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 Francey Russell

Perhaps more than any other canonical philosopher, Kant can lend himself to caricature: his metaphysics risks dislodging agents from the material world, his regulative ideals look like goals we must pursue all the while aware that we cannot achieve them, and it can seem that the Kantian practical agent is constantly stepping back, surveying both her action-context and her own motives, and making choices based on conscious rational deliberation. In *Kant on Reflection and Virtue,* Melissa Merritt is concerned to correct this latter caricature in particular. This book is on the whole persuasive and creative. She demonstrates the intuitive plausibility of Kant's claims while also clarifying their place in his larger system.

As the title suggests, the main effort of the book is to provide a new interpretation of what Kant means by "reflection" and why he argues that all judgments require reflection. Again, the risk is that the latter imperative could be interpreted as proposing an alienated and robotic picture of practical life. Against this, Merritt works to provide an account of the Kantian agent as, put simply, a reflective person, where being reflective does not designate a special activity but rather describes a way of being, or a "consistent cast of mind"[1] (5:152) that orients one's practical and theoretical engagements with self, world, and others. Merritt makes three central interventions in support of this picture. First, we need to distinguish two senses of reflection in order to understand Kant's conception of our mind as essentially reflective, where *constitutive reflection* (c) is a basic requirement for thought and experience as such, and *normative reflection* (n) involves a commitment to standards of correctness and to truth more generally, where these commitments can be realized more or less well. Second, we need to conceive of normative reflection, again not as some special action one undertakes but rather as the spirit in which one engages one's cognitive

[1] In the case of the *Critique of Pure Reason*, I follow the standard practice of referring to the 1781 (A) and 1787 (B) editions. References to all other texts works are to the Prussian Academy pagination appearing in the margins.

capacities; because of this it makes sense to conceive of normative reflection in terms of one's cognitive *character*. Third, this suggests that we should analyze healthy human understanding committed to the standards of normative reflection as a kind of cognitive *virtue*, where the cognitively virtuous person is characterized by her practical capacity to judge and act in light of her commitment to truth. This is the spine of the book and these claims are made substantial and subtle through many supporting arguments.

Merritt's proposal to understand reflection and reflectiveness in terms of character, and character as a kind of *Denkungsart* or *way* of thinking, is particularly creative and compelling. Merritt's suggestion here is that virtuous reflection involves something like *style*, a mode of being minded that involves being subtly attuned to oneself, one's situation, and other persons. To be reflective here involves having a good sense of what kinds of questions need to be asked and when. For instance, Merritt emphasizes Kant's commitment to self-opacity, and elsewhere Kant cautions against a kind of arrogance to which we are prone that involves taking moral credit for one's good conduct where, "strictly speaking," one is simply lucky enough to have escaped real temptations to vice (6:460). This is a kind of self-conceit that involves mistaking one's good fortune—what we might now call *privilege*—as evidence of one's excellent moral disposition, and Kant conceives of this as a form of dishonesty "by which we throw dust in our own eyes" (6:38). If these are risks to which human beings are prone, the virtuously reflective agent will possess a kind of humility with respect to her claims to know, and will know when to ask whether things really are as she's taken them to be. This picture recalls what Lorraine Code calls "a commonsense, practical skepticism of everyday life" (2006, 224), which involves a readiness to self-critique, an acknowledgment of one's own fallibility, and an ongoing awareness that "one is never more easily deceived than in what promotes a good opinion of oneself" (6:68). And because all this is a matter of character, the point is not that the Kantian agent is implausibly self-skeptical or mechanically subjecting all her beliefs to painstaking review; rather reflection "infuses" (KRV 159) and inflects her basic cognitive orientation, like a style. What Merritt does so well is show that Kant offers an attractive picture of reflection, humility, and self-critique that does not succumb to the caricature. At the same time, one can see the continuity between ordinary, everyday reflection and more explicitly philosophical reflection, which we may think of as a more disciplined and specialized exercise of this same basic capacity.

In what follows I will take up Merritt's discussion of affect and passion, and then self-knowledge and virtue. Merritt's work brings much light to these topics, though I will object to some of her readings of Kant.

**1) Concerning the relation between affect, passion, and reflection**

In Chapter One, Merritt demonstrates how her two senses of reflection on the one hand, and the distinction between affect and passion on the other, can mutually illuminate each other. For Kant, both passion and affect undermine the sovereignty of reason (7:251), and Merritt maintains that both mental states constitute modes of reflective failure. Yet while passion essentially involves reflection, affect essentially lacks it. But if this is so, then it looks like Kant must be drawing on two distinct notions of reflection: that passion and affect apparently dis-engage reflection in two very different ways suggests that Kant is operating with two different kinds of reflection. Merritt thus proposes that affect lacks reflection-c, while passion lacks reflection-n. While there is something right about the idea that both mental states constitute modes of reflective failure, I think Merritt does not offer quite the right analysis of affect and passion.

Affects, Kant writes, are "honest and open," involving surprise through sensation; affect is a feeling of pleasure or displeasure that *does not let one rise to reflection*, or again, a feeling that "quickly grows to a degree of feeling that *makes reflection impossible* (it is thoughtless)" (7:251). So affect undermines one's capacity for reflection in a kind of rush of feeling, as in Kant's example of the rich man feeling overwhelmed by anger when his servant breaks his goblet.

Passion by contrast is a habitual desire (or inclination) that Kant says can be conquered by reason only with difficulty or not at all. Unlike open and short-lived affects, passion is hidden and deceitful; passion "takes its time *and reflects*, no matter how fierce it may be, in order to reach its end" (7:252). The passion for sex is an example of what Kant calls an innate passion, the mania for dominance an example of an acquired or cultural passion. Crucially, Kant claims that "passions can be paired with the calmest reflection […] they are not thoughtless, like affects, nor stormy and transitory; rather they take root and can even co-exist with rationalizing." So passion is an inclination for some non-moral end that takes root and becomes habitual precisely through the machinations of reflection and rationalization. And precisely because of this, passions "do the greatest damage to freedom…passion is an enchantment that refuses all recuperation" (7:266).

From all of this Merritt concludes that both affect and passion are modes of reflective failure, which suggests there must be two corresponding modes of reflection. And again, for Merritt, affect lacks reflection-c whereas passion lacks reflection-n. Notice, though, that this means that in the grips of an affect, a person lacks even what Merritt describes as "the basic tacit handle" on herself that is constitutive of thinking and experience for a finite rational being (that is, the kind of reflection Kant is concerned with in the first *Critique*) (KRV 18 and *passim*). By contrast, in the grips of a passion, a person will take up some practical point of view, pursuing ends based on reflected-upon commitments, but she will not make *good* use of her cognitive capacities; she will fail to take an appropriate interest in her own cognitive agency, fail to be appropriately oriented by the three maxims of healthy human understanding; thus, she lacks reflection-n.

Yet this can't be the right way to understand the reflective failure involved in affect. For if reflection-c is constitutive of thought and experience as such, the very basic consciousness of I as subject, then its absence in affect would render affect paradoxically unexperienceable ("less even than a dream" [A112]), or like an utterly alien episode that cannot be knitted into one's overall experience. While affect may rise up and overwhelm in us as a surge of feeling, it does not typically obliterate such basic self-consciousness, except perhaps in very extreme cases. Some of the examples of affect that Kant cites include fright, anxiety, shame, and cheerfulness. Again while one may in some sense "lose oneself" in these affects, such self-loss would seem more akin to acting or feeling unusual or out of *character*, rather than, paradoxically, experiencing something without the very self-consciousness constitutive of experience. If affect lacks reflection-c, it either cannot be experienced *or* it is experienced as a kind of possession, and neither of these seem like plausible pictures of our life with affects.

I propose that the more apt way to differentiate affect and passion is as follows: both involve disruptions of reflection-n; yet, while affect involves a more encompassing failure or *inability* to reflect-n, passion involves what I would call the *ersatz exercise* of this capacity, that is, a habitual and perverse misuse of the capacity for reflection. While there may be extreme cases of dissociative affects that are so overwhelming that they disrupt one's capacity for reflection—for example, in certain instances fright or rapturous pleasure—but these must be either exceptions or a specific sub-category of affect, something more like a trauma-level emotional experience. Ordinary affect cannot typically undermine reflection-c, for then much of our emotional lives

would be oddly unexperienceable. Such a picture would push us towards the kind of caricature of Kant that Merritt rightly wants to avoid: this would be the emotional counterpart of the caricature of the stepping back picture of reflection, with Kant as an overly squeamish philosopher gripped by a slightly hysterical conception of the disruptive power of affects.

On my view, everyday affect renders us not completely blind, as would be involved in the absence of reflection-c, but "more or less blind," as Kant himself puts it (7:253). Here one is unable to make good use of one's cognitive capacities, hence one's capacity for good judgment is undermined. We can picture a range here: on the extreme end, this failure may be so extreme as to compromise one's capacity for cognition as such, and at the other end this failure may render good or precise or objective judgment impossible. But I would still describe this as a failure of reflection-n, not reflection-c.

Turning now to passion: we saw that passion involves a kind of reflection and reasoning, as Merritt notes. But the problem here is not that in passion one *fails* or is *unable* to reflect well (as in the throes of affect); rather the more unnerving problem is that in the grips of passion one engages in *ersatz reflection*, a perversion of its proper exercise. Merritt picks up on this perverse mimicry of the good case when she notes that the logical egoist "*mimics* the reflective person" (KRV 45) (where logical egoism is a form of prejudice, and Merritt has argued that we should analyze passion as a kind of prejudice). Kant writes that passions subject us to *delusion*, which is the "practical illusion of taking what is subjective in a motive for something objective" (7:274). This is precisely what makes passions so difficult to correct, since the passionate person is to some degree rightly oriented: insofar as she takes what is merely subjective as if it were objectively valid, the passionate person displays some concern for meeting the standard of objective validity, hence her engaging in reflection and rationalization. In the grips of a passion for, say, honor, what I seek is to be recognized by others; in fact, all I really seek is a *reputation* of honor where semblance suffices (7:272), but I *take myself* to be seeking recognition for what *I take* to be my objective value. So I have reflected on the value of honor as an end to be pursued, and my rationalizing activity allows me to delusorily believe that I have earned such honors and that others rightly owe it to me, and that this whole exchange is justified. By being steadily oriented by such a passion, I precisely do not fail to reflect and I do not make an ordinary kind of error (for example, believing falsely and sincerely that certain actions would earn me *real* moral esteem); rather, I am reflecting and reasoning while in the grips of a practical illusion, a false but encompassing

conception of what is worth pursuing and how, where this is guided by self-love rather than reason. So again, this means that while affect involves a failure to reflect-n, passion involves an *ersatz* or perverse version of it.

This leads me to make a general remark about something I'd wished to hear more about, which is how Merritt understands *illusion* in general and also self-conceit in particular, vis-à-vis her account of Kantian reflection. For Kant, all illusion involves "taking a subjective condition of thinking for the cognition of an object" (A396), or taking something that is merely subjectively valid (either valid only for me or only for human cognition) as if it were objectively valid (valid for all cognizers or true of things in themselves). In addition, for Kant "illusion is that delusion which persists even though one knows that the supposed object is not real" (7:150). So, an illusion is an erroneous way of taking something—*as if* it were objective or real—where this way of taking cannot be, as it were, simply shaken off or corrected, and perhaps is never finally overcome. Rather even as one recognizes that one's way of seeing is only subjective, one continues to see things *as if* they were objective (the way we continue to see the moon as if small even though we know that it is large).

Now there are many things one can say about the various ways in which illusion plays a role in Kant's system, but what makes this quite salient for Merritt's project is the fact that illusion is a "deformity" (to use Kant's word) to which *only* rational, reflective minds are prone. Non-rational creatures can make mistakes but they cannot be gripped by an illusion (or a passion). Thus, a complete account of Kantian reflective agency would need to clarify how we ought to understand illusion in general and self-conceit, and their place in the life of the reflective Kantian agent.

**2) Self-Knowledge and Virtue**

The ongoing work of avoiding illusion and prejudice involves the cultivation of what Merritt calls healthy human understanding as a basic cognitive virtue. The three maxims of healthy human understanding (see 5:294) specify the general frame of mind from which to judge, and describe a general commitment to unprejudiced thinking and to truth. Here, again, reflection characterizes the *way* one engages in cognition, it "infuses" one's theoretical and practical cognitive activity, and hence lodges at the level of character rather than as some specific action (like stepping back

to reflect). So, for Merritt, Kant is interested not in episodic moments of stepping back, but in accounting for our kind of mindedness as essentially involving the capacity to exercise discernment in our engagements with self, world, and others.

This way of understanding the reflective mind informs how Merritt interprets Kantian self-knowledge. For Merritt, the First Command of all duties to oneself—the command to "*know* (scrutinize, fathom) *yourself* […] that is, know your heart" (6:441)—should be understood, not as a command to introspect, but as a command to be generally reflective in one's engagements with the world, to pay attention not to oneself but to *what* one pays attention to. Thus, Kantian self-knowledge is not self-directed but *world*-directed.

Indeed, Kant expresses deep reservations about the effort to achieve moral self-knowledge by looking inward. He specifies particular ways we try to know ourselves that are doomed to fail, and he refers to such efforts, variously, as "self-examination" (4:407), "plumbing the depths [of the human heart]" (ibid.), "self-observation" (6:63), and knowing by "inner experience" (ibid.). For Kant, efforts to discern one's motives and practical principles as if they were locatable in some inner time and place is "absurd" (7:135). In the *Lectures on Ethics*, Kant refers to this method as a form of "eavesdropping on oneself" (27:365). The introspective method for self-knowledge can be conceived as a form of eavesdropping precisely because it wants to "catch" motives in their efficacious activity, to witness one's own practical reasoning while it operates unawares. Kant conceives of this effort as either already a "disease of the mind (melancholy)" (7:134) or as easily leading to "enthusiasm and madness" (7:132), and that "spying" on one's own "thoughts and feelings" (7:133) indicated a kind of self-satisfying obsessiveness masquerading as moral inquiry.

On the other hand, while there may be something dubious about picturing self-knowledge on the model of a kind of perception turned inward and while Kant himself recognized such dubiousness, the command to *scrutinize one's heart* seems on the face of it to be a matter of exacting moral self-assessment, and not with the broad character of reflective cognition with which Merritt is concerned. While only such reflective minds could be commanded to know themselves, the First Command seems in fact to be a more self-involved, critical, and perhaps episodic affair than Merritt's reading suggests. Merritt's interpretation seems to have been influenced by contemporary, "transparency" accounts of self-knowledge, according to which one knows one's own mind by looking not inward but outward (see Boyle 2011; McGeer 2007; Moran 2001). And again, while Kant did indeed reject the introspective method, it is not obvious that he thought the

command to know oneself could be satisfied by simply being reflective in an ongoing way in one's engagements with the world and others. That is, there is a pressing interpretive question: how can Kant command us to know and scrutinize ourselves *without* resorting to methods of self-observation or introspection? So, while Merritt is correct that Kant rejects the introspective method, *how* exactly we should interpret the First Command is not so straightforward.

Kant also presents another kind of self-knowledge in his presentation of the First Command. Kant insists that you must know yourself "in terms of what can be imputed to you […] as belonging originally to the substance of a human being" (6:441). Let me say something about what I think this means. In the chapter on moral motivation in the second *Critique*, Kant pursues an extended contrast of the attitude of self-conceit with the attitude of *virtue*, which he describes as *the moral disposition in conflict*, an attitude of striving (5:83) and struggle (5:84). The point here is not primarily to insist on having a pained or acutely conflicted consciousness, but rather to capture the idea that the moral law presents to us in the imperatival mode alone. And insofar as one stands in this kind of relationship with the law, Kant writes that one must acknowledge or know oneself to be, as he puts it, a *creature*, "hence always dependent with regard to what he requires for complete satisfaction with his state [and thus] never entirely free from desires and inclinations… which do not by themselves harmonize with the moral law" (5:84). Thus, the attitude of virtue involves understanding one's relationship to the moral law as imperatival and "appropriate to *our station* among rational beings as *human beings*" (5:82). Kant writes that whereas the attitude of virtue involves practical, moral appreciation of oneself as a human being, a creature, self-conceit involves mis-conceiving oneself as a different *kind* of being, one with a naturally (and yet voluntarily) good will that "requires neither spur nor bridle" (5:85). So, there is a failure of self-knowledge in self-conceit, not just at the level of individual character, but with respect to what we might call *practical, anthropological self-knowledge*.

Connecting this up with the First Command, the result seems to be that the human being stands under a command to know his station amongst rational beings, to know his mind as the kind that needs spur and bridle, hence a command to know himself, morally and practically, *as* a human being. Kant suggests that while virtue need not require any specialized expertise regarding human nature (gleaned, say, from sociology or biology), virtue *does* require the kind of knowledge of human being that comes from some experience of *being* a human being, subject to inclinations that will never by nature conform to law.

I think Merritt's reading of Kant can help us make better sense of this. Clearly this kind of self-knowledge must be available to common understanding, which suggests this is a form of self-knowledge that can be tacit, just as common understanding grasps the principles of its exercise only tacitly. Further, if "experience is the sole instructor of common understanding" (KVR 64), as Merritt puts it, then this is a kind of morally-salient anthropological self-knowledge that must be acquired over the course of concrete moral practice, resulting in a practically-guiding appreciation for the kind of fallible creature one is. Again, this would be the kind of knowledge that comes, not from working on the human sciences, but from the ongoing work of *being* a human being. Finally, this would seem to fit with Merritt's skill model of moral virtue: one can only exercise skillful moral judgment if one appreciates the kind of creature one is, including the kinds of illusions to which one is prone.

While much of Merritt's work provides a useful frame for understanding such anthropological self-knowledge, this latter idea actually reveals that Merritt misunderstands an important feature of Kantian virtue, including its difference from the *holy will*. Merritt writes that "the holy will should have the same strength [as virtue] because this strength is essentially cognitive: it is the readiness of one's commitment to morality" though there will be a "difference between the holy will and the virtuous person as regards the *content* of their commitments to morality" (KVR 203). For Merritt, strength is the same for both "inasmuch as both holiness and virtue are conceived by Kant as the perfection of practical reason" (ibid.). Thus, for Merritt, "virtue is a human ideal […] a *perfection* of the will, of practical reason" (KVR 202-my emphasis).

I think this is actually not the right way to understand Kantian virtue. For Kant, while virtue is an ideal to strive for, virtue is not the perfection of practical reason but is rather the attitude in the struggle *towards* such perfection, where perfection, the ideal of holiness, "is not attainable by any creature but is yet the archetype which we should strive to approach and resemble in an uninterrupted but endless progress" (5:83). Again, the attitude of virtue is an attitude of striving and struggle that precisely bears in mind, however tacitly, the *kind* of creature that one is and one's *station* amongst rational beings. As Kant continues, "if a rational creature could ever reach the stage of thoroughly *liking* to fulfill all moral laws, this would mean that there would not be in him even the possibility of a desire that would provoke him to deviate from them" (ibid.). But the virtuous person knows that his desires and inclinations will never, by nature, conform to the law. This need not be an especially paranoid or tortured position, but rather humble and honest and self-

critical (in just the way Merritt recommends, in fact). Thus, human beings must be *committed* and continually re-committed to morality, precisely *because* of the fact we can never be free of our desires and inclinations which do not of themselves accord with the moral law, and precisely because we know that the dear self may obfuscate this fact, making us think we do what is our duty as proud and willing volunteers. But because of this, it doesn't seem that *commitment* to morality figures in the holy will at all. Hence to my mind the difference between the holy will and virtue is not merely a matter of content, of things that need to be kept specially in mind from our point of view, as Merritt puts it (KVR 203). It is a wholly different *kind* of orientation. This is worth emphasizing not in order to get the right conception of the holy will (which as Merritt rightly points out, can only be speculative) but to secure the right conception of virtue as an attitude of moral struggle proper to our station as rational *animals*.

Let me raise one more question about Merritt's conception of Kantian virtue. Merritt argues that we should understand reflection as cognitive virtue, and virtue as a free skill, where these are skills of discernment that are open to and constituted by reflection (as opposed to a model of skill as unthinking or mechanistic habit). And, in brief, Merritt proposes that while the commitment to and respect for truth governs what will figure as salient in action and cognition, this commitment is only rendered determinate through steady practice and the concrete engagement of one's attention. That is, one's overarching and guiding commitment becomes increasingly determinate (rather than abstract and vague) to the degree that one cultivates the resources to actually and concretely judge in light of that commitment. As Merritt very helpfully puts it, the strength of one's cognitive commitment just is the extent to which one can have a concretely action-guiding through by means of it (KRV 188). This allows Merritt to offer a new way of conceiving of the difference between virtue and lack of virtue (which is different from active vice). For Merritt, both *Tugend* and *Untugend* share a commitment to morality. As Kant writes, *Untugend* can coexist with the best will (6:408), but *Untugend* lacks the resources for acting in a way that concretely realizes this commitment, which means that the commitment itself remains correspondingly vague.

I wondered if this could be seen as tracking Aristotle's distinction between character virtue and practical wisdom, where "virtue makes the goal right, practical wisdom the things leading to it" (1144a7-9). For Aristotle, character virtue concerns one's desire for, and taking pleasure in, fine things; that is, it describes a general and deep-rooted emotional orientation towards the good.

But with character virtue alone, all we can say is that one's heart is in the right place; and this is because character virtue needs to be complemented by practical wisdom, the capacity to discern, concretely, what would be really good or fine to do. While Merritt indicates that she takes Kantian virtue to be quite different from Aristotelian virtue, this sounds quite close to Merritt's idea that one may be committed to the good and yet lack the resources to determine what specifically would be good to do.

Notice also that Merritt's conception of virtue as perfection, rather than the attitude in the struggle, actually makes Kantian virtue much closer to (certain readings of) Aristotelian virtue. For if virtue is the *perfection* of practical reason, such virtue sounds close to, say, McDowell's conception of the virtuous agent, where all claims that run counter to morality are "silenced." Against such a reading, I again would argue that for Kant the claims of inclination and the dear self can never be wholly silenced; rather human beings can only strive for such perfection, while at the same time bearing in mind that ours is a mind that will forever require spur and bridge. This struggle and striving *is* virtue, it is not deficient in virtue, but it is not perfection.

*

That Merritt's book provides an occasion to think more deeply about these difficult, fascinating topics in Kant is exactly what makes it so refreshing, creative, and careful, a deeply rewarding work for anyone interested in Kant's picture of mind and morality.

## Bibliography

Aristotle. *Nicomachean Ethics.* Translated by C.C. Reeve. Indianapolis: Hackett Publishing, 2014.

Kant, Immanuel. (1787) *Critique of Pure Reason*. Translated and edited by Paul Guyer and Allen Wood. New York: Cambridge University Press, 1999.

_____. (1785) *Groundwork for the Metaphysics of Morals* in *Practical Philosophy*. Translated and edited by Mary Gregor. New York: Cambridge University Press, 1999.

_____. (1788) *The Critique of Practical Reason*. in *Practical Philosophy*. Translated and edited by Mary Gregor. New York: Cambridge University Press, 1999.

_____. (1797) *The Metaphysics of Morals* in *Practical Philosophy*. Translated and edited by Mary Gregor. New York: Cambridge University Press, 1999.

_____. (1788) *Critique of Judgment*. Translated and edited by Paul Guyer and Eric Matthews. New York: Cambridge University Press, 2001.

_____. *Lectures on Ethics*. Translated and edited by Peter Heath. New York: Cambridge University Press, 2001.

_____. (1793) *Religion within the Boundaries of Mere Reason*. Translated and edited by Allen Wood and George di Giovanni. New York: Cambridge University Press, 2003.

_____. (1798) *Anthropology from a Pragmatic Point of View* in *Anthropology, History, and Education.* Translated and edited by Robert B. Louden. New York: Cambridge University Press, 2007.

McDowell, John. "Some Issues in Aristotle's Moral Psychology," in *Mind, Value and Reality*, 23-49. Cambridge: Harvard University Press, 1998.