Merritt, Melissa

Kant on Reflection and Virtue

Melissa Merritt, *Kant on Reflection and Virtue*. Cambridge: Cambridge University Press, 2018, 234 pp., US$99.99, 9781108424714

Reviewed by David Sussman

In *Kant on Reflection and Virtue*, Melissa Merritt challenges a common understanding of Kant as promoting a highly self-conscious attitude toward the world, as both an object of possible knowledge and as an arena for action. On this view, which Merritt associates with people such as Christine Korsgaard, rational beings like ourselves must always be prepared to "step back" from their own mental life, so as to reflect the content and grounds of their beliefs, desires, intentions, etc., and to consider whether those grounds really support the attitudes based on them. Reflection, on such views, is essentially a matter of having thoughts about our own thinking, in a highly articulate and determinately conceptualized way. And as Merritt notes, there does seem to be some basis for this view in what Kant says. Kant does claim that all judgment is necessarily self-conscious or reflective, yet also that such judgment needs to be brought to an even higher degree of reflection according to the maxims of a "healthy understanding." In his moral philosophy, Kant seems to insist that a good person will always be trying to articulate the maxims upon which they are acting, so as to be able to test those maxims according to the Categorical Imperative. The resulting picture of virtue is strangely intellectualistic and narcissistic; the good person seems to be thinking, not so much about the situations she faces, and the needs, rights, and interests of other people, but about her own motives and intentions. Many commentators have recoiled from this picture, noting how such an unremittingly self-reflective life, even if psychologically possible, would leave little room for any kind of spontaneity, let alone any direct emotional engagement with the world or other people.

Merritt argues that these interpretations of Kant misunderstand what he takes reflection to be in both theoretical and practical contexts. Despite the naturalness of the reading, Kantian reflection is not really a matter of "stepping back" from our thinking, forming determinate thoughts about it, and rationally considering its justification. Instead, reflection is a kind of sensitivity or

responsiveness to rational norms which may be largely tacit, and so something revealed in the way we think, feel, and act, but not anything that requires explicit thoughts about our thoughts, let alone any kind of introspective observations of our own mental life. For Merritt, a healthy understanding does involve a kind of cognitive virtue, but that virtue need not involve any self-conscious applications of rules or concepts. She similarly argues that moral virtue need not involve any special attention to our own intentions and motives, or any explicit deliberation or maxim-testing. Instead, moral virtue is itself an element of a "healthy human understanding;" that is, such virtue is, although a matter of practical reasoning, essentially a kind of cognitive skill.

I will leave discussion of the connection between reflection and theoretical thought to the other commentators; my concerns center on the way that Merritt applies this general picture to Kant's understanding of moral virtue. I want to challenge what Merritt calls the "specification thesis": that moral virtue is a special instance of cognitive virtue (or "healthy understanding" in general). I also take issue with her further claim that such moral virtue is best understood as a kind of skill (even if only in the special sense of "free skill" that Merritt employs). However, let me be clear that I think that there is considerable truth in both of these claims, which serve as important correctives to many unfortunate caricatures of Kant. My criticisms are really only to the effect that these claims are overstated; that is, while there are indeed important cognitive dimensions to virtue, there are equally significant (non-cognitive) affective and volitional dimensions as well (at least insofar as it still makes sense to contrast the cognitive with the volitional in the first place). Similarly, I believe that Merritt is right that moral virtue is, in part, a matter of a kind of normative discernment and appreciation, rather than just a sort of continence or self-command. And I agree that such practical discernment does not require any sort of highly self-conscious thoughts let alone theoretical abstractions. However, I argue that there has to be more to Kantian virtue than this, in part because the problems we face in moral life are not merely matters of appreciating subtle moral distinctions in particular, concrete situations. I'll argue that Kant does not see all the challenges to morality to come from blurry vision alone. Rather, the deepest dangers come from perfectly clear illusions, illusions that result not from the complexity of the moral world, but from ourselves, and our ineradicable propensity to rationalization and self-deception. I will argue that it is these latter tendencies that require us to go beyond the virtues of discernment and skill that Merritt describes. In addition to these capacities, we will need a profound kind of self-knowledge that will indeed require, if not much by way of introspective discernment, at least fairly intellectually sophisticated

powers of self-interpretation. In essence, my objection is that Merritt's understanding of moral virtue is too Aristotelian, and insufficiently Christian. Her conception of such virtue might be enough for uncorrupted creatures, but not for the inescapably fallen beings such as ourselves.

In making sense of Kant's claims that all judgment is, and should be, reflective, Merritt draws a distinction between "constitutive reflection" ("reflection-c"), and "normative reflection" ("reflection-n"). Neither sort of reflection involves the kind of stepping-back or self-theorizing that, for ease of reference, I'll call "self-reflection." Supposedly, it is constitutive reflection that must accompany all our judgment, and indeed, all our sensible experience as well. Such reflection involves the way that the notorious "I think" "must be able to accompany all my representations." As Merritt interprets it, such constitutive reflection fundamentally involves implicitly seeing oneself as the source if a distinctive point of view on the world. In contrast, normative reflection involves a kind of concern with and taking responsibility for one's own "cognitive agency," as governed by the three maxims of the "healthy understanding": 1) Always think for yourself; 2) Always think from the point of view of others, and 3) Always think consistently (such that thinking for oneself coheres with thinking from the point of view of others). As Merritt understands it, none of these activities need involve any abstractions or explicit rule-following.

Merritt contends that a healthy understanding is to be found as much in the moral virtues as it is in the theoretical ones. As she sees it, to have a moral virtue is not just a matter of being determined to apply some abstract moral principles to one's life. Rather, moral virtue is a matter of a kind of cognitive sensitivity to how basic moral concerns apply to the particular features of concrete cases. This sort of fine-grained appreciation need not involve (or even entail) any ability to clearly conceptualize and articulate what is morally important in a specific case in any particularly illuminating way (that is, in a way that would be helpful to someone who didn't already share that virtue). The virtuous person is not just someone who is always formulating and testing her maxims and abiding by whatever the results are. Instead, she is a person who has a vivid kind of appreciation of the central moral value (essentially, of just what a person is), and to be able to see how this value is at play in the various features of the particular circumstance she is facing. For Merritt, this ability is essentially cognitive, although not in a sense that is supposed to contrast with the volitional or the affective. The virtuous person has a particularly rich understanding of morality, but this understanding takes the form not of propositional knowledge, but structures of

feeling and motivation. Such appreciation or attunement is cognitive in the sense that having an ear for music might be, or having a sense of humor, or a feel for the strength of a position in chess. This is a disposition by which we can do what morality requires of us; and while morality can demand a great many different things, being able to craft theories or explain oneself in abstractions is very low on the list.

As Merritt understands it, such cognitive/volitional/affective dispositions count for Kant as a kind of skill *(Fertigkeit*). This may sound surprising, but Merritt explains what Kant has in mind is what he calls a "free skill," as opposed to something more like an Aristotelian *techne*. A *techne* is such that it could, in principle, be used for any sort of end or from any sort of motive. A characteristic feature of skill in this sense is that it can be intentionally misused; the *techne* of medicine can just as readily employed to kill or torture as to heal or comfort. The "free skill" of moral virtue, on the other hand, is bound to a particular kind of end and a particular kind of motive. The virtuous person does not merely have a kind of dexterity in doing the right thing; her doing so is continuously informed by a sense of its moral importance that follows from a deep concern for it. Presumably, even if a virtuous person did try to use her moral skill for a morally bad end, she would tend to do a worse job of it than someone who lacked that virtue (when honest people are compelled to lie, they usually do so ineptly. e.g., James Mattis' remarks about recent troop deployments to the U.S./Mexico border).

I think everything Merritt has said so far is exactly right, so that virtue does indeed involve a cognitive element that cannot be understood in terms of either articulating or applying theories. However, I think that moral virtue also has distinctive volitional elements, at least insofar as it makes sense to still talk of any contrast between the cognitive and the volitional at all. When discussing moral virtue, Kant repeatedly describes it as a kind of "strength" or "fortitude" with respect to our commitment to act morally. He tells us that virtue is a matter of having a moral resolve that is powerful enough to overcome whatever obstacles that inclination (or really, we ourselves in response to inclination) puts in our path. This feature of virtue doesn't sound very cognitive; ordinarily, it would seem that one can perfectly understand why something is wrong (drinking too much at a party, committing adultery) and still have little resistance to temptation. Conversely, people who are able to overcome such temptations often don't seem to have an especially deep understanding of their wrongness, at least not deeper than more weak-willed folk.

Many ordinary, uneducated Germans managed to stand against Hitler, unlike the supposedly greatest philosopher of the 20th Century.

For Merritt, this familiar thought rests on the equation of cognition with theoretical sophistication. It is certainly true that true virtue does not require us to intellectually articulate any abstract philosophical systems. However, virtue does involve being able to attend to (and care about) what is truly important, where this is expressed in doing (rather than saying) the right things. Although Kant does insist that virtue is a matter of strength of resolve, such strength is itself an aspect (or consequence?) of clarity of vision.

This is a very appealing response, but I think it is neither true nor Kant's position. This is not to deny that having a richer and more fine-grained appreciation of the moral features of a case might indeed (and even typically) increase the strength of our moral resolve. However, I don't think this need be so; our resolve can soften or harden without any change in what we know, and our moral understanding can become richer or poorer without any corresponding change in our practical commitments. Such divergence might not be possible in a perfectly rational being, but it is in us. Consider a non-moral case: it's around midnight, and I know if I finish off the last of the pizza, I will, in about four hours, suffer terrible heartburn. I enjoy pizza, but nowhere near enough to compensate for such pain. Yet I cannot resist the siren song of the pizza; I eat it, in perfectly vivid apprehension not just of how I will be suffering later on, but how I will be ruing the choice that I am now making. I've done this a fair number of times; there's definitely some failure of determination or self-command here, but it's not that I don't fully appreciate what's going on. After all, you'd expect that after a few instances of this happening, I would finally catch wise. And indeed I have; but I keep knowingly doing this idiotic thing anyway.

I see no reason to redescribe such cases (that seem all too common) as ones where, on some level, I'm still not grasping something: at least, no reason that's prior to our commitment to the philosophical thesis in question. And indeed, Kant seems to recognize the possibility of such "clear-eyed" weakness of will, which he describes as "frailty," as one of the grades of our ineradicable "propensity to evil," which he finds:

> "expressed even in the complaint of an Apostle: "What I would, that I do not!" i.e., I incorporate the good (the law) into the maxim of my power of choice; but this good, which is an irresistible incentive objectively or ideally (*in thesi*), is

> subjectively (*in hypothesi*) the weaker (in comparison with inclination) whenever the maxim is to be followed." (*Religion within the Boundaries of Mere Reason,* 6:29).

Kant does not here suggest that there is any kind of confusion or unclarity at work; the good has indeed been taken up into my maxim without any apparent defect, and Kant does not suggest that I then suffer any difficulty in seeing how it applies in a particular case. The failing here lies not with my understanding, or my vision, but with *me* as an agent (i.e., with my will). Admittedly, Kant does not give any further explanation of how such frailty is possible for creatures like us, and so one may be tempted think that there must be some kind of cognitive failure behind it. But this is a mistake; unlike the more intellectualist views of Plato or Aristotle, Kant recognizes that there are primitive liabilities of the will, liabilities that are distinct from any defect of rational apprehension. Admittedly, the will must always operate in light of basic rational norms, as recognize by *Wille.* But the basic power so informed, *Willkür*, may or may not fully hold itself to those norms. Here it is simply up to us whether we act well or not, not to some feature of what we know or can see. (Were this not so, we could not freely do wrong, and so could never be morally culpable.) Of course, even if we allow that virtue involves an irreducibly volitional element (in terms of strength of resolve), this does not entail that virtue involves anything like fancy self-reflection, insofar as this would involve sussing out one's own motives and testing them by any sorts of abstract rules or principles. Moral virtue might then remain a kind of reflective (but not self-reflective) skill, even if it would not be purely cognitive one.

However, Kant seems to think that true virtue requires a high degree of self-knowledge, where this is a kind of conceptually sophisticated understanding of one's own motives and intentions. Merritt notes that Kant considers the "first command of all duties to oneself" to be the Delphic injunction to know oneself. However, she argues that such self-knowledge is not a matter of having any explicit beliefs about one's own mental states. Rather, a person knows herself in the relevant sense by trying to know the world; that this, by taking an interest in her own "cognitive agency" and thereby trying to figure out how things really are, as guided by the maxims of the healthy understanding. After all, the question of whether or not I believe *p* is not different from the question of whether or not *p* is case, so long as they are being posed in the first-person present indicative. Knowing *p* is then, in this sense, knowing that one believes *p*, which I can do without

ever having to make myself, and my thoughts, the direct objects of my attention. After all, Merritt argues, what is the alternative? Kant repeatedly tells us that we can never know, at least for sure, just what our real motives, intentions, or maxims are. Introspection is not just unreliable; rather, because its objects can only be given in time but not in space, introspective psychology can never become a science. It would seem that, if the self-knowledge required by morality involved such an explicit grasp of one's own mental states, we could not hope to make even the slightest progress toward virtue.

This is a highly appealing interpretation. However, it does not seem to fit with what Kant goes on to say about moral self-knowledge. Here is how Kant explains the Delphic injunction:

> "This command is *know* (scrutinize, fathom) *yourself,*"…in terms of your moral perfection in relation to your duty. That is, know your heart—whether it is good or evil, whether the source of your actions is pure or impure, and what can be imputed to you as belonging originally to the *substance* of a human being or as derived (acquired or developed) and belonging to your moral *condition*.
>
> Moral cognition of oneself, which seeks to penetrate into the depths (the abyss) of one's heart which are quite difficult to fathom, is the beginning of all human wisdom…. (**Only the descent into the hell of self-cognition can pave the way to godliness**.) (*The Metaphysics of Morals* 6:441, my boldface).

Here, Kant seems to be very clear that moral self-knowledge is not a matter of looking out at the world, of fully grasping the situation we are faced with. Rather, such self-knowledge is indeed a matter of scrutinizing oneself, i.e., one's heart, one's motives ("the source of your actions") and what is grounded in our substance (autonomous agency) or our condition (happiness, etc.). This is not a matter of looking out, but looking inward, of "fathoming" or "descending" into the depths of oneself. If Merritt were right, there would be no reason for self-cognition to involve a kind of "descent," and no reason for it to be any kind of "hell." Admittedly, attending to the moral features of a particular situation could sometimes be hellish (as with current politics), but it needn't be; surely, we may often find ourselves struck by the courage, kindness, patience, or integrity of those we are dealing with. Yet Kant believes that we are all necessarily afflicted with a radical evil, and that evil leads us to perpetually misrepresent out own motives and intentions as being far nobler than they are (even when our actions are "legal" in the sense of being in external accord with

morality). If so, it is no wonder that such self-knowledge can be a hellish experience; if Kant is right, real honesty with ourselves involves a continual experience of "humiliation" whereby all our moral pretenses (about our own virtue and self-worth) are continually being "struck down" as being self-serving shams.

For Kant, our basic moral problem is not just that we are prone to confusion, distraction, or temptation in particular situations. While we do face such challenges, the deeper problem is that, under the influence of our radical evil (or relatedly, our "self-conceit"), we endlessly rationalize and deceive ourselves about what we are doing and why. It is not just that our vision is often blurry; but rather, that we think we are seeing clearly when we are really in the grip of an illusion. Knowing more about morality, even in its particulars, is no help, because those considerations will just serve as more material with which to convince ourselves of our own virtue. If I am profoundly self-deceived (and not just confused or ignorant) about climate change or vaccinations, it won't help to just give me more information (since I'll just reprocess that in a way that reinforces my delusion). Rather, I need to come to realize something, not about the climate, but about *myself*; I need to how and why I am driven to resist these truths, and in so doing release myself from the illusion I have been casting for myself. There is something fundamentally therapeutic about this task, where such therapy involves coming up with the right kind of interpretation of oneself.

These worries about self-deception are evident in what Kant has to say about the passions. Kant does not think that our inclinations are themselves bad, and that at worst they become occasions of temptations to weakness of will through our ordinary self-love (*Eigenliebe*). The deep threat to morality is found in the passions, a kind of mutated inclination (Kant calls them "cancerous sores") that does not merely motivate us, but pretends to be an alternate source of authority to rival morality. As Merritt understands passions such as ambition, they all involve a failure of normative reflection in her sense, in which we become transfixed by a particular inclination (say, a desire for esteem), and so cannot bring that inclination into proper comparison with the rest of our desires, thereby taking the part for the whole. However, I think that for Kant the passions do not involve merely a lack of reflection, but a pervasively corrupted form of reflection that has hijacked and distorted basic rational norms, thereby pretending not just to be dominant (seizing attention), but legislative (and so commanding attention). The passions (also known as the "manias" in the *Anthropology* or the vices of the *Tugendlehre*) are all parodies of

reason, where some form of self-love asserts itself under the guise of some rational (especially moral) ideal. Envy is a corruption of equality, ingratitude of independence, vindictiveness of justice, arrogance of self-respect. Unlike ordinary inclinations, the passions are not primarily directed toward objects (food, drink, shelter); instead, Kant tells us that the passions are all fundamentally addressed to other people ("the passions are only appetites directed by men to men, not to things…and can also be satisfied only by men" (*Anthropology from a Pragmatic Point of View* 7:268, 270)); the passions are ways not just of wanting certain things, but laying claim to them as a matter of right, as a kind of entitlement. Every passion involves a kind of "illusion," in which a person convinces herself that she is recognizing some rational/moral demand, when in fact she is only engaging in some kind of cloaked self-assertion (such "inner practical illusion" consists in "mistaking a subjective element in the grounds of action for something objective"; in passion, a person becomes "the fool (dupe) of his own inclinations". (*Anthropology* 7:274, 271))). Such passions are not just cases where, as Merritt argues, one inclination merely eclipses all others in terms of salience, making a proper appreciation of the whole impossible. Instead, passion is "an enchantment that…refuses to be corrected…[passion] always presupposes a maxim, on the part of the subject, of acting in according with the end prescribed to him by the inclination. So it is always connected to his reason… (*Anthropology* 7:266-7)".

Kant tells us (as Merritt notes), that unlike the affects, the passions are "consistent with the calmest reflection" (*Anthropology* 7:265). Indeed, Kant thinks that "brooding" over the passions only strengthens them, like the flow of a river cuts it ever deeper into its bed. The problem, it seems, is that when we are in the grip of the passions, it is our very capacity to reflect that has been distorted. If the envious or vindictive person reflects more deeply about the nature of justice and desert, he will just become more envious and vindictive, since his vice consists precisely in an illusory understanding of justice in the first place. Here it will not help to attend more closely to the world, to think more seriously about what the truth is. Rather, we need to apprehend the motives that are leading us to lie to ourselves, precisely to release us from such self-cast illusions. And this, I'm afraid, does involve some pretty sophisticated theorizing about oneself and about morality.

This is not to deny that Merritt's picture of virtue might apply to some kind of rational agent. Perhaps there can indeed be a kind of "holy idiot," like Forrest Gump or Prince Myshkin, who is very virtuous despite being incapable of much by way of self-reflection. Such people would not have to go through the "hell of self-cognition," because they do not have the reflective

capacities needed to deceive themselves in vicious ways to begin with. Such self-deceit, (and with it the possibility of the passions) depends on our being able to tell ourselves narcotizing stories about our own choices. It takes a fair amount of reflective sophistication to maintain a pretense, and even more to buy into it oneself. If so, then Myshkin or Gump, who are as transparent to themselves as they are to others, have no need of fancy self-knowledge to be good. However, once a person starts to engage in such self-interpretation, moral pathologies emerge that can only be treated by more reflective self-knowledge ("A dog cannot lie; neither can it be honest"). As Wittgenstein said of philosophy in general, self-reflection is the only cure for the disease that it itself represents.

What then about Kant's famous caveats about the limits of introspection and self-knowledge generally? One point to note is that Kant never says that we can't make *any* progress in self-knowledge; even if we can never know ourselves with certainty, we may still be able to make some pretty educated guesses. That might not be much for purposes of science, but it might be enough for the practical task of moral reconstruction. In addition, Kant's doubts seem directed only at the thought that we may be acting virtuously or from the motive of duty. It's not clear that we should be equally doubtful about our judgments that we've done something wrong. Even if I can never be certain that I have acted from the motive of duty, I may still be able to tell that I have acted from self-love or self-conceit. I don't know if I have ever spoken honestly, but I'm quite sure I've lied from envy and fear of embarrassment. Such self-knowledge may be always provisional and incomplete, but it still might be "good enough for government work."

Merritt rightly observes how little faith Kant has in the deliverances of introspection. However, I don't think the kind of reflection needed for the requisite self-knowledge need be introspective in any interesting way. When I try to honestly make sense of myself, to understand what I really care about and why, I don't think I simply peer into my mental life, in the attempt to observe what is going on (as I might make sense of some external phenomenon before me). After all, that's not the only way I have to make sense of other people, either. I have a great interest in understanding my spouse, but I don't do this simply by watching her and trying to come up with the best explanation of my observations. Of course, I do think about such public facts, but I also talk *to* her. I ask questions, listen to her responses, offer alternative readings, etc. (usually as she does the same thing with respect to me). Here we are jointly constructing and challenging stories about ourselves; ones that have to answer to some outer (and inner) realities. I doubt one can come

up the absolutely correct reading of a person here; indeed, I doubt there really is such a unique, determinate fact of the matter at all. But we can certainly do better or worse in this endeavor, and achieve something that counts, for all practical purposes, as a piece of real self-understanding.

I'd like to suggest that, whatever Kantian moral self-knowledge is, it involves having something like this kind of conversation not just with oneself, but with other people engaged in the same enterprise (this may help explain why Kant insists that progress toward virtue is necessarily a collective task, and one that cannot be completed in any finite span of time). Such successful self-interpretation will not be merely (or even primarily) a matter of clairvoyant introspection, but will instead involve powers of self-reflection that incorporate sophisticated forms of psychological and philosophical theorizing. The demand for such self-theorizing is not based merely in a vestigial Platonic desire for the form of the good. Rather, we need such ever-more sophisticated kinds of self-reflection in order to expose the increasingly subtle forms of self-deception that our growing powers of self-refection themselves engender.

Perhaps such honesty and insight into oneself should still count as a kind of cognitive skill. But if so, then the notion of the "cognitive" has been stretched past its normal meaning that would have it contrast with the volitional, the affective, or the persuasive. The "skill" involved would not just be that of deftly coping with the world, but in managing to be honest with ourselves, despite our well-grounded and inescapable mistrust of ourselves. However, I do not mean to resurrect the caricature of the Kantian agent who spends all her waking hours trying to formulate and test her maxims. Merritt is surely right that, for anyone with a modicum of virtue in anything like half-decent circumstances, a proper commitment to morality expresses itself in patterns of attention, affect, and response that are properly directed toward the world rather than to oneself. The need for such particular acts of self-reflection is indeed usually an indication that something has gone seriously wrong, either in oneself or one's situation. But this does not mean that we do not have a fundamental obligation to try make sense of ourselves, if only to guard against our own ineradicable tendencies to concoct false narratives of who we are and what we are doing.

Bibliography

Kant, Immanuel. *Anthropology from a Pragmatic Point of View.* Translated by Mary J. Gregor. The Hague: Martinus Nijhoff, 1974.

Kant, Immanuel. *The Metaphysics of Morals*. In *Practical Philosophy*, translated by Mary J. Gregor, 353-604. Cambridge: Cambridge University Press, 1996.

Kant, Immanuel. *Religion within the Boundaries of Mere Reason*. In *Religion and Rational Theology*, translated by George di Giovanni, 39-216. Cambridge: Cambridge University Press, 1996.