Moral Agency as Cognitive Agency: Recovering Kant’s Conception of Virtue

JESSICA TIZZARD

University of Connecticut, USA


The aim of Merritt’s exceptional book is to offer a sustained and unified account of Kant’s conception of reflection, a conception that in turn reveals his account of virtue as a concretely engaged, world-directed skill. Such a revelation will no doubt be a surprise to many. It is, of course, well-established that Kant thinks reason and the capacity to reflect is foundational to moral agency and the possibility of virtue. But the question remains, how should we understand the transformative role that rational reflection plays in the lives of those who possess the capacity to engage and develop it? Merritt’s analysis deftly reveals the shortcomings involved in framing reason as the ability to austerely “step back” and direct one’s attention to the abstract principles that speak for or against one’s way of acting. What this picture lacks, in her view, “is when and how this reflective capacity is to be engaged, and to what end” (15). A widespread failure to recognize that Kant’s view of practical reason is deeply concerned with these questions has led to the misappropriation of his philosophy, especially as it relates to virtue. On closer inspection, Kant’s account of the proper use of reason reveals that the requirement to reflect is not discharged through separate acts of reflection—over and above our everyday practical attention—but rather through conducting this attention reflectively. This requirement is thus met “in adverbial fashion: it is met by considering how things are in the right spirit” (72).

With this approach, Merritt’s book is in a good position to make sense of a number of difficult and important passages from the Doctrine of Virtue in the Metaphysics of Morals. Her textual foundations reach well beyond this locale, creating a holistic vision of

• Assistant Research Professor, University of Connecticut; PhD University of Chicago. Contact: jessica.tizzard@uconn.edu

[Recibido: 16 de mayo de 2020
Aceptado: 22 de mayo de 2020]
Kant’s system that draws from all three Critiques, the Anthropology, the Jäsche Logic, and well beyond. But her ultimate focus is the argument, advanced through what she calls the specification thesis, that moral virtue is cognitive virtue considered in its practical species (113). The book thus builds towards the claim that moral virtue is skillfully engaged practical cognitive activity, activity grounded in “the capacity to recognize what morality requires of one and the readiness to be moved by this recognition” (152). In addition to giving a sense of what readers can expect from Merritt’s book on the whole, my aim is to explore the way this claim marks a significant contribution to literature on the Doctrine of Virtue and Kant’s conception of good moral character more generally.

The significance of reflection and the knowledge at which it aims

Merritt insists that Kant conceives of reflection, and by extension virtue, “in resolutely cognitivist terms,” adding that “it is in the interest of knowing, and that alone, that we ought to be reflective” (208). While it could seem as though this reaffirms the conception of reason focused around lofty abstract goals, in fact this characterization is meant to direct us towards Kant’s insistence that knowing the world involves being in touch with objects in the theoretically-laden, robust Kantian sense of the term. To have knowledge that is by nature objective, is to have used one’s cognitive capacities to grasp how things truly are, rather than how they merely seem to be in light of one’s subjective situation. One of Merritt’s most intricate, book-spanning arguments is that Kant thinks we cannot obtain such knowledge without being interested in cultivating how we actively attend to and engage with the world. Whether the object of knowing is a theoretical science like physics (130), tracking how things are with every-day items like pieces of clothing (89-90), a technical skill like playing tennis (168-9), or a moral duty like tending to one’s parents’ happiness (192), being a good knower requires developing the capacity to properly attend to the determinate particulars at issue. One of the most impressive aspects of Merritt’s work is her persuasive argument that, for Kant, this is an incredibly complex task that requires much more agency than it might seem.

To get a sense of how reflection figures in this aspect of Kant’s view, Merritt introduces a distinction between constitutive and normative forms of reflection in Chapter 1 (16). Each of these answers to basic requirements on thinking, with a distinction drawn between those constitutive requirements (reflection-c) “without which there could be no employment of the understanding—no thinking—at all” (23), and those normative requirements (reflection-n) that are “necessary to make good use of our cognitive capacities” (23).

Reflection-c speaks to the unifying, synthetic activity of the understanding expressed most serenely through Kant’s account of the transcendental unity of apperception in the Transcendental Deduction (B132). On Merritt’s description, “one
‘reflects’ in this sense simply by having some tacit handle on oneself as the source of a point of view” (18). The full complexity involved in achieving the sort of reflection expressed through the “I think”—as a source of synthetic unity that captures a point of view on the world—is explored more fully in Chapter 3, one of the most difficult and rewarding parts of the book. There, Merritt argues that we neglect Kant’s rich account of the engaged intellect if we think of this reflective activity as primarily sub-agential (95).

Working with important passages from the first Critique and Anthropology, Merritt poses a plausible and well-argued alternative to the view that the principles constitutive of experience should be understood as background conditions that enable the passive perception of objects. Drawing on recent literature on the nature of attention, Merritt explores the Kantian version of a developmental account of human consciousness. Its culmination is the claim that the synthetic activity of reflection must be implicated in the way one actively attends to one’s perception: “my agency infiltrates empirical consciousness, which gives me a whole over which my attention can run, more or less freely” (95). She notes, for example, that “it takes attention to so much as divide empirical consciousness into a foreground and a background” (98). This simple activity draws on the agent’s capacity for reflection insofar as the foreground one is attending to is understood to be part of some wider synthetic whole (99). Through this account, Merritt delivers a compelling and relatable version of the Kantian thesis that in thinking about objects, we cannot abstract from the activity of the subject. To perceive as a rational being capable of reflection is, minimally, to engage the cognitive capacities that are constitutive of a world of objects in the robust Kantian sense implicated by the Copernican Revolution marking his critical turn. On Merritt’s view, this engagement happens at the level of active attention that carefully selects the part(s) of a synthesized whole relevant to what the agent seeks to know or do.

In addition to these baseline requirements on thinking, reflection-n speaks to the norms guiding the proper use of the cognitive capacities implicated in reflection-c. They are what enable one to achieve an objective point of view on how things are, one that demonstrates a well-used understanding free from prejudice and inconsistency. When Kant says that “all judgments require reflection” (A261/B317), Merritt claims he has reflection-n in mind. While judging could not take place at all without the basic activity constitutive of cognitive agency highlighted by reflection-c, a judgment requires reflection-n if it is to meet its implicit aim of giving an objective take on how things are or should be. In Chapter 2, Merritt directs her reader towards the three maxims of ‘healthy human understanding’ (gesunder Menschenverstand) articulated in the third Critique and the Anthropology, arguing that Kant’s conception of reflection-n is best analyzed into these commands (54). They are: “To think for oneself,” “to think in the position of another, or: to think in the position of everyone else,” and “to think always consistently with oneself” (55). Each maxim picks out a different aspect of what it is to think well in one’s orientation towards the world, and they jointly represent the virtuous knowing subject as someone who takes

---

1 Merritt cites Kitcher (2012, pp. 24-5) as a prominent example of such a view.
responsibility for the agency implicit to thought itself by aiming towards a systematic body of knowledge. In later chapters, Merritt describes this interest as a general “respect for truth” that all cognitive agents must cultivate (156).

There is undoubtedly much to value in Merritt’s account of reflection-n and the three maxims. As one of the most accessible parts of the book, it should stand as a resource for virtue epistemologists looking to engage with the Kantian perspective—one which is oft-neglected in favour of Aristotelian accounts. That said, the most striking aspect of Merritt’s account of reflection-n is how she builds on it to argue her view that virtuous cognitive activity should be understood as a concretely engaged world-directed skill. One could easily co-opt the three maxims for the ‘stepping-back’ conception of reason, but from Chapter 3 onwards Merritt constructs a very different reading, beginning with her impressive conception of reflection-c and its involvement in the attentive perception constitutive of our everyday experience. It is to this reading in application to specifically moral virtue that I now turn.

Moral virtue as cognitive virtue

There are two main features of Merritt’s cognitivist view of Kantian moral virtue worth highlighting to show the significance of her account. The first is its aforementioned emphasis on virtue as an engaged world-directed skill, and the second is the closely related claim that the capacity to recognize determinate moral requirements is only notionally separable from the capacity to be motivated to fulfill them. Together, they present a radical view of Kant’s moral philosophy that should be seriously considered on both textual and philosophical grounds.

This account grows directly out of Merritt’s two-pronged understanding of Kantian reflection. If reflection-n provides norms for objectivity, reflection-c describes the kind of object-directed activity that is to be normatively guided, and herein lies the fertile ground for her view of virtue. Working from the conception of reflection-c as embedded in concrete acts of attention built in Chapter 3, Merritt argues that “Kantian virtue can ultimately be understood as a readiness to see facts about one’s situation as themselves requiring certain responses of attention and action” (126). One is well-situated, reflectively, insofar as one is capable of properly executing the different synthetic acts of understanding that constitute one’s cognitive point of view on the world. In the moral context, this means being able to ascertain the morally relevant factors in any given situation. Though Merritt does not go into much detail about what these factors are, I would presume they include both what needs to be read immediately from the situation at hand, as well as how these immediate factors relate to the various moral ends (mine and others’) that need to be considered in light of them. So, moral virtue would include, for example, the ability to attend to how other people are feeling and what they want, the
various social and political relations that inform the situation, and the moral duties that presently apply, ultimately leading one to a concrete representation of how these relevant factors can be accounted for through some determinate action. To this point, in reference to the duty to attend to one’s parents’ happiness, Merritt suggests that in many cases a child “will be capable of having a more concretely determinate appreciation of their own parents’ needs and desires, as well as of their reason-giving force, than of the needs and desires of anyone they know less well and to whom they are less intimately indebted” (196). She also uses the example of a virtuous teacher who grades anonymously to combat implicit bias (203). It thereby emerges that the morally virtuous person is one who is attentive to the right things for the right reasons, and who thus has a proper grasp on the moral situation as a whole.

This view of moral virtue stands in stark contrast to the introspective model that paints it, instead, as the ready capacity to undertake separate acts of reflection that introspect to uncover one’s true motives and measure them against the moral law in abstracto. Merritt sums up her more concretely engaged, world-directed “adverbial” approach to virtue in the following particularly lucid passage, worth quoting in full:

“there is no obvious need to step back from the exigencies of action in order to compare the more determinate principle on which one proposes to act against a formal law. There is only the quality of one’s engagement in the action or activity itself. Commitment to the moral law, in other worlds, need involve no explicit thoughts about the law as such, nor even about the principles on which one proposes to act, and the ideally reflective person need not be conceived in the familiar way, as the one who is always poised to ‘step back’” (178).

It must be stressed (as Merritt herself does) that this is a very controversial point. Even with the benefit of her carefully constructed account, other reviewers have been quick to question whether fidelity to Kant’s text demands that separate non-adverbial acts of reflection remain an important part of his account of virtue. I find this resistance to be striking, as Merritt provides her own persuasive and assiduously constructed reading of one of the most obvious passages in favour of the step-back and introspect model (200). This is the passage at 6:441 in the Doctrine of Virtue, in which Kant commands that the first duty of virtue to oneself is to ‘know yourself’, i.e., “whether your heart is good or evil, whether the sources of your actions pure or impure”. On the basis of a carefully built interdependence between the self-consciousness implicit in reflection-c and our engaged acts of attention, Merritt argues that I know my moral character precisely by “considering what figures in my attention as providing me with reasons for doing this or that” (201).

---

2 See, for example, Kohl (2019, p. 728), Marshall (2019, pp. 1008-9), and Callanan (2020, pp. 152-3).
Properly unpacked, this account lends support to the idea that I can grapple with the law and recognize my moral character through a world-directed mindset focused on figuring out what to do. Such an account does not preclude demonstrating the necessary concern for whether one is mistaking merely subjective grounds of judgment for objective ones, as some reviewers have worried; according to Merritt’s Kant, it is the most fundamental way to demonstrate such concern.

On my view, two types of recurring passages from Kant’s corpus speak in favour of Merritt’s reading. First, as she herself notes, Kant’s insistence that we can never actually know whether our maxims are pure or impure suggests that the duty to know oneself does not include a dedication to turn inwards and ferret out one’s maxims to measure them against the moral law (200). Second, Kant repeatedly acknowledges that the abstract formulation of the moral law which guides reflection in its practical application is simply not a feature of our everyday, non-philosophical moral consciousness. At the close of section I in the *Groundwork*, he insists that practical reason “does not think [the moral law] so abstractly in a universal form” but rather has it “always before its eyes”, using it “as the norm for its appraisals” (G 4:403-4). The value of Merritt’s view lies in its ability to explain how we can at once respect Kant’s insistence that everyday moral engagement has no need of the moral law *in abstracto*, and still show its normative force to be operative. For on her view, “the moral law is grasped *in concreto* in the recognition of how particular facts about one’s situation give one reasons to do and not do certain things” (192).

If the text alone does not unilaterally support the ‘step back and introspect’ model of virtue, one might search for a deeper account of why it still finds vocal, if qualified, support. I would argue that this is at least in part because, on the widespread view of Kant that Merritt criticizes, there seems to be no other role for the moral law. On this widespread view, Kant’s insistence that the moral law is available to all of us is taken to imply that it is easy enough to know what it requires us to do, meaning the true obstacle to virtue lies in having the strength of will to do it, both *simpliciter* and from a truly moral disposition. But in the Doctrine of Virtue and elsewhere, Kant gives us plenty of cause to question this default assumption in a way that supports Merritt’s account. In such passages it emerges that having a grasp on the abstract requirements of morality is not enough. It might be relatively easy to know that we should not ask for exceptional treatment, that we must treat others as we demand to be treated, and that, accordingly, we must cultivate the disposition within ourselves that brings us to respect others and value their happiness. But this does
not mean we will always be able to recognize when these commitments are relevant in concrete cases, and still less does it mean we will always know what to do to fulfill them. One’s general moral knowledge needs, in addition, the concrete acts of attention that strive to figure out what to do and why to do it. “Moral purity” is revealed in the skillful ability to arrive at an objective picture of the whole situation and what it calls for. Impurity is revealed insofar as one is not skilled enough to see past the subjective feelings and desires that distort one’s point of view.

Kant arguably acknowledges this role for the moral law throughout the Doctrine of Virtue, insofar as he identifies duties of virtue as wide duties admitting of latitude (MS 6:390, 411). A common way of reading this claim is to conclude that since, in many cases, moral maxims fail to stipulate what precisely to do, we must conclude that duties of virtue are vague at best, no matter how virtuous one becomes. Since the abstract commands of morality cannot fill in the details, our subjective desires provide the only way to determinate action and so there is a permissive latitude to do as one happens to want. But there is another way to interpret Kant on this point, and it stands in direct support of Merritt’s view of Kantian moral virtue as an engaged skill. Consider the following passages, both from MS 6:411:

“Ethics, because of the latitude it allows in its imperfect duties, unavoidably leads to questions that call upon judgment to decide how a maxim is to be applied in particular cases, and indeed in such a way that judgment provides another (subordinate) maxim...So ethics falls into a casuistry”;

and,

“Casuistry is not so much a doctrine about how to find something as rather a practice in how to seek truth.”

I read the first passage as implying that we are obligated to use our judgment to work out how the moral law applies in particular situations. The second passage drives home that in such contexts, it does make sense to seek determinate truths, and that grasping them is a matter of practice and using one’s cognitive agency well. So, the fact that the moral law prescribes only formal maxims instead of particular actions does not mean there are no determinate situational facts to discern in connection with one’s moral duty. It means, rather, that there is no instruction manual for one to passively absorb and deploy, that it is up to the agent to develop the skill required to have a determinate, context-sensitive

---

7 Kohl mentions this reading of Kant’s account of wide duties as a reason to doubt that there is any determinacy to be gained through skilled practical engagement with the world (Kohl 2019, p. 730).
conception of what to do in accordance with the norms of pure practical reason. The virtuous person is precisely one who has a judgment cultivated enough to do this.

This leads directly into the second controversial implication of Merritt’s account. Making sense of the passages where Kant talks about moral strength, she insists that “virtue is strength of the will, or practical reason: for Kant, this means that only a notional distinction can be drawn between cognitive and motivational aspects of virtue” (187). Rather than thinking of motivation as a separable volitional force, Merritt builds the argument that having sufficient motivation to act from the moral law depends on the determinacy of one’s practical knowledge. Having determinate knowledge, in this case, amounts to having the good moral discernment just discussed: one gains determinacy by being skillfully attentive to the relevant concrete practical factors. The difference between the virtuous and unvirtuous person is, therefore, best articulated in terms of the quality of their knowledge (192), and so must be the corresponding disparity in motivation. Merritt thus continues to question the reading of Kant that presents moral knowledge as easily won and wholly separable from the issue of motivation, arguing instead that “the determinacy of such knowledge, and so its ability to be concretely action-guiding, admits of degree” (192). I have argued for a version of this view elsewhere, in connection with Kant’s doctrine of the practical postulates and their bearing upon moral motivation (Tizzard 2020). To my knowledge, Merritt is the only other Kant scholar to claim so explicitly that practical knowledge and motivation are wholly interdependent. I think such a view is best described as the strongest or most resolute version of cognitivism as applied to Kant’s practical philosophy. Even though it seems to follow neatly from his equation of the will with practical reason (G 4:412), there is bound to be ongoing resistance to this view. It would be overly simplistic to reduce all possible objections to a single overarching point, but the widespread history of treating practical reason as a separate volitional capacity that is not as equally grounded in the faculty of cognition as its theoretical counterpart must be an important factor. Merritt provides a useful discussion of this tradition and the textual evidence against it in Chapter 4.

Further questions and concerns

Having shown the significance of Merritt’s view, I will close by raising two concerns pertaining to its details. As with any comprehensive interpretation of Kant, there are a number of places where further expansion would be helpful to Merritt’s reader. Given the nature of her account, I would especially like to know more about her view of Kant’s

---

8 For example, in the recent and valuable work *Kant on Evil, Self-Deception, and Moral Reform*, Laura Papish comes close to this view by arguing that overcoming vice and cultivating virtue requires focusing on the cognitive aspect of moral life, and building a more complete and determinate understanding of one’s moral commitments. But she still ultimately maintains that motivation and knowledge are more than just notionally separable: “an agent who lacks full understanding does not lack the wherewithal or executive power to act on her maxims” (Papish 2018, p. 196).
distinction between imperfect and perfect duties. An interesting textual puzzle emerges in the Doctrine of Virtue insofar as Kant states that “ethical duties are of wide obligation” and “imperfect duties alone are, accordingly, duties of virtue” (MS 6:390), but then goes on to talk about negative duties of omission, which are typically thought to be narrow, perfect duties (e.g., MS 6:421). The question thus arises, does Kant think that addressing the perfect duties to ourselves and others discussed in the Doctrine of Virtue requires the same cultivation of judgment that he mentions explicitly in reference to imperfect duties (MS 6:90)? Whenever the contrast between perfect and imperfect duties explicitly arises, Merritt suggests that the answer is ‘no.’ For example, on the duty to respect others, she remarks: “since everyone is owed recognition respect regardless of merit, the fulfilment of duties of recognition respect is not especially context-sensitive, and should not require any remarkable discernment” (194). Granted, one never needs to ask “should I respect this person now?”, while in contrast, it could make sense to ask, “should I attend to this person’s happiness now?” But I don’t take this obvious difference to preclude the possibility that discernment is called for in discharging perfect duties. To see this, we can look to Kant’s account of the duty not to lie. It is a well-known tenet of Kantianism that one must always avoid lying, but Kant still leaves room for the possibility that discernment will be called for in assessing whether some particular action counts as a lie or not. This comes out especially clearly in the casuistical questions that attend his analysis of the duty (MS 6:431). Moreover, in his discussion of friendship, Kant discusses a case where one friend points out another’s faults in accordance with a duty of love, but in doing so is thought by the recipient of this criticism to have violated a duty of respect (MS 6:470). Despite important differences between negative and positive duties, it thus seems that there can still be room for skillful discernment in cases of the former. This suggests the distinction between different duties of virtue could be thought of as falling along a spectrum, with the widest duties of virtue on one end and the narrowest on the other. Such a view would, I think, strengthen and further support Merritt’s account, since it speaks to the complex agency and attendant difficulties involved in the experience of concrete particulars. Insofar as the reflective character of one’s agency is relevant to freely adopted duties of virtue, but not coercible duties of right, it would make sense that all duties discussed in the Doctrine of Virtue presuppose some level of attentive discernment on the part of the agent actively pursuing them.

Second, and more importantly, I would argue that Merritt’s account of the Kantian conception of affect has radical implications for her theory of reflective agency that warrants more discussion. Her remarks on affect are introduced as a way into the discussion of reflection-c at the very beginning of the book, and I found myself wishing she would return to them in more detail as her view developed. Kant certainly ties affect closely to some kind of reflective failure, remarking that “preceding reflection, it makes

---

9 See also (Merritt 2018, p. 146) for some remarks on the difference between imperfect and perfect duties.

10 There is solid textual evidence for the spectrum view at MS 6:390: “the wider the duty, therefore, the more imperfect is a man’s obligation to action; as he, nevertheless, brings closer to narrow duty (duties of right) the maxim of complying with wide duty (in his disposition), so much the more perfect is his virtuous action.”
this impossible or more difficult” (MS 6:408). In his lengthier account of affect in the *Anthropology*, he describes them as “feelings of pleasure and displeasure that transgress the bounds of the human being’s inner freedom” (A 7:235), stating that affect is “thoughtless” in its capacity to inhibit reflection. From such remarks, Merritt draws the striking conclusion that the affected subject cannot even be said to engage in reflection: “affect involves losing even that self-consciousness that is implicit inasmuch as one judges or acts intentionally at all” (20). This leads to earnest questions about the demandingness of reflection and the regularity with which affect is encountered in practical life. To the first point, Merritt’s reading of Kant certainly builds a lot of complex cognitive activity into the notion of reflection, but it also presents this activity as implicit in our everyday openness to a world of publicly available objects. While it is easy enough to go wrong in judging about this world and thereby fail to meet the requirements of reflection, the cognitive agency constitutive of reflection—which Merritt finds implicated in even the simplest acts of attention, and stirring in us from the very first stages of life (83, 89)—seems much harder to turn off once it has come online. While I would not dispute that such occurrences are possible, they seem to be exceptional cases at the margins of life, as opposed to a recurring obstacle to virtue against which we must continually ready ourselves.

Though there are certainly passages that support Merritt’s reading of affect, it should be noted that, as is usually the case with Kant, there is room for alternative accounts. Kant himself can sometimes seem aware of the tensions involved in saying both that practical failures are failures of reason, and that affect prohibits reflection entirely. Merritt takes Kant to overstep the limits of his own account in speaking of the affected agent as someone who ‘gives himself over to a feeling of pain’ (A 7:254; Merritt 2018, p. 20). Even this description attributes too much agency to one’s failure if reflection is nowhere present in it. Merritt’s strategy in dealing with apparent inconsistencies in Kant’s account is to take literally his talk of affect as thoughtless and unfree, and downplay the remarks that implicate cognitive agency. This is a questionable strategy, especially given that Kant also describes passions as afflictions that shut out the sovereignty of reason and limit the free use of one’s agency (A 7:251, 253), while also explicitly claiming that they involve the formation of maxims and therefore reasoning to some extent (A 7:266).

In sum, the implications for thinking of affect as a total breakdown of rational activity are serious, not least of all because it leads one to wonder how someone could ever be held responsible for the behaviour it produces, or strive to overcome such weaknesses through the cultivation of virtue as a cognitive skill. It seems further discussion is needed to persuasively establish Merritt’s view of affect, especially given that she introduces a two-pronged account of reflection, and failures of affect could also be construed as a different way of failing to meet the requirements of reflection.

---

11 Such agency is even implicated in Merritt’s account of the obscure representations that form the whole against which one abstracts some part for attention (Merritt 2018, p. 96).
I expect these will be just a few of the questions and concerns posed to Merritt in the years to come, as her book is highly deserving of sustained and dedicated discussion. I am hopeful that through such efforts, the resolute cognitivism and world-directed account of virtue she recommends will become more established positions in the world of Kant scholarship.

*Special thanks to Charlie Crerar, Lionel Shapiro, Paul Bloomfield, and Yuhan Liang for helpful thoughts and lively discussion shared in a reading group centered on this book at the University of Connecticut during Spring semester 2020.

References


