confuse those parts of it that are real, and rare, from those parts that are mass produced. We can frame, as Mark Alfano is doing, a sophisticated philosophical notion, drawn from social science research. We can also reflect on what is genuinely praiseworthy and genuinely blameworthy; but this, I want to say, will carry us beyond questions of social science, evolution, and survival.
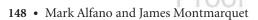
Socrates' Unanswered Question

"Thou shalt not sit/with statisticians nor commit/A social science," warned a noted poet.[23] We will adopt a somewhat more forgiving stance. There is certainly a role for the social and biological sciences in understanding moral character and morality generally. The giants of our subject—Aristotle, Kant, Mill—have generally kept clear of explaining moral conduct (sociologically, psychologically, biologically), recognizing that explaining it would run the extreme risk of explaining it away. There is another way of putting these matters, which hearkens back to what I said at the outset. We would like the most important things to be studied in the best of ways. Science has proven itself to be the best, so we would like to study moral character and responsibility scientifically. We would like what is most important to us to be very clear to us. All the same, Socrates' question—what am I doing here (in prison)?—seems no more answerable by today's than by the sciences of his day. If its answer is clear (to him), answers to comparable questions (including Hanna's) are anything but.

Study Questions

1. Is someone's character better revealed by how they live their life in ordinary circumstances or how they respond to extraordinary circumstances?
2. Arguably, Aristotle thought that friendship was "extended" in the way Alfano says all virtues are extended. Do you think virtues such as trustworthiness are extended? Explain.
3. How well would you need to know someone to be able to paint a Humean picture of their character? How many people do you know that well (including yourself)? Would it be a problem if we never knew each other or ourselves this well?
4. What does Alfano mean by saying that "the bearers of someone's moral and intellectual virtues sometimes include asocial aspects of the environment and (more frequently) other people's normative and descriptive expectations"?

5. Ethical naturalism is endorsed by some of the authors in this volume, especially Snow, Russell, Miller, and Alfano; it's rejected by others, especially Montmarquet. What's the best argument in favor of ethical naturalism? What's the best argument against it?

6. How would Alfano likely address Montmarquet's worry that in explaining moral character and responsibility in ways suggested by the social sciences, we run the risk of explaining them away?

Notes

1. Nathan is also using an extended metaphor. My point is clear nevertheless.
2. An alternative is the "psycho-functionalist" method, which disregards common sense in favor of (solely) highly corroborated scientific claims. See Kim (2011) for an overview. For my purposes, psycho-functionalism is less appropriate, since (among other things) it is more in danger of changing the topic.
3. See, for instance, Zagzebski (1996, p. 112): "It does count against a person's virtue, however, if stractions or persuasions lead him to fail to exercise it."
4. I seem to be in disagreement on this point with Christian Miller (this volume), who worries that people may not be motivated to be or become virtuous. In general, I'm even more skeptical than Miller about the prospects of virtue theory, but in this case I find myself playing the part of the optimist.
5. I am here indebted to Gideon Rosen.
6. It might also be possible to circumvent this difficulty, which anyway troubles Lewis's application of Ramsification to the mind-brain identity theory, by using only *de re* formulations of the relevant statements. See Fitting and Mendelsohn (1999) for a discussion of how to do so.
7. Experimental philosophers have started to fill this gap, but not in any systematic or consensus-based way.
8. Micah Lott (personal communication) has told me that he endorses this claim, though he has a related worry. In short, his concern is to explain how, given the alleged rarity of virtue, most people manage to live decent enough lives.
9. For an overview of the varieties of externalism, see Carter et al. (forthcoming).
10. I spell out this view in more detail in Alfano (forthcoming a) and Alfano and Skorburg (forthcoming). For a treatment of the feedback-loops model in the context of the extended mind rather than the character debate, see Palermos (forthcoming).
11. I am grateful to J. Adam Carter, Orestis Palermos, and Micah Lott for comments on a draft of this chapter.
12. See mainly (Alfano 2013a), especially his description of his approach to ethics involving reasoning in an "abductive" way (p. 5) from scientific findings to what must be the case, philosophically, for those to be true. This general viewpoint informs his treatment of epistemic character in (Alfano 2012).
13. For important discussion of these results (and defense of what has come to be called the situationist viewpoint regarding moral character), see Harman (1999, 2000) and Doris (1998), but especially Doris's extended and extraordinarily valuable treatment of these issues in *Lack of Character: Personality and Moral Behavior* (2002).
14. Among the best known of these are, respectively, the work of Isen and Levin (1972)—in which the remarkable effects of a found dime on helping behavior are demonstrated; Darley and Batson (1973)—in which the effects of 'being in a hurry because late' on Good Samaritan behavior are presented, and, finally, the famous Milgram (1974) experiments concerning obedience to authority.

15. Most commonly cited here would be Hartshorne and May (1928), in which it was shown how the likelihood of dishonesty or cheating among school children tended to vary, for any given individual, considerably from situation to situation.
16. See, for example, Kupperman (2001), Sreenivasan (2002), Montmarquet (2003), Kamtekar (2004), Sabini and Silver (2005), and Adams (2006).
17. Stephen Crane (New York: Dover, 1990).
18. In *Great Short Works of Tolstoy* (New York: Harper and Row, 1967).
19. *Treatise of Human Nature*, Bk. II, Section II.
20. For Hume (1975/2007, 8.10), the "uniformity of human action" must take account of the "diversity of characters, prejudices, and opinions." Here he adds, tellingly, that even if an action occurred with "no regular connection to any known motive"—this would provoke no very clear or definite judgment from the observer—except wonder.
21. Also in *Great Short Works of Tolstoy*.
22. (New York: Penguin, 2005).
23. W. H. Auden in *Under Which Lyre: A Reactionary Tract for the Times.*