2 Reasoning about Wrong Reasons, No Reasons, and Reasons of Virtue[1]

Neera K. Badhwar

I. INTRODUCTION

Aristotle holds that the moral virtues are enduring or stable integrated intellectual-emotional dispositions to deliberate, feel, and act rightly, that they are global or cross-situationally consistent, and that they are reciprocal or united. Guided by *phronesis*, or practical wisdom, the virtuous person always does the right thing for the right reason, in the right manner, at the right time, whatever the circumstances. The virtuous individual "will never do hateful and base actions,"[2] because he acts "from a firm and unchanging state."[3] Indeed, virtuous activities are "more enduring even than our knowledge of the sciences."[4]

Contemporary virtue ethicists are by and large agreed that actual virtue—virtue as instantiated in human beings—falls short of the ideal described here. This, however, is not a problem if we understand that the ideal in the texts is just that: an ideal to aspire to, even if no one can ever reach it.[5] Shane Drefcinski has argued that Aristotle gives us another, more realistic picture of virtue in the *NE* and *Politics,* alongside the picture of perfect virtue as an ideal to aspire to.[6] Since the text sometimes presents virtue as a state of perfection, and sometimes as a state of excellence-but-not-perfection, I agree with Drefcinski that Aristotle presents us with two pictures of virtue, ideal and actual, rather than with an entirely idealistic or entirely realistic picture.[7]

The arguments against globalism, unity, and stability over an entire lifetime have been thoroughly hashed, so I will not spend much time on them. My focus will be the requirement of right reason and the challenge it faces. The challenge, in brief, is that in many morally significant situations, most of us act for reasons we are unaware of—or for no reason at all. If we do the right thing, it is often for the wrong reason. If we do the wrong thing, it is often because features of our situation trigger automatic cognitive processes that bypass intentional control.[8] Our lack of awareness of our real reasons for, or causes of, our behavior, challenges the idea that we are in rational control of our actions and that most of us have any genuine virtues. This, say the critics, makes virtue ethics 'empirically inadequate.' But the requirement that the agent act for the right reason and be in rational control of her

actions is a requirement of almost every ethical theory: Kantian as well as rule-, motive-, virtue-, and even some versions of act-, consequentialist theories. So if it turns out that most actions, even right actions, are unwittingly done for the wrong reason, or no reason, then every ethical theory is empirically inadequate and must be modified. I will argue that neo-Aristotelian virtue ethics can best meet the challenge from descriptive psychology because it is committed to basing its requirements on human nature. If some of its requirements are beyond our capacities, then those requirements have no place in the theory, except as part of a regulative ideal. A good life for virtue ethics is, in some sense, a fulfillment of human nature. By contrast, Immanuel Kant's theory calls for a purity of motive that depth psychology has shown to be impossible. And although most consequentialist theories can trim their requirements in light of our psychological capacities, the imperative to do so comes not from any theoretical commitment to psychological realism but from the pragmatic imperative to offer a theory that most people can follow.

Too often, however, the critics' claim that people unwittingly act for the wrong or no reason is itself made without good reason because the evidence supports alternative interpretations that are compatible with the agent acting for the right reason. Or so I will argue.

In Section II, I provide an overview of Aristotle's conceptions of ideal and actual virtue in the *NE* and *Politics* and contemporary virtue ethicists' views of virtue as a less-than-perfect excellence. In Sections III and IV, I describe the experiments that lead situationist critics to argue that the virtue ethical model of rational deliberation is mistaken or that most of us don't have genuine virtues, and I challenge some of their interpretations of the experiments.

II.   VIRTUE: IDEAL AND ACTUAL

(i)   Cross-Situational Consistency and the Unity of Virtue

In *NE* Bk. VI, Aristotle argues that each virtue requires phronesis, that phronesis is one, and that phronesis entails all the virtues.[9] Hence, one virtue entails all the virtues. For example, if Generoso is generous, he must also be just, courageous, honest, and so on. This is the unity of virtue doctrine.

The unity thesis in this pure form has long been rejected by many philosophers, including those who accept Aristotle's conception of virtue as consisting of both phronesis and emotional and action dispositions to act for the right reason in the right way at the right time.[10] Common sense concurs because, after all, 'Aren't there people who are generous but not very courageous?' Implicit in the argument for the unity doctrine is the claim that the virtues, like phronesis, are global or cross-situationally consistent. It is this assumption, as I've argued elsewhere, that is the truly problematic one because it is inconsistent with recognition of our epistemic and emotional limitations vis-à-vis the high bar set by virtue ethics.[11] Globalism claims that phronesis

is one or global in the sense that we cannot have it in some areas of our lives and not others. Since phronesis entails each of the virtues, and each of the virtues entails phronesis, it follows that each virtue is also global or cross-situationally consistent. For example, if Generoso is generous towards his students, gladly giving freely and appropriately of his time to his students, he must also be generous towards his siblings, parents, children, coworkers, and so on, insofar as giving them freely of his time (or money, or praise) is appropriate. (Here and throughout I ignore Aristotle's thesis of the proper spheres of the virtues: for example, that generosity concerns only wealth, mild-temperedness concerns only unjustified slights, and so on.) If Justine is just in the courtroom, she must also be just in the classroom.

Both history and everyday observation of human nature and behavior give the lie to this globalist doctrine, and social and cognitive psychology provide experimental evidence against it.[12] We are familiar with figures famous for their unwavering dedication to noble causes at great cost to themselves—along with surprising character flaws in their lives as husbands or fathers. We also all know that our fantastic spouse or parents or children or siblings (and maybe even our own selves) are not paragons of virtue. A few rare individuals may have what virtue requires in every aspect of their everyday lives: the self-understanding and understanding of others that is necessary for a sure grasp of the true worth of things, the motivation to consistently act on this understanding easily and with pleasure, and the fine sensitivity to particulars necessary for discerning what is required in a given situation. But even these rare individuals are bound to be partisans of at least some false theories about the nature of government, law, free markets, or drugs, theories that ensure that many of their public actions, such as teaching or writing about political, legal, or economic matters, engaging in political discourse, voting, aiding this or that cause, serving on a jury, and so forth, are contrary to virtue—or *would* be contrary to virtue if they were to engage in them.

Interestingly, even Aristotle seems to reject globalism at *NE* 1115a 20–22, where he states that some people who are cowards in war nevertheless face the prospect of monetary loss with confidence, as required by courage. Admittedly, he goes on to say that such people are courageous only "by similarity" with the brave in war, who are the truly brave, because bravery should be defined with reference to the "greatest and finest danger," and only war is such a danger because it threatens death, is fought for a good cause, and allows the brave to use their strength.[13] But this argument is not very persuasive, since facing the fear of financial ruin (or, for that matter, other dangers) well also meets these criteria. So if Aristotle accepts that the virtue in question is bravery proper, his view implies that bravery (and thus phronesis) is not global because some brave-in-money-matters individuals are cowardly on the battlefield. If he insists that the virtue in question isn't really bravery, we reach the same conclusion about phronesis not being global, since the virtuous-in-money-matters individual must have phronesis

in money matters but not on the battlefield.[14] And both readings entail that, if the virtues are united, they are united only within certain spheres, such that the virtuous-in-money-matters but cowardly-on-the-battlefield individual has all the other virtues concerned with money, but none concerned with the battlefield.[15]

Contemporary virtue ethicists generally accept that globalism as a thesis about actual virtue and phronesis—as opposed to ideal virtue and phronesis—has been defeated. The defeat of globalism, however, does not spell the defeat of virtue ethics. Even if dispositions are not global, if people exhibit reliable patterns in their attitudes and behavior, if their deliberative, emotional, and behavioral responses to people and events they encounter daily in pursuit of their goals are usually consistent and predictable, they have dispositions, and if those goals and dispositions are praiseworthy, they have virtues.[16] For example, if Eidos is kind in most important everyday contexts but surprisingly unkind in a few, she is still pretty kind. To use a metaphor: if the strands in a skein of multicolored wool are mostly red, then the skein itself is pretty red, even if it contains many black, white, and green strands. This "aggregative solution," first endorsed for personality traits by social and personality psychologists such as Walter Mischel and Seymour Epstein, works just as well for character traits.[17]

(ii)   Ideal Virtue and the Power of the Situation

The question of perfection can come up even if people's virtues don't extend over all important contexts. Do context-specific virtues require context-specific perfection, or are they compatible with sometimes doing the wrong thing, or the right thing for the wrong reason? Can the actions they lead to be virtuous even if motivation sometimes lags, one's manner is a little off, or the timing is not quite opportune? The common sense answer to these questions is 'Yes.' Eidos's kindness as a teacher and mother is genuine kindness, even if she is sometimes uncharacteristically unkind or less than wholeheartedly kind. And common sense is right here since, after all, we don't expect perfection in any other human skill or achievement. Some philosophers also argue that a trait is a virtue so long as it is good enough, or *satis*.[18] Indeed, in the following passages, Aristotle himself offers a 'good enough' conception of virtue and virtuous action.

One of Aristotle's reasons for this more realistic conception of virtue is epistemic. He argues that it is hard to say exactly what the right thing to do is in each case; hence, even the virtuous individual can find it hard to always discern the mean.[19] Thus, although the generous person takes care of his own property and gives the right amount for the right reason with pleasure, he sometimes "deviates from what is fine and right" and, to his own regret, gives too much.[20] He sometimes deviates in the other direction as well, failing to spend "what it was right to spend" and is "more grieved" by this than by the opposite mistake.[21] Likewise, although the good-tempered

or mild person is disposed to be "angry at the right things and toward the right people . . . in the right way, at the right time, and for the right length of time," he sometimes errs "in the direction of deficiency, since [he] . . . is ready to pardon, not eager to exact a penalty."[22] Some deviations from perfect virtue, then, are unavoidable, given the complexity of situations and the multifaceted nature of virtuous responses.

Precisely because they are unavoidable, however, these deviations are not very serious. But Aristotle also makes exceptions for situations that "overstrain human nature."[23] For example, a man may succumb to the tyrant's demand to do something shameful in order to avoid grievous harm to his family.[24] In such situations, people are either praised or excused for their shameful action, even though it was avoidable.[25] In *Politics* 1286b 27, Aristotle goes further, excusing a (virtuous) king who allows his unfit son to inherit his throne because, he claims, denying one's son the throne is too much to ask of a human being.

Some of us might question Aristotle's claim that the right action in the latter situation would be too much to ask of a person. The point, however, is that Aristotle realizes that there are limits to what human beings can be expected to do, and that certain situations test these limits and render people's wrong actions excusable. But Aristotle makes even more surprising concessions to the frailties of human nature and the power of the situation: he allows that in situations of great temptation or passion, even a virtuous man can defraud another or commit adultery without losing his virtue.[26]

Here, Aristotle's standards are rather lower than those of contemporary critics who sometimes suggest that just one unjust act is enough to defeat someone's claim to justice.[27] Of course, some singular acts of injustice *are* enough. For example, no one would argue that murdering someone in order to inherit her fortune just once is compatible with being a just or mostly just person.[28] What's important here, however, is that Aristotle is aware that certain situations can lead even a virtuous person astray, and he regards this as being consistent with the virtuous person's continuing to be virtuous—unless and until, of course, his vicious actions become a theme of his life.

This last is a real possibility if the corrupting situation lasts long enough. A well-known situation of this kind is political power unconstrained by the rule of law, "for desire is a wild beast, and passion perverts the minds of rulers, even when they are the best of men."[29] So although virtue is enduring, it is not immune to the influence of situational factors and the passions they evoke: A virtuous man can become vicious if he remains in a corrupting situation. If asked what cognitive-affective processes lead to this outcome, Aristotle would probably reply that human beings tend to be tempted by power and that when they are surrounded by sycophants, it is easy for them to believe that they are acting in the best interest of the public. Over time, they lose all critical insight and start thinking that their tyrannical policies and actions are justified. Such blindness, indeed, is the very essence of the vicious man, who acts viciously *on principle*. Contemporary virtue ethicists,

aware of the many ways our minds can trick us, can also cite self-deception and confirmation bias: a refusal to pay attention to disturbing truths, rationalizing our bad actions as done for good reasons, avoiding those who see us as we are in all our unflattering unloveliness and seeking those who don't in order to preserve our shining conception of ourselves, and so forth.

All these are ways in which we human beings are alike. Where we differ is in what we do about these tendencies when we become aware of them. Whereas some people try to resist them by changing their actions and ways of thinking, others say, in effect, 'Well then, so be it, I'm human after all.'

How does the model of the mind emerging from cognitive science complicate this picture, according to situationist philosophers? By adding that we often act—*and cannot help acting*—contrary to our reflectively held values, for reasons we are unaware of, or for no reason at all. In innumerable everyday situations, most people behave alike, and predictably so, thanks to "depersonalized response tendencies" that *resist or bypass* "intentional direction."[30] Such "moral dissociation" between values and behavior show the limits of practical rationality and challenge the image of the (virtuous) agent as someone who typically knows what he is doing and why. In the next section, I will analyze and evaluate these claims.

III.   MORAL DISSOCIATION: MOODY HELPERS AND UNHELPFUL BYSTANDERS

The critics make their case by focusing on the subliminal influences on our attention that prevent us from acting in other-regarding ways. They describe well-replicated experiments that show the influence of common situational factors on morally significant behavior and the subjects' lack of introspective access to their influence. One such set of experiments concerns the effect of ambient smells and noises and of good or bad moods.

(i)   Moods and Other Non-Reasons

A number of studies have shown that people are more helpful when their environment is pleasant and less helpful when it is unpleasant.[31] What makes their environment pleasant or unpleasant can be as trivial as the smell of freshly baked bread or the noise of a loud (85 dB) lawnmower, respectively. In one experiment, whereas 80% of people helped a man who seemed to be injured pick up his books, only 15% did so in the presence of a lawnmower. Noise levels above 80 dB significantly affect people's behavior. More seriously, someone in an aggressive mood is likely to be more aggressive if subjected to a loud noise. Again, many studies have found that people are far more helpful when they are in a good mood than in a sad or depressed mood.[32] Similarly, happy people are more likely to notice and respond to another's need,[33] a fact noted by philosophers as far apart as Kant

and Nietzsche. Embarrassment or guilt can also lead an individual to be more helpful.

I am in no position to verify these claims by studying all the experiments myself, but they seem plausible enough. My task is simply to evaluate the philosophical arguments made on their basis. The first point to note is the problem of *the forgotten minority*. The economist William Graham Sumner spoke of the forgotten man, the man in the background whose interests are never considered when some policy is framed to benefit a vocal interest group. The forgotten minority in situationist critiques are those whose actions and attitudes are never or barely considered in framing the critique. These are the individuals who manage to do the helpful or other virtuous action notwithstanding the absence of pleasant fragrances, the presence of a 85dB lawnmower, and so on. Most situationists ignore them as though they are irrelevant to the issue of the empirical adequacy of virtue ethics, even though they are the ones most likely to have the (nonglobal) virtues being tested. Be that as it may, since it is possible even for someone who is kind in most important contexts to sometimes fail to act kindly in those very contexts, perhaps these experiments can teach us something about the situational factors and human tendencies responsible for such failures.

Situationists argue that the influence of ambient smells and noises and moods on attention and action are problematic because they are morally irrelevant, yet "hugely and secretly influential" in how we think, feel, and act.[34] They are secretly influential because when asked why they helped, people are unlikely to cite the smell of baked bread, a good mood, or a bad conscience, showing that they have little introspective access to the actual causes of their actions. They are morally irrelevant because kind behavior is not evidence of kindness if prompted by the smell of freshly baked bread, a good mood, or a bad conscience. Unlike temptations or bad reasons, Mark Alfano argues, these factors give *no* reason at all, hence they cannot be accommodated by virtue ethicists' moral psychology.[35] Accordingly, they present a bigger challenge to virtue ethics than other situational features.

Is it, however, accurate to say that those who help do so because of the pleasant smells, positive moods, or bad conscience, *instead of* because someone needs help and helping them in that situation is appropriate? Clearly, the smells or positive moods or bad conscience play a causal role in the behavior of most subjects (but remember, again, not all). But why suppose that they *supplant* rather than *supplement* the standing reasons that they, like most people, have to help people in need, when helping them would not be a heavy burden or an obstacle to their own legitimate ends? Isn't it possible that the smells or the good mood merely make it easier to act on these standing reasons?

It might be countered that someone who is genuinely kind doesn't need an extraneous factor to make it easier to act kindly: the reason of kindness is sufficient. But if the reason that pleasant smells or good moods lead to more helping behavior is that they enable people to notice occasions for helping

and open them to new experiences, as many psychologists believe, then they are among the conditions that form the causal background of most people's perceptions and actions, along with good light, well-functioning senses, and the ability to match means to ends.[36] The fact that we were not aware until recently of how much effect they have on us does not change this fact.

Suppose, however, that their role in helping behavior is motivational: pleasant smells lead to a good mood, and a good mood causes the desire to help through a sense of optimism and the triggering of positive memories.[37] Is it accurate to say, as Christian Miller does, that kindly behavior is just a "causal byproduct" of these cognitive changes?[38] This seems like an odd view. If Eidos helps P because P needs help, she helps for the right reason, even if she would not have desired to help had her mood been bad or neutral. The fact that one part of the causal history of Eidos's desire to help P at time t does not involve P does not vitiate this fact. Indeed, having a stronger-than-usual desire to help when one is in a good mood is evidence of a benevolent outlook on life. It is a common experience that when we feel happy, we want to spread happiness around; we want others to share in our good fortune, for their sake. This reason, unfortunately, is never mentioned by critics. So long as a preexisting good mood is not a *sine qua non* of Eidos's kind acts, so long as she is sufficiently helpful for other people's own sake even when her mood is neutral or bad, her increased desire to help others for their sake when she's happy *is* evidence of her kind character.

Again, it may be true that sometimes Eidos helps others only because helping them enables her to maintain her good mood.[39] But before we conclude that this shows that she doesn't have the trait of kindness, even partially and imperfectly, we would have to know about her behavior and attitudes not only in other mood experiments, but also at home, in her neighborhood, and in her social life. Indeed, just expanding our horizons enough to take in the bystander experiments changes the picture. For in these, the vast majority of people help strangers when they are alone and the situation is not too dangerous.[40] Hence, the mere fact that most people are more likely to act in a kindly fashion when they are in a good mood than in a neutral or unhappy mood (or when they feel guilty than when they don't), does not show that they have no genuine kindness. The people who have no genuine kindness are those who don't care a fig about others when they are on top of the world, feeling invulnerable, or those who can maintain a good mood only by ignoring or harming others.

A common mistake of those who take these situational and mood studies to show that people have no genuine kindness is that they assume that kindness requires helping every person with a legitimate need whenever the cost of doing so is small.[41] What they overlook is that the *policy* of doing so *would* result in a heavy cost. Even if we never ventured outside, kindness understood thus would chain us to our desks for hours every day as we responded to worthy appeals for help by mail, e-mail, or telephone—even if each appeal cost only a fleeting minute or five. Kindness does not require

us to be forever-on-duty soldiers at the service of humanity, as some consequentialist theories would have it. One of the features that makes neo-Aristotelian virtue ethics psychologically realistic is that it recognizes that we have a right to live our own lives. Although some critics claim to recognize this, they seem to forget it when they infer from the literature just reviewed that most people aren't genuinely kind.

The objection that genuine kindness can't be prevented from operating by a noisy lawnmower or a depressed mood merits a similar response. Being distracted by the noise of a loud lawnmower and wanting to escape it, instead of noticing who needs help picking up books, doesn't seem any different from not noticing who needs help when one has a severe headache, or is absorbed in tending to one's toddler. Nor does any of this seem more momentous than not being able to concentrate on one's work in the presence of loud noises, severe headaches, or clamorous toddlers. Noisy lawnmowers that prevent kind acts simply prove, once again, that we are physical beings. As for depressed moods, these are unhappy moods in which nothing seems worth doing, either for others or for oneself. Hence, not being helpful when depressed doesn't seem like a moral failing—unless the depression itself is due to some moral failing, or unless one wallows in it. Even Kant acknowledges the difficulty of acting dutifully when we are unhappy—which is why, he thinks, we have an 'indirect' duty to be happy.

Situational non-reasons, then, leave virtue ethics unscathed. Once we become aware of them, however, virtue ethics requires us to welcome their positive influence and try to withstand their negative influence.

(ii)   Bystander Effects

Since 1968, when John Darley and Bibb Latané did their first study of the bystander effect, hundreds of experiments, both field and laboratory, have shown that, more often than not, as the number of bystanders in an emergency situation increases, the chances of anyone helping decreases.[42] This is especially so when the situation is dangerous or ambiguous.[43] Various psychological processes have been proposed as explanations:[44]

- 'I'm no more responsible than anyone else,' or *diffusion of responsibility*.
- 'What if I make a fool of myself?' or *evaluation apprehension/social influence*.
- 'No one else seems to think it's dangerous, so it isn't dangerous,' or *pluralistic ignorance/audience inhibition*. The more ambiguous the situation, the more often this happens, since each individual relies on the others for cues to disambiguate it, with the result that no one does anything, thus unintentionally signaling to the others that nothing needs doing. To these explanations I add one more of my own:
- 'Too many cooks spoil the broth,' or *I'll just make things worse*. My evidence that this is a common explanation for bystander effects is

that (i) it is a widely and explicitly known reason for not helping in many well-known everyday settings, including, of course, a crowded kitchen; (ii) before I read about the bystander effect, it was a factor in my own non-helping behavior in emergencies when there were other people closer to the event, and thus in a better position to help; and (iii) social psychology has taught me that I'm not unique.

Maria Merritt et al. cite the cognitive processes that explain the bystander effect as examples of largely automatic processes to which people have little or no introspective access.[45] This claim is based on Latané and Darley's report that not one of their subjects acknowledged that their failure to help had anything to do with the presence of other people—even after they were shown evidence for the connection.[46] This nonacknowledgment is extremely odd because the striking thing about the factors cited above is that they are highly intuitive. Moreover, only two of these factors are embarrassing to the subjects. One is the thought, 'I'm no more responsible than anyone else,' *if* the rest of this thought is, 'so no one can blame me, and that's what matters.' The other is, 'What if I make a fool of myself?' Both express the wrong priorities (caring more about being blamed by others, or seen as a fool by them, than about the individual who seems to be in need). In addition, the former shows indifference to the well-being of the individual in need and the latter shows cowardice. Latané and Darley do not tell us what reasons their subjects provided for not helping, but there seems to be no alternative to the list of possible explanations noted above that isn't (more) unflattering to the subjects. What, after all, could it be? 'I couldn't be bothered?' 'I had better things to do with my time?' 'I didn't want to get involved?' 'I like to see people suffer?' At any rate, it would be hasty to conclude from the fact that people can't always tell their reasons for their actions on the spot that they can't access them, period. Sometimes, knowing one's reasons takes time.

How should we evaluate the rational and moral status of the other possible explanations for not helping: thinking that the situation is not dangerous because no one else seems to think it's dangerous, and the desire to not be the cook who spoils the broth? The striking thing about these explanations is that although they make immediate sense, they stop short of being the logical conclusions to draw from the data. The logical conclusion in each case is: 'Others might be reasoning the same way and not helping for the same reason, so I better find out if my help is needed.' But whereas doing nothing to help someone in dire need at only a small cost to oneself because of a mistake in reasoning can't be kind, it is surely excusable. It would be inexcusable only if it was due to bad motives or traits. These include *akrasia* (we know we ought to help because we can without great risk to ourselves and are better positioned to help than the other people present, but we're getting late for that movie), bad principles (thinking that we have no responsibility to strangers), indifference to others' weal or woe, a generalized malice

(wanting others to suffer), deceiving ourselves into thinking that it's too risky (when it isn't) or ambiguous (when it's perfectly clear), or some other such rationalizing maneuver to get ourselves off the hook. And no one who already knows about the bystander effect can be an innocent bystander in a situation in which someone is in apparent need of help and there are others around who might or might not help.

In the next section, I turn to the most disturbing experiment in social psychology, an experiment that shows that most of us are highly limited in our virtues.

IV.   MORAL DISSOCIATION: OBEDIENCE IN STANLEY MILGRAM'S EXPERIMENTS

Before Milgram conducted his pilot experiment, he described the experiment to 110 people—faculty, students, psychiatrists, and others—and asked them how many would go all the way.[47] One in a thousand, said the psychiatrists.[48] The results shocked Milgram himself. In Experiments 2, 5, and 8, in which the experimenter is in the same room as the subject and the 'learner' (a confederate) can be heard but not seen, 65% of subjects proceed, unwillingly, even agonizingly, to shock a screaming, innocent individual, ostensibly to death, on the orders of the experimenter.[49] Milgram provides several explanations for this, including the absence of, and distance from, the learner,[50] the physical presence of the experimenter,[51] the status of the experimenter, and the loss of a sense of agency and responsibility. In several permutations on Experiment 2, Milgram changed one of the first three factors to test for its importance in explaining the subjects' obedience (Experiments 3, 4, 7, 14). Each of these conditions reduced obedience, but only in Experiment 14 did every subject stop as soon as the learner shouted at him to stop—for the learner was the experimenter himself, and the one who gave the orders was ostensibly a subject.[52] The exalted status of the experimenter outweighed the fact that he was physically absent from the room, whereas the relatively lowly status of the confederate outweighed the fact that he was present in the room. Yet when asked later, none of the subjects thought that the experimenter's status had anything to do with their behavior.[53] In Experiment 16, one experimenter took the role of learner again while another gave the orders. Here, the physical absence of one and the presence of the other broke the tie between their equal status: 65% continued to the end, just like in the baseline experiments.[54]

What should we conclude from these facts? That the vast majority of people can be led to behave egregiously when ordered to by someone they regard as a trustworthy authority figure, if he's looking over their shoulder.[55] According to Milgram, the presence of the experimenter acts like a physical force on the subject. A great deal of empirical work since Milgram published his studies reportedly supports his view that the experimenter's physical

presence and status were important factors in obedience.[56] Merritt et al. use this work as well as work on empathy and perspective-sharing to argue that subjects' other-oriented attention is misdirected by these features. An emotional response to someone in pain can lead either to an understanding of his point of view and empathy for him or to a self-focused personal distress and a desire to escape the situation.[57] Which it will become—empathy for the target or concern for oneself—depends on whether or not the subject (believes that she) shares the target's perspective, such as his values, interests, or background.[58] The fact that the subjects have volunteered to participate in an experiment makes them partial to the experimenter's perspective. Further, as the experimenter's explanation of the procedure repeatedly uses the words "punishment" and "learning," semantic priming may also be playing an unconscious part in influencing the subjects.[59] Consequently, the subjects' other-oriented attention is directed toward the experimenter instead of the learner. They seek to please the experimenter, who insists that the shocks are not dangerous and that the experiment requires them to continue. Their emotional response to the learner's screams fails to lead to empathy for him and remains a self-focused personal distress and a desire to escape the situation.[60] Hence, although most of the subjects find it unnerving to continue to shock the learner, they cannot stop so long as the individual with authority standing next to them doesn't "permit" it.

Merritt et al. hypothesize that (like priming) perspective-sharing is an automatic mediating cognitive factor that bypasses people's intentional control and leads to behavior that is incongruous with their reflectively endorsed values.[61] Hence, the highly integrated model of deliberation defended by Aristotelian theories must be modified.[62]

That a trusted authority figure has the power to define the situation for most people and get them to (more-or-less) accept his perspective and do his bidding against their will, without even threatening them with consequences, is an important insight. Once stated, however, we can see its roots in familiar facts of human life. Most of us started life with implicit trust in our parents and teachers, a trust that led us to obey them even when we couldn't understand why we should obey them. Most of us are also familiar with the fact that some people, especially authority figures, have an imposing presence, a presence that makes it hard to disobey or even contradict them. Also intuitive is the finding that when people feel that they have the same values (or background or interests) as another person, or when they are made to feel like valued participants in an important enterprise, as they are in the Milgram experiments, their sympathies and concerns get directed towards that person.

Other factors that every model of deliberation, whether Aristotelian or not, must take into account include our implicit egoistic and in-group biases, our tendency for self-deception, and the factor that Milgram himself regards as crucial in explaining the experimental results: the tendency for people in a subordinate position to see themselves as mere instruments of

the authority figure instead of as responsible agents.[63] (Surprisingly, this and self-deception rarely get discussed in the situationist literature.) But how exactly should we understand the influence of these tendencies or biases on us? Here are two possibilities suggested by Merritt et al.[64]

(i) These automatic influences on social cognition, emotion, and behavior spell the bed rock of human nature in the sense that no amount of virtue and phronesis, and no amount of self-awareness or self-control, can break through them and thwart their pernicious influence. They are the wild, untamable parts of human nature that can trip us at any time.

This claim is easily falsified by pointing to the 35%+ defiant subjects in Experiments 2, 5, and 8. Even more importantly, in Experiments 2 and 8, one subject stopped at 135 volts, before the learner demanded to be released, and nine stopped at 150 volts, when the learner made that demand. In Experiment 5, in which the learner claimed to have a slight heart condition, one subject stopped at 90 volts, and six stopped at 150 volts. These seventeen subjects stopped as soon as they had clear moral reason to stop. Even though the other defiant subjects were taken unawares to some extent, sixteen stopped before or at 300 volts, when the learner screamed that he was not going to answer any more questions.[65] Further, many defiant subjects gave the right reason for stopping: they were not going to hurt someone against his will just for the sake of an experiment. Had this value been uppermost in the obedient subjects' minds during the experiment, and had they trusted their own judgment more, they would not have obeyed either. In light of the behavior and attitudes of the defiant subjects in these experiments, there is no reason to think that *everyone* in Experiment 14, in which the experimenter becomes the learner, was motivated only or primarily by the status of the learner, or that *everyone* was ignorant of his own reasons for stopping.

(ii) These factors *can* take us unawares and lead us to misconstrue situations and to act in contrary-to-virtue ways, to a greater or lesser degree.

This is consistent with the facts, both experimental and nonexperimental. It has been shown repeatedly that automatic cognitive-affective processes (ACPs) are not entirely involuntary or unconscious, nor controlled processes entirely intended or fully conscious.[66] This explains how seventeen subjects could stop at or before 150 volts in Experiments 2, 5, and 8. Although we can't know for *certain* that they would have defied a trusted religious or political authority's seemingly heinous orders, we have excellent reason to think that they would have. To a greater degree than anyone else, they based their decisions on the objective evidence instead of the experimenter's judgment. Whatever tendencies they had to obey the experimenter, seeing themselves as his instruments, with no responsibility for their actions, whatever

tendencies they had to agree with his perspective thanks to the experimental set-up and the semantic priming in his explanation of the procedure, they managed to do the right thing at the very first opportunity, refusing to be budged by the experimenter's repeated prods to continue. The ACPs that led most subjects astray in the experiments or that delayed the right response in the other defiant subjects had *no* influence on these few. Perhaps they were just having a good day because they had run across and taken to heart Chico Marx's "Who you gonna believe, me or your own eyes?" But isn't it more likely that their actions were due to their (nonglobal) virtues of justice, courage, compassion, independent-mindedness, and integrity?

If character can render certain automatic tendencies powerless or weak, it can also render them powerful. In postexperimental interviews with the subjects in Experiment 5, the experimenter asked what level of shock subjects would be willing to accept for themselves.[67] Only seven were willing to accept 450 volts. Milgram found similar or worse results in other postexperimental interviews when this question was asked.[68] These subjects clearly did not act on the principle that "such as that is for me, so it is for him, nothing less."[69] The 'learner' became a mere shadow-figure. Some of the obedient subjects also formed a kind of alliance with the experimenter against the learner, whom they saw as stupid or stubborn. And some asked during the experiment who would be held responsible if something happened to the learner. After being assured that only the experimenter would be held responsible, they continued with the experiment. During interviews after Experiments 1–4, many of the obedient subjects stated that they did what they did only because they were "helpless," thanks to the experimenter's commands. When the experimenter asked how much responsibility they assigned to themselves, to the experimenter, and to the learner, the fifty-seven obedient subjects taken as a group assigned only 36% responsibility to themselves, 38% responsibility to the experimenter, and 25% to the learner.[70] By contrast, the sixty-one defiant subjects taken as a group accepted the lion's share of the responsibility. Perhaps the obedient subjects were just having a bad day. But isn't it possible that the actions of at least some of them were due to akrasia, or pusillanimity[71] and a weak sense of justice and compassion?

Evidence from other experiments also shows that character makes a difference to the influence of situational factors and the ACPs they trigger. Thus, several experiments have shown that subjects who have the chronically accessible trait concepts, honest and dishonest, are able to reliably distinguish between honest and dishonest behaviors under conditions that do not allow conscious processing of the behaviors.[72] In other words, those who take honesty and dishonesty seriously enough to habitually evaluate behaviors in these terms have the discernment to perceive and correctly evaluate behaviors as honest or dishonest—even when their attention is fully absorbed by a different task.[73] Other research has shown that people who have internalized norms of fairness and equality have lower levels of

prejudice, both implicit and explicit,[74] and that greater understanding of the target of prejudice, greater awareness of their own implicit prejudice, and more positive emotions towards the target of prejudice can reduce people's implicit prejudice.

To a virtue ethicist, all this is welcome confirmation of one of her basic premises: through practice, virtue can become second nature, and to the extent that it does, we can respond appropriately to a variety of situations in the dynamic, complex world in which we live. Human nature in itself is neither good nor bad. Our innate biases and susceptibilities to various situational factors are simply part of the structure of our cognitive and affective machinery. It's what we—and our upbringing and moral luck—make of them that is good or bad. Thus, our egoism can be manifested primarily in not wanting to be blamed for harming others—or primarily in not wanting to harm others and failing in our own eyes. Our in-group biases can be manifested in denigrating members of out-groups just because they are members of out-groups—or simply in building and maintaining friendships and community with those who share our values. Unusual situations pose subtle and unexpected deliberative and motivational challenges. It is in these situations that, contrary to the advice of some situationists, we most need to remind ourselves of our values. If, for all our care, we think that certain situations will test our limits, phronesis tells us to either avoid those situations in favor of situations more conducive to virtue—or seek help to tie us to the proverbial mast. No amount of care, however, and no theory, can guarantee perfect virtue, because self-knowledge, knowledge of the world, and emotional and deliberative habituation is always imperfect.

NOTES

1. Thanks to Nancy Snow and Franco Trivigno for their comments on an earlier draft.
2. Aristotle, *Nicomachean Ethics,* ed. Terence Irwin (Indianapolis, IN: Hackett Publishing Company, 1999), 1100b34–1101a1. All references to Aristotle's *Nicomachean Ethics* are to the Irwin translation.
3. Aristotle, *Nicomachean Ethics,* 1105a35.
4. Aristotle, *Nicomachean Ethics,* 1100b13–15.
5. See Julia Annas, *The Morality of Happiness* (New York: Oxford University Press, 1995); Annas, *Intelligent Virtue* (New York: Oxford University Press, 2011); and Daniel C. Russell, *Practical Intelligence and the Virtues* (Oxford: Clarendon Press, 2009).
6. Shane Drefcinski, "Aristotle's Fallible *Phronimos,*" *Ancient Philosophy* 16 (1996): 139–154.
7. Howard Curzer in "How Good People Do Bad Things: Aristotle on the Misdeeds of the Virtuous," *Oxford Studies in Ancient Philosophy* 28 (2005): 233–256.
8. Maria W. Merritt, John M. Doris, Gilbert Harman, and The Moral Psychology Research Group, "Character," in *The Moral Psychology Handbook,* ed. John Doris and The Moral Psychology Research Group (New York: Oxford

University Press, 2010), 355–401; Mark Alfano, *Character as Moral Fiction* (New York: Cambridge University Press, 2013); Christian B. Miller, *Moral Character: An Empirical Theory* (Oxford: Oxford University Press, 2013).

9. Aristotle, *Nicomachean Ethics,* 1145a 1–3.
10. Peter Geach, *The Virtues* (Cambridge: Cambridge University Press, 1977), 161–162; J. L. Ackrill, *Aristotle the Philosopher* (Oxford: Oxford University Press, 1981), 137; Neera K. Badhwar, "The Limited Unity of Virtue," *Nous* 30, no. 3 (1996): 306–329.
11. Badhwar, "Limited Unity."
12. See, for example, John M. Doris, *Lack of Character: Personality and Moral Behavior* (New York: Cambridge University Press, 2002); Badhwar "The Milgram Experiments, Learned Helplessness, and Character Traits," *The Journal of Ethics* 13, no. 2/3 (2009): 257–289; Alfano, *Moral Fiction,* and Miller, *Moral Character*.
13. Aristotle, *Nicomachean Ethics,* 1115a 30–32, 1115b 4–6.
14. Terence H. Irwin, "Disunity in the Aristotelian Virtues," *Oxford Studies in Ancient Philosophy* Supplementary (1988): 62, n. 2.
15. Badhwar, "Limited Unity."
16. Robert M. Adams, *A Theory of Virtue: Excellence in Being for the Good;* Badhwar, "The Milgram Experiments"; Badhwar, *Well-being: Happiness in a Worthwhile Life* (Oxford: Oxford University Press, ~~my book will be published around the same time as this book.~~ 2014).
17. Doris, however, doesn't find this solution satisfactory~~,~~ *Lack of Character,* 71–75.
18. Christine Swanton, *Virtue Ethics: A Pluralistic View* (New York: Oxford University Press, 2003); Russell, *Practical Intelligence*.
19. Aristotle, *Nicomachean Ethics,* 1109b 14–26.
20. Ibid., 1120b 3–4, 1121a 1–4.
21. Ibid., 1121a 15.
22. Ibid., 1125b 1–~~1~~1126a 3.
23. Ibid., 1110a 25–26.
24. Ibid., 1110a 5–8.
25. Ibid., 1110a 20–25.
26. Ibid., 1134a16–23; see Drefcinski, "Aristotle's Fallible Phronimos," 151–153.
27. For example, Alfano, *Moral Fiction,* 31–32.
28. Badhwar, "The Milgram Experiments," 263.
29. Aristotle, *Politics,* in *The Complete Works of Aristotle: The Revised Oxford Translation*, ed. Jonathan Barnes (Princeton, NJ: Princeton University Press, 1984), 1287a31–32.
30. Merritt et al., "Character," 370.
31. Merritt et al., "Character," 358; Alfano, *Moral Fiction,* 43–46; Miller, *Moral Character*.
32. Doris, *Lack of Character,* 30–32; Alfano, *Moral Fiction,* 46–50; Miller, *Moral Character,* chapters 3, 4, 6.
33. Alfano, *Moral Fiction,* 46.
34. Ibid., 44.
35. Ibid., 43–44.
36. Ibid., 44–50; Miller, *Moral Character,* 69–70.
37. Miller, *Moral Character*, 69–73.
38. Ibid.
39. Ibid., 68–69.
40. Bibb Latané and Steve A. Nida, "Ten Years of Research on Group Size and Helping," *Psychological Bulletin* 89, no. 2 (1981): 308–324.

41. Doris, *Lack of Character,* 31–32; Alfano, *Character,* 36; Miller, *Moral Character,* 75.
42. Latané and Nida, "Ten Years of Research."
43. Peter Fischer et al., "The Bystander-Effect: A Meta-Analytic Review on Bystander Intervention in Dangerous and Non-Dangerous Emergencies," *Psychological Bulletin* 137, no. 4 (2011): 517–537.
44. Bibb Latané and John Darley, *The Unresponsive Bystander: Why Doesn't He Help?* (New York: Apple-Century Crofts, 1970); Latané and Nida, "Ten Years of Research."
45. Merritt et al., "Character."
46. Latané and Darley, *Unresponsive Bystander,* 124.
47. Stanley Milgram, *Obedience to Authority: An Experimental View* (New York; Perennial, 2009), 27–31.
48. Ibid., 31.
49. Ibid., 34–35, 55–57, 61–63.
50. Ibid., chap. 4.
51. Ibid., 59–63.
52. Ibid., 95, 99–104.
53. Ibid., 103–104.
54. Ibid., 95, 107–110.
55. Ibid., 59–62.
56. See Merritt et al., "Character," 382–383.
57. C. Daniel Batson et al., "Influence of Self-Reported Distress and Empathy on Egoistic versus Altruistic Motivation to Help," *Journal of Personality and Social Psychology* 45, no. 3 (1983): 706–718. Here I follow Batson et al. in distinguishing empathy from personal distress, whereas Merritt et al. distinguish sympathy from personal distress and use 'empathy' to cover both.
58. Merritt et al., "Character," 381.
59. Ibid., 384.
60. Ibid., 384.
61. Ibid., 381–383.
62. Ibid., 379.
63. Milgram, *Obedience,* 145–147.
64. Merritt et al., "Character," 356, 371, 388.
65. The number of subjects who stop before or at 150 volts in Experiment 3, in which the learner is present in the same room as the subject, goes up to eleven, and in Experiment 4, in which the subject is required to forcibly place the learner's hand on the plate starting at 150 volts, to seventeen. I find these numbers to be less significant than the numbers in 2, 5, and 8, because the proximity of the learner to the subject makes obedience viscerally harder without providing the subject with any new reasons.
66. John A. Bargh, "The Ecology of Automaticity: Toward Establishing the Conditions Needed to Produce Automatic Processing Effects," *American Journal of Psychology* 105, no. 2 (1992): 182–183. Although Merritt et al. acknowledge this point in the abstract ("Character," 373–374), they seem to forget it when they warn about the pervasive influence of automatic cognitive processes.
67. Milgram, *Obedience,* 57.
68. Ibid., 57.
69. Josiah Royce, *The Religious Aspect of Philosophy: A Critique of the Bases of Conduct and of Faith* (Chicago: Houghton Mifflin, 1885; Google e-book), 55–59.
70. Milgram, *Obedience,* 203.
71. Aristotle, *Nicomachean Ethics,* 1123b 10–12ff.

72. Bargh, "Ecology of Automaticity," 189.
73. For a related argument, see Nancy E. Snow, *Virtue as Social Intelligence: An Empirically Grounded Theory* (New York: Routledge, 2010), chap. 2.
74. Laurie A. Rudman, Richard D. Ashmore, and Melvin L. Gary, " 'Unlearning' Automatic Biases: The Malleability of Implicit Prejudice and Stereotypes," *Journal of Personality and Social Psychology* 81, no. 5 (2001): 856–868, 861–862, 865–866.

REFERENCES

Ackrill, J. L. *Aristotle the Philosopher.* Oxford: Oxford University Press, 1981.
Adams, Robert Merrihew. *A Theory of Virtue: Excellence in Being for the Good.* Oxford: Oxford University Press, 2006.
Alfano, Mark. *Character as Moral Fiction.* New York: Cambridge University Press, 2013.
Annas, Julia. *The Morality of Happiness.* New York: Oxford University Press, 1995.
———. *Intelligent Virtue.* New York: Oxford University Press, 2011.
Aristotle. *Politics* in *The Complete Works of Aristotle: The Revised Oxford Translation.* Edited by Jonathan Barnes. Princeton, NJ: Princeton University Press, 1984.
———. *Nicomachean Ethics.* Edited by Terence Irwin. Indianapolis, IN: Hackett Publishing Company, 1999.
Badhwar, Neera K. "The Limited Unity of Virtue." *Nous* 30, no. 3 (1996): 306–329.
———. "The Milgram Experiments, Learned Helplessness, and Character Traits." *The Journal of Ethics* 13, no. 2/3 (2009): 257–289.
———. *Well-being: Happiness in a Worthwhile Life.* Oxford: Oxford University Press, 2014.
Bargh, John A. "The Ecology of Automaticity: Toward Establishing the Conditions Needed to Produce Automatic Processing Effects." *American Journal of Psychology* 105, no. 2 (1992): 181–199.
Batson, C. Daniel, Alice Isen, Karen O'Quinn, Jim Fulz, and Mary Vanderplas. "Influence of Self-Reported Distress and Empathy on Egoistic versus Altruistic Motivation to Help." *Journal of Personality and Social Psychology* 45, no. 3 (1983): 706–718. doi: 10.1037/0022-3514.45.3.706.
Curzer, Howard. "How Good People Do Bad Things: Aristotle on the Misdeeds of the Virtuous." *Oxford Studies in Ancient Philosophy* 28 (2005): 233–256.
Doris, John M. *Lack of Character: Personality and Moral Behavior.* New York: Cambridge University Press, 2002.
Drefcinski, Shane. "Aristotle's Fallible *Phronimos.*" *Ancient Philosophy* 16 (1996): 139–154.
Fischer, Peter, Joachim I. Krueger, Tobias Greitemeyer, Claudia Vogrincic, Andreas Kastenmüller, Dieter Frey, Moritz Heene, Magdalena Wicher, and Martina Kainbacher. "The Bystander-Effect: A Meta-Analytic Review on Bystander Intervention in Dangerous and Non-Dangerous Emergencies." *Psychological Bulletin* 137, no. 4 (2011): 517–537. doi: 10.1037/a0023304.
Geach, Peter. *The Virtues.* Cambridge: Cambridge University Press, 1977.
Hursthouse, Rosalind. *On Virtue Ethics.* Oxford: Oxford University Press, 1999.
Irwin, Terence H. "Disunity in the Aristotelian Virtues." *Oxford Studies in Ancient Philosophy* Supplementary Vol. (1988): 61–78.
Latané, Bibb and John Darley. *The Unresponsive Bystander: Why Doesn't He Help?* New York: Apple-Century Crofts, 1970.
Latané, Bibb and Steve Nida. "Ten Years of Research on Group Size and Helping." *Psychological Bulletin* 89, no. 2 (1981): 308–324. doi: 10.1037/0033-2909.89.2.308.

Merritt, Maria, John Doris, and Gilbert Harman. "Character." In *The Moral Psychology Handbook*. Edited by John Doris and The Moral Psychology Research Group, 355–401. New York: Oxford University Press, 2010.

Milgram, Stanley. *Obedience to Authority: An Experimental View.* New York: Perennial, 2009.

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

Royce, Josiah. *The Religious Aspect of Philosophy: A Critique of the Bases of Conduct and of Faith.* Chicago: Houghton Mifflin, 1885; Google e-book.

Rudman, Laurie A., Richard D. Ashmore, and Melvin L. Gary. "'Unlearning' Automatic Biases: The Malleability of Implicit Prejudice and Stereotypes." *Journal of Personality and Social Psychology* 81, no. 5 (2001): 856–868. doi: 10.1037//0022-3514.81.5.856.

Russell, Daniel C. *Practical Intelligence and the Virtues.* Oxford: Clarendon Press, 2009.

Snow, Nancy E. *Virtue as Social Intelligence: An Empirically Grounded Theory.* New York: Routledge, 2010.

Swanton, Christine. *Virtue Ethics: A Pluralistic View.* New York: Oxford University Press, 2003.