*PLEASE DO NOT QUOTE WITHOUT PERMISSION*

**Becoming Good: Narrow Dispositions and the Stability of Virtue**

Rachana Kamtekar

1. Introduction

Aristotle frequently reminds the audience of the Nicomachean Ethics that the goal of their inquiry is not to know what virtue is but to become good (ii.2 1103b28, cf. 1095a5); knowing what goodness is is valuable only because it gives us a target at which to aim (1094a23). I begin with this reminder to point out both a difference between Aristotle’s and a contemporary virtue ethical interest in virtue, and a similarity between Aristotle’s and any ordinary person’s interest in it. First the difference: Aristotle faulted his predecessors—the sophists and Plato—for thinking that knowledge of some (roughly) theoretical kind was sufficient for virtue; he instead emphasized that the good condition of our appetites and emotions, necessary for us to desire and feel as we should, was essential to virtue. Contemporary virtue ethicists, by contrast, value the idea of virtue as a condition in which rational judgments as to what is the right thing to do harmonize with a desire to do it and pleasure in doing it, because it promises the good person a psychologically healthy life in which there is no neuroticism-inducing gap between justification and motivation. But, second, the similarity: one doesn’t have to be a virtue ethicist or even a virtue theorist to want to be good—that is why Aristotle can assume it as a goal he shares with his audience, and that is why we are in crisis after the last century’s findings by experimental psychologists about the inputs to our behaviour of which we are not and perhaps cannot be aware. Forget about being good and harmonizing your desires and feelings of pleasure and pain with your rational judgments, these findings seem to suggest, you can’t even get your rational judgments to determine your behaviour, because most of the time you are on automatic pilot. Most of your actions are the output of automatic processes that operate below the level of consciousness, and these automatic processes are rigid and specialized responses to certain triggers in the environment, perhaps selected for in our evolutionary history because adaptive overall, but for all that, hardly sensitive to all the sorts of considerations required for delivering morally appropriate action. And there is nothing you can do about this, because rational processing is too slow and costly to be on-line all the time.

Responses divide into what I’ll call the ‘Aristotelian’ and the ‘Nudge’ responses. The Aristotelians propose that our automatic processes can be educated by just the sorts of processes Aristotle envisioned for the habituation of our appetites and emotions in the development of virtue. The Nudges rejoin that the cognitive limitations of these automatic processes make such an education impossible and call instead for the development of policies and institutions that acknowledge these limitations and frame our choices and actions so that we end up acting as we would, on conscious reflection, like to (‘opt-out’ rather than ‘opt-in’ policies for organ donation, a ban on candle-lit dinners with other people’s spouses, flies painted on urinals, and so on). Although the Nudges are relatively pessimistic about the prospect of developing virtue—on their view, the best moral agent’s reflective judgments are somewhat alienated from her in-the-moment motivations, her in-the-moment motivations need social and environmental backing, and perhaps she is neurotic rather than at peace with herself—their interest to bring it about that *our* actions are ones we would approve (rather than merely that good outcomes obtain) still retains something of the hope of becoming good.

A concrete example will help to sharpen these alternative responses. Research shows that knowledge of a student’s race or ethnicity makes a difference to many instructors’ assessment of their performance on ‘subjective’ grading, even when the instructors’ explicit attitudes are not prejudiced. In light of this research, should you as an instructor decide to add a certain number of points to the score of students belonging to that race or ethnicity? (Blind grading may not be an option; even if you suppress students’ names, you may soon recognize their styles.) Suppose that you thereby end up treating the students more fairly—scoring them according to their accomplishment—even though your actions are alienated from your in-the-moment and case-by-case judgment (so you deviate from some of your moral ideals in order to achieve others). Now we can all agree that it would be even better if you had no tendency to deviate from assessment based on performance because of a student’s race or ethnicity—you would be better, and you would act fairly even in cases where you hadn’t formulated a corrective policy. But, the Nudges say, you do not have such virtue, and you cannot achieve it, so you should settle for the policy, for it at least enables you to act in conformity with the fairness that you value. If the Nudges are right about our abilities, then we should do as they say, given that our practical interest is in whatever capacities an agent has and can develop to make it the case that her actions and feelings conform to the values she avows—if not dispositions to judge, feel, etc. as one should in the moment of action, as the Aristotelian would have it, then to act on principle and resist conflicting inclinations, as a Kantian would recommend, or to engineer the circumstances for action so that we are not likely to act contrary to our values, as some of our contemporaries propose. Of course these are capacities of very different sorts: those that are exercised in the moment of action (‘quick’ or ‘hot’ capacities such as perception, an automatic judgment of right or wrong, inclination) and those that are exercised more slowly in a cool hour (self-knowledge, calculation and reasoning for the formulation of policies to be activated in the moment of action); moral knowledge and psychological knowledge (about oneself in particular but also about people in general); whole-hearted attitudes as well as tactics for self-manipulation; and so on. Given that our desire to be good entails a desire for conformity between our avowed values and our actions, we ought to be open to suggestions about what kinds of capacities we should develop, and how to develop them, whether by imitation and practice, imagination, reasoning from principles, consistency reasoning, self-manipulation, etc., in order to bring about this conformity.

Some philosophers have conceived of the psychological condition in which reason sets policies and then manages the sub-rational elements by force or manipulation as a relationship between a human policy-maker dealing with a non-human, or animal, set of motivations. But Aristotle suggests a kinder picture: think of the policy-setter as a father and the automatic dispositions as children. Fathers do not only lay down the rules of the household based on what is right and ensure compliance by force; they base their rules on a compassionate understanding of what is easy or difficult for their children. We do this with ourselves when we adjust our expectations of ourselves, promising ourselves a reward for undertaking some unappetizing task, or setting the clock fast to correct for our tendency to squeeze in too many things just before the moment of departure. Perhaps we will grow out of needing the reward or the clock manipulation, perhaps not.

Instead of taking sides on how educable or uneducable are the automatic processes or non-reasoning motivations that generate most of our actions, in this paper I comment on three points that seem to me to have received insufficient attention in the contemporary discussion. First (section 2), I argue that our interest in becoming good (as good as possible) should be divorced from an interest in being better than others, or better than the norm, and that if it is, then we need to be careful about the bearing of some experimental psychology on the discussion, for much of that literature assumes that virtue is something distinctive. Instead of conceiving of virtue as distinguishing some people from others, we should conceive of it as correcting certain human tendencies that are destructive of social and personal well-being. Second (section 3), I examine the idea that virtue can be developed from situation-specific dispositions through thoughtful generalization and integration with other such dispositions, and I discuss some difficulties for and alternatives to this proposal. Third (section 4), I suggest that rather than seek a general recipe for the development of desirable dispositions, we focus on the sources of stability or instability for these dispositions one by one.

2. Virtue need not be distinctive

 Both historically and in recent discussions, virtue has been assumed to be distinctive. Ordinarily, we might call someone ‘brave’ on the basis of actions that stand out from the norm insofar as they involve taking a great risk or facing a great danger without exhibiting much fear. Classical Greek and Roman attributions of virtue also foreground being outstanding (with respect to situations of danger, temptation, etc.); indeed, it is precisely for this reason that Socrates rejects Meno’s definition of virtue, ‘desiring good things and being able to acquire them’: the desire, he says, is common to all (Meno 77b-78b).[[1]](#footnote-1) In a different context, Aristotle uses what most people can resist or endure as the standard against which to measure weakness of will with respect to pleasures and pain (EN VII.7.1150a10-b5). And in our time, the father of personality psychology Walter Mischel echoes this conception of character when he defines it as ‘a construct or abstraction to account for enduring behavioral *consistencies and differences’* (1968, 5-6; my emphasis). Mischel’s definition makes it clear that what social and personality psychologists study is behavioural *variation*, and in particular, they study what behaviour that departs from the norm *covaries with*.

In more recent work, Mischel and colleagues argue that although most of us converge on the norm for many traits, a more situation-specifying look at our behaviour reveals distinctive *patterns of variation* and departures from the norm.[[2]](#footnote-2) Bolek may behave more aggressively than the norm in situation 1 (when criticized by an adult, say), less aggressively than the norm in situation 2 (when teased by a peer), and with average aggression in situations 3, 4, and 5, but although his aggression across these situations is average, he will in the future reliably behave with above-average aggression when criticized by an adult, with below-average aggression when teased by a peer, and so on. By contrast, Lolek may also be average in aggression across situations but distinctive and stable in his below-average aggression in response to adult criticism and above-average aggression to peer harrassment. Bolek’s greater-than-normal aggression in the adult-reprimand situation and less-than-normal aggression in the peer-teasing situation constitute a ‘signature’ (and Lolek has a different signature). How is this different from attributing to Bolek and Lolek conjunctions of narrow dispositions? The constituents of character, at least according to Mischel and the philosophers who follow him,[[3]](#footnote-3) are cognitive and affective items; the path from patterns of behavioural variation to their causes in goals, priorities, etc. would go something like this: Bolek has the ‘if-then’ signature, ‘if reprimanded by an adult, argue back’; ‘if teased by a peer, grin and bear it’. We might then characterize Bolek’s character as follows: he has the goal of standing up to authority and being liked by his peers; he prioritizes these goals over others, such as not being in trouble and enjoying high status in his peer group; he construes adult reprimand as an opportunity to stand up for himself and peers’ teasing as goodnatured . . .

Even Adams (2006), who insists that ‘virtue is not a competitive sport’ (123) supposes that the elements out of which the virtues are built, narrow dispositions, are displayed in behaviour more often than the average (122-25). While Adams uses the probabilistic qualification in order not to require disposition-manifesting behaviour on every occasion (correcting an expectation that is especially implausible for virtues of imperfect obligation), building ‘more often than the average’ into the requirement for behavioural manifestation builds in distinctiveness.

I object to building distinctiveness into the very definition of virtue, rather than allowing that whether a virtue is common or rare, the norm or distinctive, is a contingent matter, depending on things like which dispositions are supported by the environment, education, etc. The reason is that it makes the fact of whether or not you possess a disposition depend on whether or not others do, rather than only on how you act and feel, in the relevant situations.[[4]](#footnote-4) Sabini and Silver make a related criticism of determining the presence or absence of virtue on the basis of behavioural correlations:

Suppose there were a situation in which it was very difficult to behave in an honest way . . . extreme torture, threats to the lives of one’s children-- . . . where the temptation to lie was high. Now let us imagine that in this circumstance 20 percent of the population did tell the truth . . . [N]ow imagine another situation in which very few people would be seduced into lying. So the 20 percent who told the truth in the first situation once again tell the truth, but an additional 60 percent of the population would, too. What is the correlation in honesty between those situations? .25. However, if 90 percent tell the truth in the second circumstance the correlation between it and the first would be .176.[[5]](#footnote-5) (2005, 543)

Of course scientists must investigate purported causal claims empirically via correlations, but we must then think about what those correlations tell us. Sabini and Silver’s example makes it clear that a low correlation between honesty in these two situations is no evidence of inconsistency in those who are honest (according to situation 1), but rather of inconsistency in those who are dishonest (in situation 1); and there is no reason to suppose that a dishonest person is going to be dishonest whenever possible. Dishonesty is rather a lack of commitment to honesty, evidenced when there is some motivation to be dishonest (I’ll come back to this characterization of vice as lack in section 4). Studies of honesty also show that in e.g. test-taking situations, there is a correlation between IQ and not cheating; thinking about this correlation, Hartshorne and May rightly ask, why would someone cheat if they knew the answers?

 It might be objected that distinctiveness is more central to virtue than I have claimed, for virtue and vice terms are terms of praise and blame, and there is no cause for praising or blaming people for what is common.[[6]](#footnote-6) A non-moral analogue might help: we do not praise literate people in societies which have achieved near 100% literacy, even though we recognize literacy as a very good quality to have. However, even here praise is not irrelevant: we congratulate children on their progress in reading, even though we do not continue to congratulate them as adults. And such praise to the children is not just manipulative; it is an acknowledgement of their effort in learning to read, and then reading more difficult texts, and of a genuine accomplishments. So we may praise people for an improvement relative to their former state, even though we would not go on praising them for the achievement once it is a stable state. Further, virtue seems to merit praise insofar as it is up to us, but whether we are better than others is not entirely up to us—whereas whether we are better than we were seems to be much more so. Finally, we may want to be wary of our ordinary use of praise. We do often use praise in judgments that compare people to one another, and insofar as how you stand in relation to others is neither up to you nor a reliable measure of the goodness of the qualities you actually possess, we should not worry if our theory of virtue departs from our intuitions about praise. Indeed, to borrow Aristotle’s line about happiness (EN I.12), it may be that virtue, insofar as it makes us good, ought to be prized rather than praised.

I conclude that insofar as our interest in virtue derives from our desire to be good, live well, or do the right thing, we should not require it to be distinctive. Our interest in distinctiveness should go only so far as behavioural co-variance is a means of empirically confirming or disconfirming causal efficacy. For I am no less honest if everyone else refrains from lying, and I am no more courageous if it turns out that most people are easily dissuaded from doing what they believe they should by fear.

Someone might object that the virtues are not only moral ideals but also, along with the vices and other dispositions, bases for explaining and predicting behaviour, and if they are to be such bases, their causal efficacy cannot but be understood in terms of their contribution to behavioural variation. It is true that we appeal to character traits to explain remarkable actions (whether remarkably good or bad or remarkable just because unusual), especially when our explanatory interest is combined with an evaluative one and we wish to praise or blame. However, when our interest is purely explanatory, and when we know agents well, especially if we do not disapprove of their actions, we do *not* usually explain their actions by appealing to their character-traits; instead, we explain their actions by reconstructing their reasons.[[7]](#footnote-7) And explanations of actions in terms of agents’ reasons need not be restricted to behaviours that depart from the norm. So it would seem that philosophers and psychologists who think that the basis of behaviour lies in goals, priorities etc. should not be attached to thinking of virtues and vices as distinctive dispositions.

3. Developing Narrow or Embedded Dispositions into Virtues

It seems to be a robust finding of social psychology that while people do not in general have broad-based dispositions such as the virtues are supposed to be—that is, dispositions to behave consistently with e.g. honesty across a range of different sorts of situations (respecting property, refraining from cheating, telling the truth)—they do have stable narrow dispositions—dispositions to behave in the same way in situations that are very similar, such as honesty in test-taking. The best evidence for this claim comes from the famous studies on honesty among schoolchildren conducted by Hartshorne and May in the 1920’s, according which while the correlation between not lying, not cheating on tests, not cheating on homework and not stealing was only .227, the correlation between not cheating on tests on different occasions was .721.[[8]](#footnote-8)

What can we say about these stable narrow dispositions? Nisbett and Ross claim that the stability is due to the similar situational contributions to behaviour across instances.[[9]](#footnote-9) Harman judges that stable narrow dispositions ‘do not count’ as character traits, on the grounds that they are indistinguishable from strategies developed for one situation and repeated .[[10]](#footnote-10) Doris concludes that while local traits are descriptively adequate and capable of predicting behaviour, they are too fragmented to be virtues, for the possessor of sailing-in-rough-weather-with-friends-courage might also be the possessor of sailing-in-rough-weather-with-strangers-cowardice.[[11]](#footnote-11) Russell 2008[[12]](#footnote-12) argues that the investigation of trait-situation-behaviour connections in subjective or psychological or idiographic terms turns up character-traits on Mischel’s Cognitive-Affective Processing Systems (CAPS), which are broad dispositions.

 In the last few years number of philosophers have proposed that the virtues can be built up out of narrow dispositions; here I summarize three such proposals. Adams 2006[[13]](#footnote-13) ‘take[s] it as a working hypothesis’ that: (1) we first acquire ‘modular’ behavioural dispositions—dispositions that are independent of one another and are domain specific—that may also be brought under the scrutiny and control of practical thinking, and (2) these ‘modular’ dispositions can be added together to form a more inclusive composite disposition to behave consistently across a wide variety of situations (125-27). For example, physical courage and moral courage, which deal, respectively, with physical and social dangers, can emerge independently of one another and of e.g. justice (p. 183), but they must be integrated with one another and temper and be tempered by the other virtues in order to add up to Virtue (206FF), which Adams defines as an admirable pro-attitude towards the good. We can presumably extend this account to narrower dispositions still: ‘courage-in-rough-sailing-weather-with-friends’ can emerge independently of other dispositions, but can become part of Virtue as a whole by being broadened and evaluatively integrated with other admirable pro-attitudes towards the good. Beginning with Mischel’s CAPS traits, Snow 2010[[14]](#footnote-14) sketches a way that local traits may generalize to become global by the assignment of similar meanings to objectively different situations, a response that is rooted in the trait itself. So for example, a person becomes irritable by interpreting not only an offhand remark by store clerk as rude, but also a casual remark by an acquaintance at lunch, and so on (pp. 31-33). We can, however, take control of such processes; so the irritable person who wants to change can punish herself for expressions of irritability, thereby activating the ‘Behavioural Inhibition System’ (Snow models this on work documenting self-correction against prejudice[[15]](#footnote-15)). However, this is not a general account of the development of character, only a proof of the possibility of exercising control over automatic processes in order to bring actions in line with avowed values. According to Annas 2011[[16]](#footnote-16), who likens the acquisition of virtue to the acquisition of a practical skill where the learner seeks to understand what is to be done and why, one begins to learn the virtues in embedded contexts, for example, one learns to be generous in the context of buying friends presents or in the context of buiding a house for the homeless by following the instruction or example of a role model, and then, because one aspires to understand the virtues and to be virtuous oneself, and with the help of explanations from one’s teachers, one learns to identify what makes a virtuous action virtuous (21-23). ‘A child, for example, is told to be honest in contexts of sharing treats and toys without cheating, not taking things in shops without paying for them, and similar circumstances. . . When he learns that his parents are scrupulous about honesty with money, but lax about honesty with the truth, this will produce confusion about what honesty is, and a need to clarify it. He may refine his conception of honesty or conclude that honesty takes ultimately diverse forms.’ (36-37).

Abstracting from the differences in these accounts of virtue acquisition, it seems that they recognize at least two distinct processes: (1) the acquisition of narrow dispositions/ virtues embedded in contexts, and (2) reflection (e.g. generalization from a narrow disposition to respond in a certain way in a narrowly circumscribed situation to others like it, or registering ethically salient differences between superficially similar situations) or integration (of independently-developed narrow dispositions, in the case of congruence by noticing what they have in common, and in the case of conflict by tempering or revising priorities). I would like to raise some questions about how to characterize (1) and (2).

First, while the narrow dispositions of (1) may be or become automatic, is there an in-principle cognitive limit on them? The skill analogy tells both ways: on the one hand, thinking of all the information processed by a virtuouso pianist, and also the way he uses his technique to express his thoughts and feelings (cf. Annas 11), one might think, the limit, if it exists, is probably very high; on the other hand, since the virtuoso’s skill is the counterpart to virtue rather than to a narrow disposition, perhaps the narrow dispositions out of which it is composed themselves could be quite limited—the great cognitive reaches of the skill come from their integration.

Second, how important is ethical evaluation in stage (1) of the development of virtue? Adams says that the first stage of virtue acquisition is the relatively effortless acquisition of ethical concepts by way of a linguistic practice which requires only the low-level agreement and functioning most societies possess (213-14).[[17]](#footnote-17) (Determining what true virtue is and correcting, tempering and integrating one’s modular dispositions to acquire true virtue is, he acknowledges, much more difficult, but Adams identifies some facilitators: principled reasoning to yield greater conformity between one’s ethical judgments and one’s conduct; aesthetic and other experiences of juxtaposed competing values to yield integration in desire and feeling; the right sorts of social institutions, e.g. a loving family, a just state, an institution dedicated to persistent ethical discourse in the context of care for whole persons.) Annas seems to accord the virtue terms some motivational force in the early acquisition of the virtues: ‘A boy will learn to be brave, initially, by seeing a parent chase off a dog, say, and *registering that this is brave*.’ (22, my emphasis)

These questions raise my third question about (1): when we develop a narrow disposition or a virtue in an embedded question, just what are we learning, and how are we learning it? The question has a motivational as well as a cognitive aspect. On the cognitive side, we must be learning something more general than ‘in this [very] situation, do this [very] action,’ because we only act in this [very] situation once. So we must be acquiring some sense of a type of situation and type of action to do in that situation. What type is that? If we think about cognition alone, it seems more likely that we will acquire the sense of a specific type (e.g. ‘return found wallet’) than a more general one (e.g. ‘do the just thing’). But if we think about motivation, e.g., if a child receives a reward for acting in a certain way he may interpret the action as a means to the reward,[[18]](#footnote-18) and either acquire the narrower disposition to return the found wallet when there’s a reward or reinforce the broader disposition to do what promises a reward.

It may be objected that this account of virtue acquisition is too intellectualist even in stage (1), let alone (2) (Annas anticipates that her account of virtue acquisition may seem too intellectualist, 28ff.). But even very early imitative behaviour—imitation, whether of a parent or teacher or other adult, being one important part of early virtue-acquisition, so of (1)—is oriented towards learning, long before the advent of reflection, reason-giving, and articulate accounts.

The last fourty years of research on imitation shows it to be an interpretive, knowledge-seeking, activity from a very early age. The work of Andrew Meltzoff shows that even infant mimicry—the imitation of tongue protrusions and mouth opening, for example—involves the infant mapping the model’s body and face onto her own. Babies and toddlers follow the gaze of caretakers and interactive peers as well as of computer-simulated blobs who react to them.[[19]](#footnote-19) Alison Gopnik reports that babies pay closer attention to novel means of causing events and imitate these: Gergely and his colleagues (2002) had a caretaker turn on a light switch with her head in the presence of 14-18 month olds and the children paid little attention when her hands were full or wrapped in a blanket, but stared and tried to copy her when her hands were free. Carpenter (2002) observed that by 14 months children will not simply imitate behaviour when they see it is a means to an end but will instead do whatever is required to achieve that end. So if children observe an adult trying to open a box—pulling at the lid, twisting the handle, etc.—they do not repeat these actions but simply open the box, whereas if they see the adult only twisting the handle repeatedly, they imitate the adult and themselves twist the handle repeatedly. Gopnik argues that these behaviours show children to be ‘lay scientists’, developing a causal understanding of the world.[[20]](#footnote-20)

Children imitate not only in order to find out about (what other people know about) the world; they also do so in order to learn about other people’s attitudes—beliefs, desires, feelings. Maringer, Krumhuber, Fischer and Niedenthal (2011)[[21]](#footnote-21) discovered that people interpret smiles (fake or genuine? friendly, pleased or superior?) by imitating them and interpreting their own imitation smiles. When subjects had to hold a pen in their mouth, inhibiting the movement of their own facial muscles, they could no longer ‘read’ others’ smiles and so had to rely on their beliefs about the typical meaning of smiles in situations such as theirs. Children mirror the facial expressions and brain activity of distressed others and thereby learn how others feel—e.g. becoming distressed themselves around age 1, then trying to comfort the distressed others around age 2.[[22]](#footnote-22)

This research shows that we engage with actions under a description from a very young age, in order to find out about the world and other people. It also shows that at least for this purpose and at a very young age, we only have to see our models as enough like us for the imitation to to be worth our while; for this purpose, even an interactive blob is worth imitating and learning from. It also shows that children will not just imitate mindlessly unless they reflect consciously about what to imitate, whether doing this conflicts with doing that, or not doing this conflicts with having thought it was the thing to do in some similar past situation. The acquisition of narrow dispositions looks like it should be an intelligent, discriminating process that is responsive to our standing motivations—although unlikely to bring concerns about consistency with it.

Aristotle is a source of rich, testable, speculation on the relationship between how actions engage us motivationally and our acquisition of a disposition to perform actions of that type. According to his account of habituation, we first do acts that are such as the virtuous person would do, e.g. on instruction (by a teacher, by the law), and by doing such acts we develop a corresponding state of character: we not only do virtuous acts, but choose them and choose them for themselves. In ‘Aristotle on Learning to Be Good’, Burnyeat[[23]](#footnote-23) explains that the pleasure of performing a virtuous action can be known only from performing it (like the pleasure of skiing can only be known when one skis), and so the agent can only choose virtuous actions for themselves—for the fine actions they are—once she has, through experience—come to see their fineness. Notice that this Aristotelian account is neutral with respect to the broadness or narrowness of the disposition to be acquired. Perhaps Aristotle thinks that ‘just’ (as opposed to only ‘return found wallet’) is a learnable type because just and unjust action have a characteristic motive—fairness vs. greed in the distribution of the goods of fortune. If Aristotle is right, the disposition a child develops through acts of doling out the kindergarten snack fairly might not be restricted to doling out kindergarten snacks, but might extend right away to cases in which a motive to prefer herself in a distribution is to be overcome--whether or not adults use the terms ‘fair’ and ‘unfair’ in the situation, and whether or not she reflects on what she is doing, but just because of the way it feels. In that case, the action-type for which she is developing a disposition is not ‘doling out snacks’, but rather, ‘not being selfish’.

For (2), I again have three concerns. First, the appropriate transition from a narrow disposition to a broad and integrated virtue seems to require that we both see situations in new domains as relevantly like[[24]](#footnote-24) the situations in which we developed our narrow dispositions and respond with actions that are appropriate to those new domains—not simply repeating the sorts of things we did in the original domain, because it is not action types (facing danger, making a donation) that are virtuous (notoriously, rushing into danger to save someone else may be courageous or foolhardy; a large donation may be generous or ostentatious), but actions when they are appropriate to the situation. This is difficult because the morally significant similarities and differences between two situations may not be obvious at all and because non-moral similarities and differences may make the morally relevant action-possibilities salient or invisible.

Second, what is the correct scope of generalization from narrow disposition to virtue? In the case where you first learn to be generous in buying a gift for your schoolfriend, how and to whom ought you extend your generosity? Universalizing your narrow disposition seems unsustainable (you will run out of time, money, energy, or whatever else you are generous with; further, you will likely be a social liability, making others share the burdens of your generous impulses), but restricting your generosity to your friends and family risks activating an ‘us’ vs. ‘them’ or ‘in-group’ vs. ‘out-group’ attitude (which comes easily to us but likely conflicts with virtue).

 Third, the world we live in constantly serves up situations that seem to demand a virtue-involving response. It is not just the person collapsed outside the train station one night on your walk home, but the many homeless individuals you see every day, or if you don’t see them, a couple of clicks on the internet and they are there before you, and a few more clicks enable you to send them help via charitable organizations already working for them. If you do not think donations are the right response, you can volunteer time—working either to help those individuals directly or to alter the social and economic system that results in their being in this bad position. If you think only they can help themselves, well, then, there is no shortage of other causes that a virtuous person would respond to: climate change, for one, or cruelty to non-human animals. But even people who devote their whole lives to some kind of virtuous activity ignore some demands of some virtues; perhaps they must do so in order to be effective, or to sustain their activity in accordance with virtue. For example, many people who specialize in being ‘green’, who put considerable resources into setting up solar panels, biking to work, and recycling, seem not to respond to world hunger; many people who rescue animals are quite unforgiving of people. The phenomenon of moral bookkeeping—e.g., after a big recycling run, I let myself turn the air-conditioning down to 76—lately documented as yet another of our species’ moral foibles, may reflect some awareness of our own limited moral energies and an effort to sustain our commitment to what we think most important. This is a less than excellent way to be, but one might thereby avoid even worse ways of being.

Here is a example, occasioned by discussion of the bystander effect. The audiences of bystander behaviour often express shock or outrage at the inaction or unwillingness to get involved on the part of bystanders. There’s a man lying slumped on the ground at Victoria and no-one stops to ask if he needs help—unless he’s wearing a suit and tie. But of course the risk of taking on problems you can’t handle is much lower if all that’s wrong with someone is that he’s fainted on a busy streetcorner. You can intervene, make a phone call, hail a cab, and he’ll be on his way, and out of your life. If he has no-one to whom you can make a phone call, on the other hand, and no money to pay the cab, and maybe needs someone to check on him at the hospital, you’ll have taken on more than you can handle. And worse, instead of not stopping when so many others are doing the same, you’ll be uniquely bad because you know how desperate things are for him, and despite his great and clear need, you’ll be walking away. I’m not saying that bystanders think of all this when they fail to help someone in clear need, but they may be responding to a lot of implicit information about types of intervention and possible consequences, which the person who intervenes has to discount. It is not clear to me whether the attitudes I am reporting are morally defective, or psychologically healthy, or both. There is something odd about thinking of virtue as a disposition to respond to certain morally relevant features of the world, and then, even though the world is quite bad, expecting the domain-specific virtues to be capable of integration with one another (here, for the record, Adams rejects but Annas embraces a unity of virtues thesis).

This case and the preceding discussion of difficulties in generalizing and integrating are intended to make vivid the conflicts between our desires to be good, live well, and do the right thing and to raise the question whether virtue can deliver all these. If it can, might it do so by lowering our standards for right action? Imagine Masha, who, having become a social worker because she believes the work is important, and because that’s what her mother did, finds herself depressed. She finally quits and turns to photography, which gives her great joy. Should she now just accept that she is not as other-regarding as her mother and enjoy her life as a photographer (perhaps donating money to social work), or should she make herself continue at social work in the hope that her other-regarding feelings will develop with experience? Virtue as the morally appropriate response to the world seems to be in tension with virtue as the developed state of an individual human being’s capacities. (Adams’ ‘virtue is persisting excellence in being for the good’ seems to avoid this tension because it allows other ways of being for the good than bringing it about, and because it understands excellence as admirability, which can attach to the less-than-the-best.)

4. Virtue and stability

The previous section considered the development of virtues from narrow dispositions for the reason that narrow, but not broad, dispositions are experimentally confirmed contributors to behaviour. Sometimes this point is put in terms of stability—narrow, but not broad, dispositions are stable—and sometimes in terms of stability over time vs. consistency across situations—our dispositions are stable over time but not consistent across situations. But the question is worth asking, why should we think that the same thing holds for all dispositions, no matter what their content? And here there is an interesting difference between the contemporary and ancient discussions, for the ancients claim that it is virtue that is stable, not dispositions in general, and they offer reasons why virtue in particular should be stable.[[25]](#footnote-25) It’s worth looking at their reasons, because if we are interested in the development of virtue, we might want to know if there are different ways to approach different virtues depending on what their reinforcers are.

1. Virtue makes its possessor happy. It is and feels good to be courageous, prudent, just and temperate, because such a person is untroubled by unsatisfied desires and in the possession of what is truly valuable. Acquiring and maintaining that good condition requires the performance of courageous, prudent and just acts. So the virtuous person receives, from within himself, positive reinforcement for acting on virtuous dispositions, which is a source of stability for those dispositions.
2. Virtue is based on knowledge, and knowledge is stable (because systematic, consisting of mutually supporting grasps of the truth; because infallible when exercised in judgment). In the absence of knowledge of what is good it might appear to me best to write philosophy papers one day and work with refugee teens the next day, with the result that I end up doing neither well. On the other hand, if I know what’s best, I won’t be buffeted about by appearances; rather, an accurate assessment of what’s best will instead guide my actions.

Both these reasons return us to the question Socrates asked Protagoras: is virtue one? The prospect of ‘broadening’ narrow dispositions and ‘tempering’ them with one another rather than remaining ‘fragmented’ looks much more promising if there isn’t inevitable conflict between the goods we rightly pursue (e.g. autonomy and well-being).

The virtues likely have two other sources of stability that other traits may not:

1. In a reasonably functional society, at least some virtuous behaviours are rewarded with trust and increased cooperation or even praise. So, for instance, honest behaviour ordinarily earns one a reputation for honesty and encourages others to cooperate with one, so one has a reason to behave honestly even in cases where dishonesty would be more profitable. This is not to say that the virtuous person has these thoughts: the virtuous person might just benefit from being, and signalling that she is, honest by blushing when telling a white lie, for example (slightly modified from Frank 1988, Passions Within Reason). However, this is not the case in dysfunctional societies. Barbara Demick’s Nothing to envy describes how in the North Korean famine at the beginning of this century orphaned children died like flies unless they were ready to steal (as one observer put it, ‘The good die first’), and how adults, in order to eat, had to give up on their ideal of working as service to their country and instead learn to sell whatever people would buy.
2. We have reasons of self-respect to commit ourselves to the virtues that we don’t for other traits: I can respect myself whether I’m talkative or quiet, but I can’t if I’m a cheat.[[26]](#footnote-26) Now commitment to honesty is different from avowing the value of honesty, for even sincere avowals of value can be accompanied by shallow commitment. Avowals may be made on the basis of values that structure subjects’ lives, in the sense that they are engaged in an ongoing effort to live up to a given value, but also in response to reminders of moral standards that they don’t think much about. In the latter case, a countervailing value (e.g. avoidance of embarrassment, as suggested in Sabini and Silver 2005) or the absence of a strategy for specifying the action called for given a long-term goal or broad value (‘implementation intention’) can easily result in action that conflicts with values one avows. But the subsequent regret and guilt people feel when they see how bad the action was can give them reason to commit to the values they avow.[[27]](#footnote-27)

Claims (1) –(4) in particular may be true or false, but consideration of the sources of stability for virtuous dispositions is absent from the current discussion about the stability of character-traits (the alternatives about which there is discussion, ‘situation’ and ‘character’, are obviously too broad-brush). On (3), note that not everything that is reinforced (especially socially) is a virtue: e.g. the desire to please is likely socially reinforced but it is not a virtue. And there are traits that may well be virtues that are not socially reinforced: I suspect that unrestricted compassion and generosity do not receive much social reinforcement: their demands are enormous and they lead us to feel guilty for not meeting them so we often avoid thinking about them and collude with one another in doing so; they generate socially awkward situations (I recall a philosopher objecting to Peter Singer’s plastic belt on the grounds that everyone else was wearing leather; we’ve all seen the perfect social event marred by the kind-hearted soul who brings along the one person who wasn’t invited exactly because he would spoil everyone’s fun). The uneven social reinforcement for helping behaviours may be yet another explanation for why dispositions to help are especially subject to situational effects (as in the Milgram and Isen and Levin studies). There may also be dispositions that are hard-wired, like the fear of embarrassment or the Big Five personality traits. And there may be dispositions that have a self-reinforcing structure, like grit. My proposal is just that we think more about *why* the dispositions we are interested in cultivating may be stable or not, rather than hoping to build the virtues out of narrow dispositions in general on the grounds that the psychologists tell us *that* they are (as a whole) relatively stable.

1. So I need to take back my 2004 claim that distinctiveness is not a feature of virtue in traditional virtue ethics (‘Situationism and virtue ethics on the content of our character,’ p. 468). No doubt ancient philosophers would have denied that one’s possession of virtue depends on others’ non-possession, but so (I think) would most contemporary philosophers, when this is spelled out. The unquestioned assumption of the ancients of significant moral inequalities among people may have blocked consideration of whether distinctiveness was a necessary or contingent feature of the virtues. Even in a society designed to produce virtue in citizens, like that described in Plato’s and Aristotle’s political writings, citizen virtue was understood contrastively (e.g. guardians vs. producers, citizens/free men vs. slaves). [↑](#footnote-ref-1)
2. Shoda, Y., Mischel, W., and Wright, J. 1994. ‘Intraindividual stability in the organization and patterning of behavior: Incorporating Psychological situations into the idiographic analysis of personality’. Journal of Personality and Social Psychology vol. 67, pp. 674-87. Mischel, W. and Shoda, Y. 1995. ‘A cognitive-affective system theory of personality: Reconceptualizaing situations, dispositions, dynamics, and invariance in personality structure’. Psychological Review vol. 102, pp. 246-68. [↑](#footnote-ref-2)
3. e.g. Russell (2009, Practical Intelligence and the Virtues, 260-61) and Snow (2010, Virtue as Social Intelligence, 9), Railton (2011), ‘Two cheers for virtue; or, might virtue be habit-forming?’ in M. Timmons (ed.) Oxford Studies in Normative Ethics vol. 1, 295ff. Doris recognizes that whether virtue is rare or common ought to be a contingent matter (2002, Lack of Character, 177n.23). [↑](#footnote-ref-3)
4. There may be some such dependence, if the dispositions in question are dispositions to cooperate vs. free-ride on cooperators, and cooperative behaviours would cease if too many free-rode, but even this would allow for quite a bit of variation in the distribution of virtues and vices. [↑](#footnote-ref-4)
5. J. Sabini and M. Silver, ‘Lack of Character? Situationism Critiqued’ Ethics vol. 115 pp. 535-562 [↑](#footnote-ref-5)
6. Thanks to Philip Reed for pressing this point. [↑](#footnote-ref-6)
7. Knobe, J. and Malle, B. (2000), ‘Self and other in explanations of behavior: 30 years later’ in Psychologica Belgica 42, pp. 113-30. Their formulation replaces the Nisbett and Jones’ idea that first-person explanation tends to be situational and third-person dispositional—which is not far off, but misses the connection between knowledge/access to a person’s thoughts and reasons-explanations. [↑](#footnote-ref-7)
8. Hartshorne and May (1928), Studies in the Nature of Character, vol. 1, The Nature of Deceit, pp. 122-125 [↑](#footnote-ref-8)
9. L. Ross and R.E. Nisbett (1991), The Person and the Situation [↑](#footnote-ref-9)
10. G. Harman (2000), ‘Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error’ pp. 165-78 in Explaining Value and Other Essays in Moral Philosophy, at 167 and 175. [↑](#footnote-ref-10)
11. J. Doris (2002), Lack of Character, pp. 62ff. In ‘Weighing globality, empirical content and empirical adequacy’ (ms.), Federico Lauria and Alain Pé-Curto argue that narrow dispositions ought to be subjected to situational variation just as broad dispositions are, and that if it emerges that hyper-narrow dispositions are better confirmed by the data than are merely narrow dispositions, these hyper-narrow dispositions ought to be subjected to situational variation, and so on. Their point is that the desiderata for a psychological explanans include both empirical adequacy and empirical content, and the narrowing of dispositions yields adequacy at the expense of content. This does not seem to be an unwelcome result for any situationist, for why should a situationist be attached to the existence and causal efficacy of dispositions of any particular breadth? For the record, Doris 2002 is not troubled by the possibility that situations alone, rather than local traits, account for behavioural reliability (p. 64). [↑](#footnote-ref-11)
12. In my 2004 I noted that Nisbett and Ross explain the correlation between situation and behaviour in terms of subjects’ construal of the situation rather than the situation as it is objectively (p. 470) and that they advocate an ‘idiographic’ conception of character according to which rather than starting with a given trait and expecting each person to have that trait to some degree or other, the starting point should be the individual, and consistency should be expected for the traits valued by that individual (p. 476). Because the idiographic conception begins with what the individual values, I thought it common ground between some traditional conceptions of character and contemporary social psychological conceptions. Russell goes much further, arguing that psychology confirms the existence of broad-based character-traits when consistency is sought between the subject’s point of view and his behaviour. [↑](#footnote-ref-12)
13. Robert Adams (2006) A Theory of Virtue : Excellence in Being *For* the Good [↑](#footnote-ref-13)
14. Nancy Snow (2010) Virtue as Social Intelligence [↑](#footnote-ref-14)
15. Devine and Montieth (1999) ‘Automaticity and Control in Stereotyping’ in Chaiken and Trope (eds.) Dual-Process Theories in Social Psychology (Erlbaum), 185-210, cited in Snow. [↑](#footnote-ref-15)
16. Julia Annas (2011) Intelligent Virtue. [↑](#footnote-ref-16)
17. [↑](#footnote-ref-17)
18. Lepper, M., D. Green, and R. Nisbett. 1973. ‘Undermining Children’s Intrinsic Interest with Extrinsic Reward: The Overjustification Hypothesis’ *Journal of Personality and Social Psychology* 28: 129-137. [↑](#footnote-ref-18)
19. Meltzoff (2013) ‘Origins of Social Cognition: Bi-directional self-other mapping and the ‘Like me’ hypothesis’ in Banaji and Gelman (eds.) Navigating the Social World: What infants, children, and other species can tell us (Oxford), 139-44. [↑](#footnote-ref-19)
20. Gopnik (2009) The Philosophical Baby [↑](#footnote-ref-20)
21. Emotion, vol. 11, pp. 181-87 [↑](#footnote-ref-21)
22. Harris (1989) Children and Emotion: The development of psychological understanding. [↑](#footnote-ref-22)
23. Burnyeat (1980) ‘Aristotle on Learning to Be Good’ in Rorty (ed.) Essays on Aristotle’s Ethics, University of California Press, pp. 69-92 [↑](#footnote-ref-23)
24. I realize that this is vague. However, understanding ‘like’ in the sense of calling for the exercise of the same virtue/virtuous response prevents situational features from justifying a given virtue-involving response. The best I can do is to say ‘like’ pertains to considerations: somebody is in danger; I am able to do something; etc. [↑](#footnote-ref-24)
25. The question might be asked: isn’t vice also stable? If the question is about the consistency of behaviour in very similar situations over time (‘stability’ on the situationist stipulation) or about the consistency of behaviour in different but still virtue-relevant situations of different kinds, then I think the answer is ‘no’. The virtuous person chooses virtuous actions for themselves—for their own sake, or for what makes them the virtuous actions they are. The vicious person does not choose vicious actions for themselves, but rather for their promise of bringing him some good he wants. The vicious person will not cheat you if cheating you offers him no prospect of gain. If, on the other hand, the question is about the stability of the relationship between reason on the one hand and appetite or passion on the other, the vicious person’s reason, Aristotle says, approves his appetitive and passionate desires (unlike the continent and incontinent persons), which is why he can be said to choose his vicious actions (unlike the person who acts through anger without any involvement from reason, and so does not choose his action although he acts willingly NE v.8). [↑](#footnote-ref-25)
26. But see Batson et al. 1999 ‘Appearing moral to oneself without being so’, Journal of Personality and Social Psychology vol. 77, 525-37 and even more, Ariely (2012), The (Honest) Truth about Dishonesty (HarperCollins), both of whom attest the widespread ability to maintain our moral self-concept while availing ourselves of the advantages of immoral action—Batson conceives of this as lowering one’s ethical standard to accord with one’s self-serving behaviour, Ariely as refraining from maximally self-interested immoral action in order to count oneself motivated by moral concerns. [↑](#footnote-ref-26)
27. (1) – (4) are systemic sources of stability, but there may be idiosyncratic ones as well: if it ‘works’ for a child to confess that he’s broken the window, in the sense that the consequences are less harsh than if he tried to conceal the fact and was found out, he will be inclined to be open about other bad outcomes for which he was responsible, as compared to if he sees he’s also being punished as the bearer of bad news. [↑](#footnote-ref-27)