SGIR Review

Published by the Society for German Idealism and Romanticism

Volume 2, Issue 1 *New York City, NY 2019*

ISSN 2577-0233 (print)
ISSN 2577-025X (online)

SGIR Review

Symposium Editor
Lydia Moland, *Colby College*

General Editor
Gerad Gentry, *Lewis University* and *The University of Chicago*

SGIR Advisory Board
Frederick Beiser, *Syracuse University*
Daniel Breazeale, *University of Kentucky*
Hannah Ginsborg, *University of California–Berkeley*
Kristin Gjesdal, *Temple University*
Johannes Haag, *Universität Potsdam*
Rolf-Peter Horstmann, *Humboldt-Universität zu Berlin*
Stephen Houlgate, *University of Warwick*
Michale Forster, *University of Chicago* and *Universität Bonn*
Eckart Förster, *Johns Hopkins University* and *Humboldt-Universität zu Berlin*
Manfred Frank, *Universität Tübingen*
Paul Franks, *Yale University*
Pablo Muchnick, *Emerson College*
Terry Pinkard, *Georgetown University*
Anne Pollok, *University of South Carolina*
Konstantin Pollok, *University of South Carolina*
Tobias Rosefeldt, *Humboldt-Universität zu Berlin*
Sally Sedgwick, *University of Illinois at Chicago*
David Wellbery, *University of Chicago*
Rachel Zuckert, *Northwestern University*
Günter Zöller, *Universität München*

The *SGIR Review* is published by the Society for German Idealism and Romanticism and registered with the United States Library of Congress. It includes symposium reviews and responses written for SGIR Author-Meets-Critics Sessions at annual American Philosophical Association conferences (Eastern, Central, and Pacific) as well as stand-alone reviews of recent publications. The *SGIR Review* cooperates with other journals in the field such as the *Kantian Review*. At times, the *SGIR Review* provides these journals with one of its own reviews from a symposium book review series. In such cases, we note the copyright transfer in the bibliographic portion of the review and reproduce it here with the permission of the given publisher (e.g. *Cambridge Journals*).

The aim of the *SGIR Review* is to provide scholars with a source for in-depth discussion of recent work related to German Idealism and Romanticism.

© SGIR Review, New York City, NY - 2019

Table of Contents

Preface ii

SGIR Author-Meets-Critics Book Symposium, 2019 Eastern APA:

1. Dean Moyar on Andreja Novakovic's *Hegel on Second Nature and Ethical Life* 1

2. Mark Alznauer on Andreja Novakovic's *Hegel on Second Nature and Ethical Life* 13

3. Andreja Novakovic on *Hegel on Second Nature and Ethical Life*: Response to Dean Moyar and Mark Alznauer 27

SGIR Author-Meets-Critics Book Symposium, 2019 Central APA:

4. Colin McLear on Melissa Merritt's *Kant on Reflection and Virtue* 42

5. Francey Russel on Melissa Merritt's *Kant on Reflection and Virtue* 60

6. David Sussman on Melissa Merritt's *Kant on Reflection and Virtue* 72

7. Melissa Merritt on *Kant on Reflection and Virtue*: Response to Colin McLear, France Russell, and David Sussman 84

Novakovic, Andreja

Hegel on Second Nature and Ethical Life

Andreja Novakovic. *Hegel on Second Nature and Ethical Life*. Cambridge: Cambridge University Press, 2017, 234 pp., US $95.36, ISBN 1107175968

Reviewed by Dean Moyar

Andreja Novakovic's approach to Hegel's practical philosophy takes criticism, and criticism of criticism, as some of its central themes. Though my role here is critical, I have no inclination to engage in the overblown "critical criticism" that Novakovic takes to be one of Hegel's main targets. For I am in broad agreement with Novakovic's interpretation. I concur wholeheartedly that the *Philosophy of Right* can and should be supplemented with material from the *Phenomenology of Spirit*. I find her claims about the various standpoints – ordinary, reflective in a bad abstract sense, critical, and scientific – highly illuminating and productive for getting a grip on Hegel's project. She has anticipated and countered many of the lines of possible dissent from her interpretation, yet there are several points on which I am uneasy with her claims, or find that I am not sure that I understand them. I am going to work backwards through her text, starting with her take on Hegelian science and ending with her treatment of individual rational agency. In good Hegelian recollective fashion, this will allow me to read back the consequences of her later claims into the basic issues with which she starts.

Novakovic orients her reading of Hegelian science around an interpretation of the famous *Doppelsatz* from the *Philosophy of Right* Preface. "What is rational is actual; and what is actual is rational." In partial agreement with Robert Stern's "neutral reading," she argues that no substantive normative claims are being made in the *Doppelsatz* (176-77). She thus advocates along with Stern a *methodological* reading of the *Doppelsatz* according to which Hegel is just announcing his philosophical rationalism. Novakovic disagrees with Stern in that she holds that the *Doppelsatz* is evaluative. Stern holds that the actual is rational in so far as we can comprehend it, but that such comprehension does not necessarily bring with it a positive evaluation of that actuality.[1] Novakovic, by contrast, holds that what in scientific terms we find to be actual and rational is

[1] Stern, "Hegel's Doppelsatz: A Neutral Reading."

thereby evaluated as good. In locating its rationality we are just thereby locating its goodness and actuality. She goes on to make the striking claim that the much-celebrated and much-criticized *Doppelsatz* is in fact *trivial*, or true by definition. I believe that what she means in calling it trivial is that there is simply no difference on Hegel's view between actuality and rationality. Many scholars have noted that with "actuality" Hegel does not mean whatever reality happens to hold at a given point in time, since actuality is already a normative standard of some sort. Novakovic takes this line one step further in flat out identifying actuality and rationality, often writing "actuality/rationality" to emphasize the point. She opposes any attempt at "prying actuality and rationality apart" (178, note 23). But instead of using this identity to give a progressive normative reading like the one Stern criticizes, Novakovic argues that the *Doppelsatz* does not have progressive consequences. This leads her to criticize the alternative version of the *Doppelsatz*, much cited by progressive interpreters, from the lectures as incoherent. The alternative version from the lectures – "The rational becomes actual, and the actual becomes rational" – is incoherent on her view because it presupposes a difference between the two and Hegel is committed to their identity.

While I agree that the *Doppelsatz* is evaluative and is fruitfully read as a methodological claim, I think there is something amiss with the triviality claim. Hegelian identity typically implies difference as well, so even the simplest seeming identity claims will turn out not to be simple at all. This seems to be especially true in the case of the *Doppelsatz*. Just look at how Hegel introduces the claim in the *Philosophy of Right* Preface. He is discussing Plato's *Republic* as "a proverbial example of an *empty ideal*," and after arguing that Plato aimed to ward off the rising drive of "free infinite personality" with a "particular *external* form of Greek ethics," he then comments, "But he proved his greatness of spirit by the fact that the very principle on which the distinctive character of his Idea turns is the pivot on which the impending world revolution turned."[2] Then comes the *Doppelsatz*. Can Plato's greatness of spirit be established by his appreciation of a trivial identity claim? Hegel reads Plato as wanting to establish that the forms, encompassed by the form of the Good (which together Hegel assimilates to "the Idea" in roughly his own sense), are the proper basis of politics, of the actual life of the state that is to be ruled by philosopher-guardians.[3] He also

[2] Hegel, *Philosophy of Right*, 20.

[3] See the long remark to Hegel, *Encyclopedia* §552.

writes in E §6 that this *Doppelsatz* captures the religious conception of the divine governance of the world, or providence (which is in fact the form better suited to representational consciousness). He is saying that the Idea, seemingly a theoretical abstraction, is necessarily and emphatically practical.

I believe that there is a difference implied in the *Doppelsatz* between the *rational* as the conceptual or logical and the *actual* as the organized world of freedom. Consider how he continues in the *Philosophy of Right* Preface a few (important) sentences later: "For since the rational, which is synonymous with the Idea, when it enters into its actuality at the same time enters into external existence [*Existenz*], emerges in an infinite wealth of forms, appearances, and shapes and surrounds its core with a brightly coloured covering in which consciousness at first resides, but which only the concept can penetrate in order to find the inner pulse, and detect its continued beat even within the external shape."[4] This passage suggests that the rational stands for the *logical* Idea, and that actuality is split between that logical dimension and the dimension of external existence. If the rational can enter actuality then it presumably has some standing prior to actuality. If there are better and worse actualities, it seems that rationality would be a way to determine which is better and why.

Now I am not sure whether my objection here reflects a deep disagreement. Novakovic does, after all, have an important place for a dynamic development within the social order. She writes that "the ethical world is undergoing an objective process of actualization, a process of *becoming* actual/rational, and that our philosophical comprehension of the actuality/rationality of the ethical world corresponds to that stage in this developmental process" (178). Given this endorsement of "developmental unfolding" (178, n23), it is unclear to me why exactly we should deny that the actual *becomes* rational. I take it that she holds that "the ethical world" is different from "the actual" because the former has a certain merely positive or contingent character. I agree with her that when we are looking for rationality in the world, finding the rational is the same as finding what is actual. But actuality includes the contingent and messy appearances (or at least a relation to the appearances), whereas the rational need not. So there remains a sense in which the actual can become more rational.

[4] Hegel, *Philosophy of Right*, 20-21, translation altered; Nisbet adds to the confusion by rendering "*indem es in seiner Wirklichkeit zugleich in die … tritt*" as "becomes actual."

In the end I suppose I am just not sure what the status of the logic is in her account. Novakovic takes it to be important that actualization is itself a kind of criterion that rules out certain kinds of formal deductive strategies (such as Fichte's). In part I cannot gauge the radicality of her thesis that the rational just is what is actualized, what proves itself in history. She may be adopting the thesis that the logic itself is grounded in historical actuality. Or, she may be adopting the view that the account of the Idea in the Logic is so indeterminate that the full rationality of the Idea of Life or the Idea of the Good is only to be found in the philosophy of spirit.[5]

While she does not address the logical idea of life directly, life does play an important role in her discussion of "unsustainability" in Chapter 3's treatment of immanent criticism and practical contradiction (141ff). She writes that "it is often not easy to see that two principles are incompatible until we try to live by them" (143). It is in acting on them, living with them, that their contradiction comes to light. Novakovic draws this claim from the method of experience in the *Phenomenology*. As examples of the method she cites the master-servant dialectic and the breakdown of Greek ethical life that Hegel sets out through a reading of Sophocles' *Antigone*. She thinks that Hegel does not utilize this method in the *Philosophy of Right*, but she argues that the case of poverty in civil society is a case of practical contradiction in the modern world. Her other main example is the practice of segregation in the post-Civil War United States and the rejection of that practice expressed in the Supreme Court's Brown vs. Board of Education (of Topeka) decision. The actual living out of the "separate but equal" principle demonstrated its unsustainability. Such cases, and such judgments, are the model of "immanent criticism" that we can salvage from Hegel's method even though it is not Hegel's mode of argument in the *Philosophy of Right*.

In trying to pin down this conception of criticism, Novakovic distinguishes it from a model of "internal criticism."[6] The internal critic takes an institution to task for "failing to live up to its own ideals" (116), and thus "objects to practices for violating the standards they espouse" (116). The ideals at issue are sound but they are not being realized in an appropriate way. Novakovic

[5] This would take us quickly to the issue of how to read Hegel's claim that the *Phenomenology* is the "deduction" of the standpoint of science. One could read this claim as saying that the Logic depends on the experience of consciousness for its objective validity, which would be a strong sense in which the rational (the logic) is dependent on the processes of life and history.

[6] She draws here on Jaeggi, *Zur Kritik von Lebensformen*.

argues that this internal model is different from immanent criticism in two main ways: first, internal criticism only works when the failure of the application of standards "has no implications for the validity of the standard itself" (116). That is, sometimes a failure of standards in practice brings out a fault in the standards rather than in the application, and that kind of failure is better accounted for on an immanent criticism model of practical contradiction. The second difference between the two forms of criticism concerns certain epistemic assumptions. On the internal model, we already know what the standards are that we appeal to in criticizing the institution. By contrast, immanent criticism brings out commitments to standards that the members of the institution might not have been aware of. Something fundamental is revealed in the process of immanent criticism itself. This second element brings to light Novakovic's view that immanent criticism is a social process – she writes of "what it means for a society to undergo immanent critique" (132).

Novakovic's leading examples of immanent criticism in the *Phenomenology* come from crucial, but relatively early (temporally and conceptually) stages in the development of *Geist*. The master's attempt to dominate fails (though it is unclear whether he actually learns this lesson) and Antigone and Creon learn the interdependence of the divine and human laws, setting the terms for Roman right. The practical contradiction that Novakovic appeals to is quite strong: the very application of the norm inverts itself, so that you end up, in following the norm, doing the opposite of what you intended.

Novakovic considers and rejects the idea that such immanent criticism is only something in the past, a relic of more immediate phases of history.[7] But what of the norms that are stated in the Abstract Right and Morality sections of the *Philosophy of Right*? In one sense I think there is an immanent criticism of the norms – there is an inversion at the transition points to the next norms. Right becomes wrong, morality becomes evil. But it does not seem that these transitions count as affecting the norms themselves. Rather, they highlight the need to supplement those norms with others, or to bring them into a different context.

My concern is that the norms of personhood and subjectivity, and their attendant rights or freedom, have such a bedrock status in the modern world that the immanent criticism model may no longer be effective or, for that matter, desirable. It is not that criticism is impossible; it is just

[7] Perhaps I should say "partly rejects," since she does say that the modern system of right, with its multiple freedoms, makes it less liable to full-blown contradictions. See 147.

that such criticism will look more like internal than immanent criticism. And if the immanent criticism model is supposed to be able to radically undermine our current standards, we should think twice about accepting it.

With this issue in mind I would like to ask: is the Brown v. Board of Education decision really an example of immanent criticism rather than internal criticism? Novakovic's main reason for thinking of it as immanent criticism is that the practice of a system committed to "separate but equal" demonstrated that the principle is conceptually incoherent (144-46). Only in that demonstration and its empirical evidence does it become clear to the people advocating segregation that it is "inseparable from, racist assumptions about the superiority of one race over another" (145). But the reconstruction of the scenario required to make this work seems implausible: we would have to imagine that the segregationists were good-faith agents who honestly believed that equality was possible with segregation, but that upon seeing the practice play out came to understand that it was not only not equal, but that their own racist assumptions were to blame. On this retelling the Supreme Court just clarified this situation for them, showed them the practical contradiction, and thereby extinguished the institution. If we read this case on the internal criticism model, on the other hand, the reconstruction would be more plausible. The segregationists were simply, and obviously, never committed to the value of equality for all. The standard of equality is sound, and it is the application in the case of separate but equal that was faulty. Without that federally enforceable claim of equality the social dynamics in the segregated South might have been sustainable for quite some time.

Even if we did only have internal criticism at our disposal, I think that would not reflect so badly on Hegel's philosophy. Hegel did, after all, broadly support the French Revolution and the Napoleonic reforms, and he did so on the basis of the freedom, or right, that he makes essential to his own story (though his critique of the absolute freedom of the revolutionary ideology complicates this picture – more on that below). Perhaps what Novakovic thinks would be missing without the resources of a more radical immanent criticism is something on the side of normative force rather than on the side of content. I have the sense that part of what she is after is a kind of contradiction that has *transformative force* for the institutions themselves. If a society "undergoes" immanent criticism then it will be aware of, and transformed by, the realization of a better set of norms. I sympathize with this concern, for look again at the Brown v. Board of Education case. I think this was internal criticism backed up by federal power, but it is not clear that even today,

sixty-five years later, the decision has been embraced. School segregation is still with us even though it lacks legal validation. So I agree that something more is needed, but I am not sure that immanent criticism is the right way to think of the move to a normative awakening.

One could look for this transformative force in Hegel's treatment of culture, *Bildung*, which is the subject of Chapter Two of Novakovic's book. She introduces the concept as a way of mediating between mere habit and a more reflective form of social engagement. It would seem – and this is why communitarians often look to Hegel for inspiration – that Hegel attempts to solve the problem of the emptiness and thinness of modern morality with a return to a thicker and more particularistic cultural identification. Rightly noting that Hegel does not dwell on this issue in the *Philosophy of Right*, she turns to the *Bildung* chapter of the *Phenomenology* to fill out the concept. She focuses her account of Hegel's chapter on the idea that the self-cultivation and self-alienation of *Bildung* are intimately connected. She holds that the main lesson from early modern culture is "its irreducible malleability" (85), and she notes that this does not exactly help with the positive question of how attachment to shared cultural norms is established. For a more constructive account she turns to Hegel's conception of work, especially the work of the servant in the famous Chapter IV of the *Phenomenology*. Her account focuses on the "distance" that the worker acquires in producing an object, and she aligns that with "reflective distance" to show its relevance to the broader argument she is making. She writes, "Only when he is confronted with something other than he that is at the same time of his own making can be attain self-knowledge and find confirmation of his own self-conception" (90). This is structurally similar to the dynamics of alienation and externalization of the Bildung chapter, for both have a structure that increases reflective self-consciousness through an activity in which one is at once distanced from oneself and able to see oneself in the world. Novakovic helpfully connects these aspects of the *Phenomenology* to Hegel's treatment of *Bildung* in Civil Society, where the subject achieves a universal standpoint that is moderated by inclusion in a particular estate. As modern agents we cannot help but have a reflective attitude towards our own cultural participation, though she stresses again that in Hegel's view these forms of cultural participation will be lived out as a second nature, that is, without constant reflection on "the roles themselves as roles" (104).

While I think this account is quite right as far as it goes, I am a little troubled by Novakovic's selective use of the *Phenomenology* account. Hegel himself presses the crucial issues in the sections that follow the initial phase of "Culture" in ways that complicate her overall

reconstruction. The specific treatment of alienation that culminates in the account of *Rameau's Nephew* is only the first phase of "Culture." The central part of the chapter concerns faith and insight, or superstition and the Enlightenment. Hegel's relationship to the Enlightenment is complicated, but given the affinity of reflection and Enlightenment thought it would be a natural discussion for Novakovic to thematize. It is true that this episode is not repeated in any clear way in the *Philosophy of Right*, where he is much more concerned with what he sees as anti-rational appeals to feeling and conviction. But the criticism of excess rationalism in the *Phenomenology* is not a criticism he revokes, as his later lectures on religion clearly show.[8] The Enlightenment views religion as the exercise of power by priests and despots who deceive the masses into accepting a false view of the world. Hegel criticizes this Enlightenment posture on the grounds that religion, and in particular religious practice, is not the kind of attitude about which a people can be deceived. The meaning that religion finds in its worship is immune to the attacks based on a theoretical stance, on appeal to natural science or to historical evidence. This is an external criticism that does not enter enough into the practice's viewpoint to criticize it effectively. I think that Novakovic could have used this point to refine her account of rational reflection and immanent critique. For the kind of critique that the Enlightenment does successfully bring to bear is closer to what she calls immanent critique: the faithful keep two sets of books, two accounts of value, and this is a practical contradiction that leads to the collapse of a certain kind of otherworldly-oriented religion. But not of all religion, for Hegel does maintain the truth of Christianity, specifically Protestant Christianity. One question for Novakovic's account is whether that element of Hegelian philosophy of spirit would be one she would hope to save, or that Hegel needs to save, from excessive reflection.

The abbreviated reading of the Culture chapter also misses the positive resolution of Enlightenment's culture of reflection. Novakovic could have fleshed out the link of culture and civil society if she had taken into account this positive contribution – namely *utility*. Within the logic of the *Phenomenology* Hegel does endorse this truth as the moment of relation within the practical domain, the element of determinate existence or finite value. There is a utilitarian or consequentialist dimension to Hegel's practical philosophy, and the account of mediation between

[8] See, for instance Hegel, *Lectures on the Philosophy of Religion* III, 246-247, where we find: "The Enlightenment – that vanity of understanding – is the most vehement opponent of philosophy. It takes it very ill when philosophy demonstrates the rational content in the Christian religion …"

individual and universal elements is hard to comprehend without taking it into account. In my view this element makes Hegel's overall conception much more dynamic or flexible, and thus rather less subject to certain forms of critique. The account that Novakovic gives of work does capture an important element of seeing oneself in the world. But Hegel is also perfectly happy with employing less tangible forms of expression. He thus celebrates the fact that we serve our state through the payment of taxes, in money, rather than through providing a specific service.[9] He seems to embrace this melting of obligation into the monetary form, and he thus seems to think that attachment to particular cultural forms is compatible with a much thinner way of participating in politics.

In the *Phenomenology* Hegel had recognized that the triumph of Enlightenment utility leads to a revolutionary stance in which all value is seen to be subject to political adjudication. I mention this because I have some trouble locating Novakovic's position on the rationality of civil society. There is an aspect of typical ideology critique that comes close to the logic of the French Revolution and its abolishment of a separate civil society. Novakovic clearly wants to put a check on that logic while also preserving a more radical form of critique than the *Philosophy of Right* seems to offer. Habit and second nature are moderating forces in her account, yet I am not sure if they give us grounds – and if she wants them to give us grounds – to criticize civil society or to criticize the Marxist critics of civil society.

Hegel's analysis of the French Revolution and reign of terror is familiar, but the resolution of that analysis remains somewhat obscure. It is that resolution that I would like to thematize in my final discussion – of Chapter 1 – as I think it is central to how we read the status of morality in the overall account. The Revolution sought to subject the entire world to the standard of abstract universality contained in the idea of the general will. The Revolution identified the individual and universal and could not come to terms with the difference, the particularity, that any developed ethical world must possess. Hegel portrays its failure as resulting in the moral worldview of Kant and Fichte, and, I have argued elsewhere, in his own conception of conscience.[10] The problem is that this seems to leave us in the *Phenomenology* stranded at the end of "Morality" with no clear way of coming back to Ethical Life. Novakovic thinks morality cannot be a legitimate successor

[9] Hegel, *Philosophy of Right,* §299.

[10] Moyar, *Hegel's Conscience*.

to earlier shapes of spirit. As she writes, "The perspective of conscience is not a full-fledged perspective at all, not a point of view that grants us substantive insight into how to proceed, should our norms come into question" (107).

Yet morality and conscience could play a greater role in uniting ordinary practice and philosophical knowledge than Novakovic accords them, and could in fact be enlisted to support key elements of her reading (her use of "true conscience" in Chapter 1 already moves in the direction, but I think that more could be said). In the *Phenomenology* account it is Kant's moral teleology, centered around the Highest Good, that is the focal point of the resolution of the absolute freedom problem. The narrow morality of the categorical imperative is not the conception at work here, for that would indeed be hard to think of as another "shape of a world." It is instead the idea of the Highest Good, which is Kant's way of getting religious ideas such as God and immortality back into the picture. It is also one of the sources of Hegel's *Doppelsatz*, for it contains the core idea of providence that Hegel aligns with his own view. There is thus a path to thinking that Hegel's view of his own ethical world is a kind of reinterpretation of Kant's Highest Good.

We can better appreciate the way that Morality in the *Phenomenology* is a successor world to the French Revolution when we see that the move to the Good in the *Philosophy of Right* follows a similar revolutionary logic, and that the Good and conscience are the main structuring elements of *Sittlichkeit*. There the issue that generates the Good is the *conflict* between abstract right and individual welfare. Whereas the *Philosophy of Right* presents the Good as generated by the right of necessity, the *Phenomenology* presents the Good as the outcome of a revolution that attempted to *unite* abstract universality with individual welfare. The critique of the French Revolution does I think count as a form of immanent critique, yet one that resituates the standard (of universal freedom) within a constellation of freedoms. The moral worldview is the successor to the general will because it can incorporate into a holistic teleology of freedom the particularity of determinate, differentiated social reality. Morality is a higher universality that recognizes that universal purposes are only attained through finite willing. Switching back to the *Philosophy of Right* transition, the Good, which Hegel calls without irony the "final purpose of the world [*Endzweck der Welt*]" (§129), is an all-inclusive concept of freedom. The rest of the *Philosophy of Right* is an attempt to work out the realization of the Good. Right and welfare are supposed to be united in this conception, but beyond the general demand to do what is right and to promote welfare, there is not much here to guide action. That is why in the *Philosophy of Right* Hegel turns to conscience

as the *actuality* of the Good. The institutions of *Sittlichkeit* are "the living Good" (PR §142) in which agents of conscience inhabit contexts in which they do not have to make exceptions of themselves in order to be ethical.

Novakovic's Chapter 1 discussion of habit and character provides much new insight into essential aspects of Hegel's view, but I was somewhat surprised by her opposition in Hegel's name to John McDowell's Aristotelian view. She argues that Hegel's views on the codification of laws in the legal system imply that his overall ethical view is one of acting on principles rather than the situation specific holistic judgment of the Aristotelian virtuous person. But I do not think that the legality issues carry over so directly to ethical agency. Hegel has a role for the universality of principles in deliberation, certainly, but that universality is only a moment; action, actuality, requires something more specific. What Hegel refers to as the "evaporating" element of conscience is always there, but in a more moderated way in stable ethical contexts. It is required in order to arrive at a holistic judgment of the situation. Principles have prima facie obligatoriness, but Hegel is much more willing than Kant, for instance, to consider defeaters, and to accept that one may have to neglect certain morally relevant considerations in favor of others. Novakovic stays that McDowell has "strikingly little to say" about criticizing our second nature (63), but it must also be said that Hegel has strikingly little to say about what principles exactly are to guide our conduct. There is no list of perfect and imperfect duties, but only some very general demands and some formal rights that seem to leave much in the hands of ethical individuals. Of course Hegel polemicizes in favor of the law against mere conviction, yet he also memorably argues against the principled reasoning of law-giving and law-testing reason. I would like to hear more about the relation of positive law and ethical living on Novakovic's view, for I worry that she courts a kind of ethical legalism in her insistence on codifiability. Hegel hopes that we habitually obey the law, but all his ethical institutions are actual in practical judgments of individuals.

Conscience plays such a major role in the *Phenomenology* because it is the locus of the power of self-consciousness to bind oneself to quite disparate elements of a moral landscape. In the *Phenomenology* Hegel touts conscience as bringing *content* to the moral picture, and in the *Philosophy of Right* he argues that it brings *actuality* to the Good. But how does this work? The answer in the *Phenomenology* refers back to the three main moments of action in the development of spirit. Hegel states this quite definitely in presenting the concept of conscience in section 641 and in recapitulating the practical development in "Absolute Knowing" in sections 792-93. I

believe this lines up nicely with Novakovic's appeal to recollection at the end of her book, but since she does not foreground conscience in her discussion of recollection I am not sure whether we are on the same page here or not. This may be mostly a matter of terminology, but it is an important one, because if we do not hold on to Hegel's near identification of conscience and philosophy in "Absolute Knowing" it is harder to fill out Hegel's claim in the *Philosophy of Right* Preface that philosophers and ordinary ethical agents appreciate that rationality is not something apart from the world but rather something always actualized through our actions on the Good.

Bibliography

Hegel, G.W.F. *Elements of the Philosophy of Right*. Edited by Allen Wood and Translated by H.B. Nisbet. Cambridge: Cambridge University Press, 1991.

Lectures on the Philosophy of Religion, Vols. I-III. Edited by Peter Hodgson. Berkeley: University of California Press, 1984-85.

Gesammelte Werke, Ed. Hartmut Buchner and Otto Pöggeler. Hamburg: Felix Meiner Verlag, 1968–.

Jaeggi, Rahel. *Zur Kritik von Lebensformen*. Berlin: Suhrkamp, 2014.

Moyar, Dean. *Hegel's Conscience*. Oxford: Oxford University Press, 2011.

Stern, Robert. "Hegel's Doppelsatz: A Neutral Reading." *Journal of the History of Philosophy* 44:2, 2006, 235-266.

Novakovic, Andreja

Hegel on Second Nature and Ethical Life

Andreja Novakovic. *Hegel on Second Nature and Ethical Life*. Cambridge: Cambridge University Press, 2017, 234 pp., US $95.36, ISBN 1107175968

Reviewed by Mark Alznauer

Introduction

In the Preface to the *Philosophy of Right,* Hegel alludes to a well-known dictum of Francis Bacon's: "a little philosophy inclineth man's mind to atheism, but depth in philosophy bringeth men's minds about to religion." Bacon claims that to avoid falling into error about religion you should either never philosophize about it at all or you should continue until you have arrived at first principles. In *Hegel on Second Nature and Ethical Life*, Andreja Novakovic argues that Hegel's account of ethical life has the same ambivalent structure: he thinks that if you want to avoid error about ethical life, you should either refrain from explicitly reflecting on it or you should persist until you have arrived at a full philosophical comprehension of it. As with Bacon, the greatest danger faces those in the middle—the half-philosophical individuals who reflect too much to be content with ordinary beliefs about our duties and rights, but too little to see the inherent rationality of our form of ethical life. She argues that it is just this intermediary group—the atheists of the ethical world—who are the primary audience for the *Philosophy of Right*, and that the main task of the book is to help them remember something that they knew before they started reflecting, a task it accomplishes by providing a *science* of right as opposed to a mere *theory* of right.

This provides Novakovic with a powerful and judicious way of framing Hegel's project. It foregrounds Hegel's radical re-conception of the basic task of ethical philosophy, and immediately raises a series of difficult interpretive and philosophical questions, questions that Novakovic pursues with admirable clarity and with much success.

The first question is: in what sense do the non-philosophical, those who never explicitly reflect on their duties and rights, *already know* what is right and wrong? Her answer is that even if unphilosophical citizens never "step back" from ethical life, their habitual or customary behavior is not to be seen as blind or mechanical but as involving an implicit form of reflection: a sense of

the meaningfulness of their social rights and obligations, of what makes it worthwhile to continue to uphold them.

A natural follow-up question is: what is the danger of half-philosophy or theory--why do our initial attempts to explicitly reflect on our ethical life lead to a *forgetting* or *loss* of ethical knowledge? Novakovic argues that the problem with reflection is that it discounts or brackets comparatively unreflective modes of knowing in favor of an attempt to derive ethical standards from "exclusively subjective sources" (11). She suggests that if such abstract reflection does not become aware of its own limitations, it leads to a general skepticism about ethical life.

This gives rise to a third question, one very pressing given Novakovic's particular emphasis on second nature: if explicit reflection is usually pernicious in this way, does this leave us with no resources for criticizing society? On this point, Novakovic gives us what she thinks Hegel's answer *should have been* given what he had to say about immanent critique in the *Phenomenology*—which is that critical reflection is only warranted when the occasion demands it: when we experience genuine contradictions in our ethical life, contradictions between the principles that have implicitly guided us heretofore, and a reality which proves that they cannot be actualized. In these cases, though, it is not a specifically philosophical critique that is called for, but a form of critique that is continuous with the experience of the embedded, pre-philosophical perspective.

A final question thus poses itself: if philosophy is not needed either to navigate a rational social order, or to criticize an imperfectly rational one, then what is it good for? Why provide a philosophy of right at all? She answers that the *Philosophy of Right* primarily addresses itself to the plight of the half-philosophical; it offers a way for them to recollect the knowledge that they had—and still have at some level—but which they have bracketed and suspended in their attempt to provide a reflective justification for ethical life in terms of certain abstract principles. And it does this by scientifically demonstrating that the same abstract principles they refer to in order to criticize ethical life actually presuppose ethical life for their actuality and validity.

As this summary suggests, *Hegel on Second Nature and Ethical Life* covers a great deal of ground—there is no way to treat all the topics it addresses. I will focus on what I take to be its most far-reaching claim, which is that Hegel's *Philosophy of Right* is doing something radically different than the other moral theories of its time, something that has no real precedent in the history of Western moral philosophy. Although something like this has suggested before by other

commentators, I am unaware of any other statement of this position which develops it as systematically and with so much textual fidelity.[1]

Although the division is a little bit artificial, I think it is useful to divide Novakovic's account of what Hegel is doing into a negative and a positive proposal. The negative proposal involves reading Hegel as denying that we need a certain traditional kind of philosophical ethics or moral theory—one oriented towards providing standards which can help individuals navigate ethical life. The positive proposal is that this does not leave us with a view of philosophy as irrelevant to practical matters or as merely therapeutic, but gives us an entirely new conception of the ethical task of philosophy: that of providing a systematic account of our own moral beliefs which shows whether they are actualizable in the social world in which we live.

The negative proposal

We can start with the negative proposal. Novakovic argues that Hegel rejected one still common way of understanding the task of philosophy in the domain of ethical life—namely, the task of formulating a moral theory. At the risk of anachronism, let me point to Mill's utilitarianism as a good example of the kind of theory that she claims Hegel rejected. For Mill, the central task of a moral philosophy is to provide a criterion of right and wrong action. Mill argues that since the morality of an action is nothing but the application of a general law to the individual case, we need to know the general law or principle that determines the moral rightness and wrongness of actions in order to achieve the end of action. Although he admits that the principle he is going to put forward (happiness) is not entirely new, but already tacitly influences our decisions, he is confident that a philosophical treatment of this issue will help resolve certain dilemmas individual agents face in real life. For Mill, philosophy is directly relevant to the ordinary individual agent because it can help her make better decisions about what to do.

What is the evidence that Hegel rejected moral theory in this sense? Novakovic points to several passages which suggest just this. She shows that Hegel denies that philosophy should issue instructions on how the world ought to be; that he seems to praise uncritical identification with the ethical order; and that he clearly suggests that everyone already knows the truth about right (111-

[1] See, e.g, the brief treatment of Hegel in the title essay of Raymond Geuss's *Outside Ethics*.

15 and 161-67). In a notorious passage from his lectures on world history, Hegel puts the point even more unambiguously:

> The individual's morality will then consist in fulfilling the duties imposed on him by his social station; these can be recognized without difficulty, and their particular form will depend on the particular class to which the individual belongs...To try to define duty in itself is idle speculation, and to regard morality as something difficult to attain may even indicate a desire to exempt oneself from one's duties.[2]

These passages make it clear that the task of providing an abstract universal criterion to help ordinary individual agents make decisions is nowhere on Hegel's agenda. Indeed, he thinks that even looking for such a criterion can be a sign that we are unserious about our duties.

Although it is easy to show that Hegel rejected moral theory in the Millian sense, it is less easy to understand *why* he rejected it. According to some of Hegel's critics, the reason Hegel thought moral theory was unnecessary was because he thought ethical life was already perfectly rational and so there is no need to worry about whether we actually ought to comply to the duties of our social station. This is sometimes described as an excessively conservative position but I am not aware of any conservative philosopher or political theorist who ever held such an optimistic view of modern ethical life. At any rate, Novakovic rightly emphasizes that Hegel's stated position is not that *no* criticism of ethical life is necessary, but just that no *philosophical* criticism of ethical life is necessary.[3] Hegel says there is "quite a lot" worth criticizing in any given state, he only insists that it does not take a philosophical license to do so.

A second possibility is that Hegel rejected moral theory because he thinks of moral action as involving a situational awareness of what to do that is inculcated through habit rather than the merely cognitive subsumption of individual cases under general laws that Mill invokes. This would make Hegel's alternative to moral theory akin to that form of Aristotelian virtue ethics defended by John McDowell in a series of essays. Although Novakovic grants a strong similarity between Hegel and McDowell—namely in their shared belief that habit is a necessary component of ethical knowledge—she also sees an important difference, one that makes it unlikely that

[2] Hegel, *Lectures on World History,* 80.

[3] See Hegel, *Encyclopedia Logic,* §6R.

Hegel's rejection of moral theory comes from this source. Whereas McDowell thinks that moral action is ultimately *uncodifiable*, something that cannot be captured under any set of rules, Hegel is convinced that any situational awareness can and must be reproducible in a rule-like form.[4] So although Hegel agrees with McDowell's Aristotle that habit as essential to ethical knowledge, that is not because ethical knowledge is ineffably particularistic, or because no one can understand the relevant principles without the right upbringing. Hegel's point is more restricted: it is that "we cannot be said to know our duty unless we demonstrate a commitment to doing it, and we only demonstrate such a commitment through the habit of the ethical" (25). If this is Hegel's position, though, it would appear that moral theory is fully possible in the sense that the general laws of moral action can be formulated and *understood*, it is just that we cannot *know* what this theory teaches as our duty until we those duties have become second nature to us.

As this suggests, Novakovic's discussion of this issue is quite subtle, turning on a fine distinction between understanding and knowledge. Her basic point, that Hegel does not reject moral theory because he thinks ethical knowledge is uncodifiable is convincing (to me, at least). After all, Hegel does not say that the problem of defining duty is that it is enormously *difficult* to fully capture all of the nuances of every situation we might find ourselves in—he says, instead, that it is *easy* to know what you ought to do since such knowledge has already been unambiguously promulgated in public laws, morality, and religion. But Novakovic goes on to make a significantly more controversial interpretive claim: which is that Hegel thinks one cannot *know* one's duty unless one has the habit of conforming with it. Holding that position would make sense if, like McDowell, we were understanding ethical knowledge as situationally specific, or as a skill like knowing how to write (an example she treats in some detail). But if we are conceding that ethical knowledge *can* be fully codified in terms of general rules applied to individual cases then it is unclear to me in what sense these rules can only be genuinely known by someone who habitually complies with them. If our paradigm for ethical knowledge is a general principle like "don't break contracts," then are we supposed to think that as someone loses the habit of conforming with his contractual obligations he is somehow losing *knowledge* of that general principle? In what sense of knowledge would that be true?

[4] Novakovic's evidence for Hegel's rejection of uncodifiability is mostly drawn from Hegel's comments on law, but she plausibly argues that the point can be generalized to ethical knowledge (see Novakovic, *Second Nature*, 63).

A third possibility is that Hegel's skepticism about our need for moral theory is more like that of Bernard Williams than it is like that of McDowell. Novakovic does not discuss Williams in great detail but she indicates that he is the contemporary philosopher whose critique of moral theory is closest to Hegel's (111-12, 167). In *Ethics and the Limits of Philosophy*, Williams offers a wide variety of reasons to reject what he calls "ethical theory"—which he defines as an account of ethical thought which implies a general test for the correctness of basic ethical belief (or else implies that there cannot be such a test).[5] But his primary complaint, to simplify things a bit, is that such theories rely on a notion of moral obligation that is both philosophically dubious and which leads to a distorted, overmoralized view of ethical life: one that fails to recognize that the "machinery of everyday blame" is not self-standing, but only makes sense when surrounded by system of customs and institutions.[6]

Williams himself notes that this an essentially Hegelian point; he says Hegel was the first philosopher to worry that moral or ethical theory is "too far removed…from social and historical reality and from any concrete sense of a particular ethical life."[7] But there is an important difference between his reasons for worrying about this false abstraction and Hegel's. For Williams (and again, I simplify) the basic problem is a recognizably Nietzschean one: once you see how blame actually operates in social reality, you see that the moral system depends on various fictions about individual responsibility that cannot be sustained.[8] Williams thinks these fictions have some salutary consequences—they encourage people to identify more completely with the ethical order—but he thinks they also lead individuals to misunderstand their own fear and resentment against wrong-doing as the voice of a law that transcends history. Williams is not sure whether our practices of assigning praise and blame could fully survive a "reflective and nonmythical understanding" of these social mechanisms, but he is convinced that moral theory is no longer of any help to us in navigating these difficulties.

[5] To be more precise, Williams says "[a]n ethical theory is a theoretical account of what ethical thought and practice are, which account either implies a general test for the correctness of basic ethical beliefs and principles or else implies that there cannot be such a test" (Williams, *Ethics and the Limits of Philosophy*, 72).

[6] Williams, *Ethics and the Limits of Philosophy,* 194.

[7] Williams, *Ethics and the Limits of Philosophy*, 197 (also see 104).

[8] See Williams, *Ethics and the Limits of Philosophy*, 191-96.

For Novakovic's Hegel, the ultimate problem is different. The reason it is dangerous to abstract principles of right from the social and historical reality from which they have arisen is because it is only in this context that such principles could prove their legitimacy or justification in the first place (169 and 171). Although this goes against some still-influential ways of reading the *Philosophy of Right*, it seems exactly right to me as an interpretive matter. Rather than starting with certain abstract principles which are presumed to be self-justified, and then showing that they obligate us to form certain institutions, Hegel is starting with certain principles we might wrongly oppose to ethical life and showing that they require certain specific institutions or practices.[9] This fits Hegel's claim that ethical life is both the *foundation* and *presupposition* of the previous spheres of right—and Novakovic makes a powerful case for it.[10]

The devil, though, is in the details. The more difficult question is this: why does Hegel think that no principles of right could have legitimacy or justification on their own, independently of social and historical reality? Novakovic's way of answering this question is familiar enough to sound distinctively Hegelian but quite original in its specific details and implications. She says: "[F]or Hegel, the legitimacy of any principle of right depends on its actualizability. It is a rational principle, so one worth pursuing only if it can be actualized" (171; also see 141). The basic thought she attributes to Hegel is that no principle can be deemed rational or binding unless it can be actualized, or consistently enacted, and that this is something that can only be known from an empirical analysis of historical reality (177).

Let me register one qualm about this before I move on. There is an important difference between saying that a given principle of right is only legitimate if it can be actualized, and that it is only legitimate if the conditions for its actualization are already more or less present. I take it that Novakovic is saying the former, but not the latter.[11] This suggests she is primarily worried about cases where the principle being invoked is entirely unrealistic, could never possibly be realized in a society. But there are other cases Hegel also seems to be worried about. For example,

[9] Novakovic strongly associates the former method with Fichte, but the same constructivist method is also clearly at work in Kant's *Rechtslehre*.

[10] Hegel, *Philosophy of Mind,* §408 Z.

[11] I am hesitant to attribute this position to Novakovic because she also thinks we can only *know* that a principle can be actualized *after* it has become actualized (Novakovic, *Second Nature*, 179). This is how she captures Hegel's claims about the retrospectivity of philosophical insight (the owl of Minerva flying only at dusk, etc.).

he says it was a mistake for Napoleon to impose a rational constitution on Catholic Spain.[12] The problem here is not that the rational principles codified in the Bayonne Constitution could not be actualized *period*—they had already proved to actualizable in France—but just that the conditions for their actualization were not present *in Spain*. This suggests that the actualization principle Hegel is invoking might be stricter than the principle Novakovic relies on, which entails only that a principle needs to be *actualizable* to be rational, not that certain conditions need to be *actualized* for a principle to be justified.

We will return to the question of whether this view of the actualization principle is strict enough in a moment, but let me now summarize Novakovic's interpretation of Hegel's negative proposal—Hegel's account of the insufficiency of traditional moral theory. On her reading, Hegel's problem with moral theory is that it cannot determine by itself whether the principles or criteria it appeals to are legitimate, and it cannot do this because their legitimacy depends on their actualizability, whether they can be consistently enacted in an actual social community. And moral theory cannot address this latter question, she thinks, because the question of whether a principle can be actualized is clearly an empirical one: it "awaits the unfolding of historical experience" (171).

The positive proposal

It is easy to see, given this account of the limitations of moral theory, the kind of positive proposal that would be needed to fix the problem. We need to replace abstract moral theory with a more empirical and reconstructive approach, one that can determine whether a given principle of right can actually be successfully realized. This, according to Novakovic, is what we get in the *Philosophy of Right*. One of the great achievements of her book is that it offers a highly original reading of the *Philosophy of Right* which shows how it can be read as carrying out something like this project. Her reading of Hegel is akin to Honneth's recent attempt to 'reactualize' the *Philosophy of Right,* but she shows greater care in squaring her reading with Hegel's own methodological commitments, commitments that Honneth himself explicitly brackets or ignores.[13]

[12] Hegel, *Philosophy of Right,* §274 A; Hegel makes similar comments about the limitations of an *a priori* code of law in *Philosophy of Mind,* §552.

[13] This would make Novakovic's account an example of what Honneth calls the "direct strategy" for re-appropriating the *Philosophy of Right* (see Honneth, *Pathologies*, 4-5).

In order to facilitate comprehension of Novakovic's reading of the *Philosophy of Right*, and set the stage for my questions, let me provide a little bit of technical terminology. As anyone who has read the *Philosophy of Right* is aware, the book is structured in the following way. In the Introduction, Hegel provides a partial deduction the *concept* (*Begriff*) of right. He then proceeds from the abstract concept of right to a series of increasingly more concrete determinations in the development of this concept which he terms *shapes* (*Gestalten*) of right, like "the right of property, contract, morality, etc." (§32 R). These shapes culminate or result in a final shape, ethical life, which is characterized as the *actuality* (*Wirklichkeit*) of the concept of right. Any adequate interpretation of the *Philosophy of Right* will need to provide answers to three questions: where does the *concept* of right come from? How do we get from the concept to the various *shapes* of the concept like the right of property? How do we determine that one of these shapes counts as the genuine *actualization* (or truth) of the concept? Novakovic's book offers original and provocative answers to all three questions.

The first question concerns the basic concept of right—where does it come from and what is its status in the ensuing argument? Novakovic argues that because Hegel thinks any principle is only legitimate insofar as it has proven to be actualizable, he must regard the principle he starts with as provisional, as something to be justified only by the ensuing argument that it is actualizable. But she recognizes that Hegel tends to suggest that the starting principle of any science, the concept of the science, is only provisional when taken in independence from the other philosophical sciences—properly considered, the fundamental concept of any science has its proof in the sciences that precede it (185-6).[14] This appears contradictory, but she thinks these claims can be reconciled by viewing the system as a circle of circles (187). Each science is tasked with vindicating its own starting point, though that starting point is provided by another science.

It seems to me, though, that the appearance of a contradiction here is driven by an equivocation about what it means for a concept to be justified or legitimate. Novakovic's position seems to be that *if* the starting concept of the *Philosophy of Right* were considered as justified in the sense of being fully established prior to the argument that ensues from it, *then* Hegel would have no choice but to regard the concept of right as binding on individuals independently of actuality—thus lapsing back into moral theory in the problematic sense. But why couldn't Hegel

[14] See, e.g., Hegel, *Hegel's Aesthetics,* 24-5.

simultaneously think that the concept of right is conceptually necessary given the sciences that precede it, and that right requires certain social conditions in order to exist as valid and binding on individuals? Is there any necessary incompatibility between claiming that the concept of right can be legitimately binding on individuals only under certain socials conditions and the claim that the concept of right is the only legitimate starting point for a science of right? They both seem like perfectly Hegelian claims to me.

The second question concerns how one progresses through the various shapes of right. On Novakovic's interpretation, it really doesn't matter where you begin or in what order you proceed. In principle, you could begin anywhere, with whatever principle of right you would like to examine, any commitment we currently avow (184, 188, 196, 200). But she notes that in practice, Hegel's accounts have a more determinate structure: they move from more abstract shapes to more concrete. For example, the *Philosophy of Right* starts with the most abstract standards of right and ends with the most concrete manifestation of right, ethical life. She suggests that what motivates Hegel to proceed in this fashion is his stated desire to liberate the half-philosophical from "the shackles of some abstraction." By starting with one abstract principle and then showing that this principle requires the precise institutions which already exist in order to be valid, Hegel is able to remind the half-philosophical of something they already know, but which lies below reflective awareness (190). The *Philosophy of Right* does this by showing that the more abstract standards lack *determinacy* and thus are inapplicable to reality unless they are supplemented by other practices and principles (197). In other words, he is reminding us that the standards we use to criticize ethical life only function as applicable criteria of evaluation by virtue of this taken-for-granted context (199).

I think this is a perceptive, powerful, and convincing way of understanding the basic task that the *Philosophy of Right* sets for itself; it is an improvement on the currently prevalent view that the *Philosophy of Right* is itself supposed to reconcile the ordinary or non-philosophical individual to the modern social order.[15] As Novakovic rightly points out, if a society is truly rational, its citizens should already be reconciled to it; they should not need philosophy to accomplish that task (202). But it seems to me that the very cogency of Novakovic's reconstruction of the argument of *Philosophy of Right* militates against her claim that, in principle,

[15] See Hardimon, *Hegel's Social Philosophy*.

it does not matter where you start or in what order you proceed, for it suggests a quite intricate ordering of principles from more abstract to more concrete. Indeed, if *all* of Hegel's sciences in fact proceed in this manner, which she herself seems to concede at one point, then it seems quite unlikely that Hegel does not have a more general reason for proceeding from abstract to concrete, one that goes beyond the specific rationale Novakovic provides, which only fits the *Philosophy of Right*. This suggests that the order in which Hegel proceeds might itself be a necessary feature of the scientific method as Hegel understands it (and, indeed, there are places where he seems to say just that).[16]

Again, I think Novakovic resists this interpretation of Hegel's method because she is worried that if we find *too much* deductive necessity in the progression of shapes of right, we will lack any way of distinguishing his method from that of Fichte and the other moral theorists. But an argument, say, that the right of property necessarily requires the development of property law and a judicial system in order to be fully actual (*wirklich*) does not imply that the property rights are valid independently of social circumstances. It seems to me that Hegel could argue that the progression from the abstractions of abstract right to concreteness of ethical life is necessary, even *a priori* necessary, without lapsing into moral theory in the problematic sense, which involves viewing principles like the right of property as authoritative independently of the conditions of social reality.[17]

The third question concerns how we know that the concept of right, or any particular principle of right, has been fully actualized. For Novakovic's Hegel, we can only know that a principle has been adequately actualized through historical experience which shows that it can be consistently enacted. But she is explicit that even a positive verdict here is not definitive, for experience is never sufficient to prove that new problems or contradictions with the implementation of a principle will not emerge in the future. This gives her Hegel an appealingly modest profile. On her interpretation, modern ethical life is not the complete and final realization of a certain number of principles of right that are themselves necessary forms taken by the concept of right. Instead, the modern state is a provisionally adequate realization of certain principles of

[16] See, e.g., Hegel, *Lectures on the History of Philosophy*, 54.

[17] To be clear, the question of whether any given society has the correct institutions to support a given right *is* irreducibly empirical, but that does not mean there is no way of logically deducing the dependence of a given right on certain institutions.

right that we provisionally avow. It may turn out that we need new principles in the future, or that reality presents new obstacles to the realization of our existing principles—both these options are explicitly kept open.

There are certainly some passages in Hegel which lend support to this interpretation, and Novakovic exploits these to full effect. But rather than enter into a full debate about whether this reading does justice to Hegel's stated ambitions for the *Philosophy of Right*, let me just raise a single philosophical worry about this position: which is that although the requirement that a principle of right can be actualized might be a *necessary* condition for its legitimacy, it does not appear to be a *sufficient* condition. Hegel himself seems to admit the need for more than actualizability when he says that a determination of right "may be shown to be entirely *grounded in* and *consistent with* the prevailing *circumstances* and *existing* legal institutions, yet it may be contrary to right [*unrechtlich*] and irrational [*unvernünftig*] in and for itself."[18] This suggests that the actualization test, at least as Novakovic interprets it, is too weak to offer a full justification of our principles. For even if we can say that the principles we happen to avow are adequately realized in our own circumstances and institutions, this gives us no reason at all to think that those principles are themselves genuinely rational.

Conclusion

There are many interesting aspects of *Hegel on Second Nature in Ethical Life* that I have not touched on, and that are worthy of further discussion. I have focused on Novakovic's suggestion that Hegel's practical philosophy has been misread as a moral theory—an attempt to identify the ultimate moral principles that determine right and wrong action—when in reality it is an attempt to replace moral theory with something entirely novel: a kind of normative social theory with empirical content. Although it is hard to completely ignore Hegel's explicit denial that individuals need philosophy to tell them what their duties are, many interpretations attribute to Hegel what is ultimately only a modified version of this task: one that supplements an account of what we ought to do with an account of the institutions we ought to bring into existence.[19] Indeed, it has been hard to see what the alternative to moral theory of this sort could even be, if not either

[18] Hegel, *Philosophy of Right*, §3R.

[19] Most recently, see Westphal, "Hegel, Natural Law & Moral Constructivism."

an extreme and unpalatable form of ethical conventionalism or an Aristotelian ethical particularism that is hard to square with Hegel's ambition to provide a science of right.[20]

Novakovic, however, has offered a clear and persuasive alternative, arguing that Hegel takes the task of a philosophy of right to be primarily an attempt to remind half-philosophical individuals—the atheists of the ethical world whom reflection has alienated from their own better wisdom—that the very principles they appeal to in order to criticize modern life are only justified and determinate by virtue of precisely those ethical practices and institutions that they are being turned against. This not only fits the text better--making sense of passages that are often soft-pedalled—it also provides us with a clear idea of what could replace moral theory as it is traditionally understood. In discharging my duties as a critic, I have identified places where I think her re-interpretation goes beyond what Hegel's texts strictly warrant, jettisoning too much of the conceptual necessity Hegel seeks to find in the domain of right. But this dispute takes for granted that the fundamental claims she is making here about the radically unconventional nature of Hegel's practical philosophy are both important and correct. I hope that the book will find a wide readership.

Bibliography

Geuss, Raymond. *Outside Ethics.* Princeton: Princeton University Press, 2009.

Hardimon, Michael. *Hegel's Social Philosophy: The Project of Reconciliation.* Cambridge: Cambridge University Press, 1994.

Hegel, G. W. F. *Elements of the Philosophy of Right.* Translated by H. B. Nisbet. Cambridge: Cambridge University Press, 1991.

Hegel, G. W. F. *Encyclopedia of the Philosophical Sciences in Basic Outline, Part I: Science of Logic.* Translated by K. Brinkmann and D. Dahlstrom. Cambridge: Cambridge University Press, 2010.

Hegel, G. W. F. *Hegel's Aesthetics: Lectures on Fine Art: Volume I.* Translated by T. M. Knox. Cambridge: Cambridge University Press, 1975.

[20] A third possibility which Novakovic does not neglect to engage with is Dean Moyar's suggestion, in *Hegel's Conscience*, that Hegel's ethical teaching bottoms out in an appeal to "true conscience" (see Novakovic, *Second Nature*, 46-7).

Hegel, G. W. F. *Lectures on the History of Philosophy 1825-6*, *Volume I: Introduction and Oriental Philosophy.* Translated by R. F. Brown and M. Stewart. Oxford: Oxford University Press, 2009.

Hegel, G. W. F. *Lectures on the Philosophy of World History: Introduction.* Translated by H. B. Nisbet. Cambridge: Cambridge University Press, 1975,

Hegel, G. W. F. *Philosophy of Mind.* Translated by A. V. Miller. Oxford: Oxford University Press, 1971.

Honneth, A. *The Pathologies of Individual Freedom.* Translated by L. Löb. Princeton: Princeton University Press, 2016.

Moyar, Dean. *Hegel's Conscience*. Oxford: Oxford University Press, 2011.

Westphal, Kenneth R. "Hegel, Natural Law & Moral Constructivism." In *The Owl of Minerva: Journal of the Hegel Society of America*, Vol. 48, Numbers 1-2 (2016-17): 1-44.

Williams, Bernard. *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press, 1985.

Hegel on Second Nature in Ethical Life: Response to Dean Moyar and Mark Alznauer

Andreja Novakovic

Introduction

Many thanks to Mark Alznauer and Dean Moyar for their generous, insightful, and challenging comments. I am grateful for and amenable to many of their suggestions. Instead of going through each of their questions one by one, I will try to combine a few and address them together. First, I will begin with a brief introduction to the book and an explanation of how its subject matter unfolded. Second, I will address those questions having to do with ethical rules and moral theories. Third, I will address those questions having to do with Hegel's method, specifically the bridge between the *Logic* and the *Philosophy of Right*. And fourth, I will speak to the role of immanent critique, religion, and conscience.

The book began as an effort to understand Hegel's reasons for privileging unreflective forms of social participation, specifically habit and custom, in his account of ethical life. This seemed at first glance to be a relatively minor topic when compared to the more well-known features of Hegel's *Philosophy of Right*, such as his account of institutional spheres like civil society, and of freedom more generally. But what I discovered in the process is that habit and custom bear on many other aspects of ethical life, including civil society and freedom. In fact, getting clear about their place in ethical life brought me to rethink Hegel's entire project in the *Philosophy of Right*.

His remarks about habit and custom – such as §151 of the *Philosophy of Right,* which was my starting point – usually fell under readers' radar, at least at the time when I began thinking about the topic, though interest in it has significantly increased in recent years. Those who did notice them usually noted Hegel's Aristotelian heritage, explaining Hegel's reference to habit as his agreement with Aristotle that well-adjusted social participants are going to take pleasure in the right sorts of things, which habituation is supposed to accomplish. This role of habit seemed innocuous enough and did not raise many worries, for it did not seem difficult to accept that it is probably better that people also enjoy performing the duties associated with their

roles. In any case, it did not seem to require accepting the stronger claim – which I think remarks such as §151 do suggest – that it is indeed better that people perform those very roles habitually, as a matter of habit.

There was, however, a growing sense that Hegel's references to habit point to an ambivalence at the core of ethical life, captured in the phrase "second nature." We acquire a second nature by engaging in certain sorts of activities, and what we acquire when we acquire a second nature is itself a product of human effort. This makes habit a "second" nature, the work of freedom. But a second nature resembles "nature" to the extent to which habit becomes something fixed, static, repeated without deliberation, and hence something detached from one's intellect or will. This seeming ambiguity between freedom and unfreedom expressed in the phrase led to the relatively standard verdict that habit needs to be surpassed, supplemented, or perpetually criticized, in order to avoid the danger of a mindless perpetuation. This is the feeling of unease that my book tries to dispel. I argue that Hegel is not ambivalent about the role of habit (and custom) in ethical life, as long as these are understood in the right context.

One of my main claims is that the habit of the ethical, which is the habitual performance of one's social roles, is actually a perfectly adequate expression of freedom, specifically subjective freedom, which here means the freedom to do what I see as good, to act in accordance with my convictions. I suggest that there is even an intimate connection between habit and subjective freedom, since I can only be said to be genuinely convinced that a certain principle is good, worth upholding, if I have the corresponding habit of heeding it. This is not to deny that we can also develop bad habits, that some habits are simply ethically neutral, and that habit can also devolve into "mere habit". What I wanted to challenge was the assumption that there is something suspect about habit in virtue of its form. I show that habit is not only an adequate, but it is in fact the best expression of subjective freedom so understood.

That said, I also acknowledged that Hegel's conception of social participation is not fully captured by this account of habit, since there are also more overtly reflective ways of relating to ethical life that are appropriate under certain circumstances. So my interest in the unreflective aspects of ethical life led me to reconstruct many other aspects of his *Philosophy of Right* in its light – the role of culture and cultural participation, the exercise of critical reflection about practices and their underlying norms, and finally philosophical comprehension itself. I chose to include the phrase "second nature" in my title because I saw it as encompassing this range of

topics, foregrounding, for example, that ethical life is a product of work, that it must be inhabitable to those tasked with perpetuating it, and that our participation in it (when it is functioning properly) is for the most part spontaneous, effortless, and unproblematic.

But this is clearly not always the case. So another part of my project was to bring into view situations that call for a critical stance and to consider the broader question of how open Hegel's account is to social criticism. What I suggested is that assessing the place of critical reflection in ethical life involved assessing the prospect that such situations will continue to arise, situations marked by *contradiction* that undermine unreflective forms of social participation. And this required going beyond Hegel's official text. Even if the *Philosophy of Right* is not read as debilitating or discouraging critical reflection, it does not discuss very many contradictory occasions that have the power to call for more radical forms of criticism.

It was also important to me to show that Hegel's own philosophical reconstruction of ethical life makes a very poor model for social criticism, since it is not meant to be critical – at least not of ethical life. In this regard I depart from readings of Hegel's project in the *Philosophy of Right*, which take it to be an early paradigm for a "critical theory" of society. But if it is not a critical theory of society, I had to offer an alternative interpretation of Hegel's method. What I argue, in short, is that Hegel's philosophical reconstruction of ethical life is, surprisingly, both normative and quietistic, because its aim is to show that we can and should trust our ordinary forms of proceeding within ethical life. This required laying out a fuller picture of pre-philosophical social participation, and hence of what it is involved in inhabiting ethical life as a second nature, on Hegel's behalf.

Theories and Rules

In Alznauer's helpful reconstruction of my argument, he points out that Hegel (on my reading) rejects moral theorizing, and that he does so not just because it is idle and needless, but because there is something misguided about it. So his first question is *why* Hegel believes this to be the case. One option would be to say that Hegel, similarly to McDowell, holds that ethical knowledge cannot be codified and that it is this that makes moral theorizing misguided, since the aim of moral theorizing is to identify some set of rules to guide judgment and action. But as Alznauer points out, this option is not available to me, since I argue that ethical knowledge is for Hegel codifiable. I take as my evidence Hegel's position in the domain of positive law,

according to which custom and habit can be translated into a law-like form without losing their rich content, although doing so need not come at the cost of following the law habitually or customarily.

Both Moyar and Alznauer find the conclusions I draw from this to be puzzling. As Moyar asks: is this not to ascribe to Hegel an "ethical legalism" that is unsupported by the text, which does not offer any codification of such rules, laws, or principles that are meant to guide judgment and action? And as Alznauer asks: why do I also argue that rules can only be genuinely known by someone who follows them habitually, if I am at the same time insisting that this knowledge is codifiable? To use Alznauer's example: "If our paradigm for ethical knowledge is a general principle like 'don't break contracts,' then are we supposed to think that as someone loses the habit of conforming with his contractual obligations, he is somehow losing *knowledge* of that general principle?" So let me take this opportunity to clarify this relationship between rules, habits, and knowledge.

To address Moyar's question first, I am trying to make sense of Hegel's idea that knowledge of what to do is a function of looking to the requirements of roles. As Hegel emphasizes, it is not difficult to find out what I ought to do in a rational social order because this is prescribed by the relevant role. And I suggest that roles come with a *code*, though it is often an informal one, and so significantly different from a legal code whose breach warrants punishment. Although Hegel is acutely aware that enacting this code requires the exercise of judgment and thus a sensitivity to the specific circumstances, he is extremely reluctant to liberate this judgment from all rule-like constraints, because this would grant too much to individual discretion. In other words, Hegel rejects the ideal of moral expertise, according to which acquiring ethical knowledge involves becoming an expert judge in ethical matters. He thinks that judgment is important, but constrained by something universal and universally intelligible: a more or less explicit, more or less formal code that can be articulated as such, as a set of rules or principles.[1]

The reason Hegel does not include such a code in his text and does not provide a list of rules or principles is because he thinks that it is, just like positive law, simply too contingent on

[1] I should note that I am not saying that we always conceive of this code as a code when we are following it, only that it *could* be so conceived, if need be. I think my view is compatible with Moyar's claim that the universality of principle is only a moment in deliberation. In fact I emphasize that when we have made a habit of following this code, it plays no explicit role in deliberation at all.

changing circumstances. Two examples he gives in the Preface are Fichte's argument that a passport ought to include a picture of the person whose passport it is, and Plato's argument that nurses ought to rock babies back and forth. While both of these might be perfectly good principles under specific conditions, and hence suitable codes of conduct for the roles of citizen and nurse, this is not something for philosophy to adjudicate. So Hegel can hold both that ethical knowledge is codifiable *and* that this code must be so context-sensitive that it makes little sense to try to incorporate it into a philosophy of right, which is supposed to be able to distinguish between that which is essential to ethical life and that which is ineradicably changeable in it.

To turn to Alznauer's question, if I place so much significance on rules and principles, why do I also say that ethical knowledge is achieved only by those who have a habit of following them? What is the relevance sense of knowledge here? One familiar alternative would have been to say that knowing rules involves knowing how to apply them and that this "know-how" is something that only the habit of applying them can yield. But while this is probably true, it was not the sense of knowledge I had in mind. Rather, I was interested in what it takes to know that the rule is good and I suggested that one thing that it takes to know that a rule is good is to be convinced that it is good. So it is this conviction, often tacit or implicit, that I think only the habit of following the rule can definitively demonstrate, since it is this habit that proves that I am genuinely committed to its adherence. Being convinced that a rule is worth following goes beyond understanding what would count as following it, which even someone who is not convinced of its goodness can do. So, in what sense does someone who begins to break contracts no longer *know* the rule "do not break contracts"? Well, in the sense that this person's conviction has faltered.

Let me return to the opening question – what are Hegel's reasons against moral theorizing? – and consider the second option available to Hegel, which Alznauer identifies as Bernard Williams'. As Alznauer notes, I am sympathetic to this comparison between Hegel and Williams, though my sympathy lies elsewhere than Alznauer suggests. What Alznauer stresses is that Williams thinks our practices of blame cannot withstand reflective scrutiny because they are based on Nietzschean fictions that work only as long as they remain unexposed. But the aspect of Williams that I find more relevant for my purposes is that moral theorizing impoverishes the evaluative resources available to social participants because it leads them to seek one master-principle from which all others are to be derived, or through which all others are

to be justified. This is something that Hegel, like Williams, opposes, for Hegel thinks that different domains of social life are going to operate according to significantly different principles.

So one reason against moral theorizing is that ethical life is always richer, more varied and internally differentiated, than a moral theory can hope to capture. As we have seen, it includes many highly context-dependent codes of behavior, which can be better or worse, given their context, but which cannot be evaluated in a context-independent way. But as Alznauer rightly notes, the even deeper problem is that every principle at work in ethical life is in a certain respect context-dependent, because no principle relevant to it is self-legitimating in abstraction from all empirical considerations. Even such principles as "obey contracts" and "respect private property" – which Hegel thinks are, unlike "carry a passport photo" and "rock babies back and forth," essential to ethical life – can only be evaluated in light of the practices to which they give rise.[2] This means that the main reason against moral theorizing, at least in the way it has been traditionally conducted, lies in Hegel's method. To cite Alznauer, "the basic thought [I attribute] to Hegel is that no principle can be deemed rational or binding unless it can be actualized, or consistently enacted, and that this is something that can be only be known from an empirical analysis of historical reality."

Logic and Right

Next, I want to turn to those questions concerning my interpretation of Hegel's philosophical method. One big issue haunting Hegel's *Philosophy of Right* is its relation to the *Science of Logic*. Although some interpreters have treated the *Philosophy of Right* as a self-standing treatise in moral, social, and political philosophy, others have made a convincing case that its central concepts presuppose the analysis Hegel has given of them in the *Logic*. A familiar example is Hegel's concept of "actuality". Readers used to wrongly assume that Hegel's infamous *Doppelsatz* – that "what is rational is actual; and what is actual is rational" – is meant to be vindication of the status quo, whereas Hegel's account of "actuality" in the *Logic* shows that the term refers to something much narrower than whatever happens to be in place. Actuality is that which has actualized its concept, so which has met the standard appropriate to it.

[2] Alznauer is right that "actualizability" could be interpreted as too broad to serve a test, since a principle might be actualizable in France, but not in Spain, making it in an inappropriate principle to which to hold the Spanish. In the book I discourage seeing "actualizability" as anything akin to a test that can be implemented any place, any time.

This means that there will be much in social life that will fail to count as actual and so will fall outside of Hegel's philosophical reconstruction. Hegel's own example is the corrupt state, which would not be an actual state, although it exists in abundance. It also means that, while the *Doppelsatz* is true by definition, it has a methodological purpose because it restricts the subject matter of a philosophy of right to that which has proven actual, hence rational. This is the version of the methodological reading that I outline and defend.

My reading of Hegel's method is intended to explain how Hegel can have it both ways, how he can connect the *Philosophy of Right* to the *Logic* and still consider it self-standing. I try to do this by taking seriously his metaphor of a "circle of circles," which suggests a bigger system that consists of "sciences" that are relatively self-enclosed, though they glean their starting point from each other. I accept Alznauer's point that it is perfectly compatible to say that the concept of right, for example, is the result of another science which legitimates it as the starting point, but that this concept is only binding on individuals under certain social conditions, so that it requires ethical life for its actualization. In short, a philosophy of right could have inherited the concept of right from another part of Hegel's system, and yet still be faced with its own distinct task of showing that this concept has a legitimate place in ethical life.

But I also argue that, given the metaphor of a circle, Hegel thinks that it does not matter where we start, which concept or principle of ethical life we take as our point of departure, because we will always arrive at the same conclusion – that this concept or principle presupposes ethical life and is only actual in its context. Alznauer asks me whether this is not to neglect a *necessary* feature of Hegel's method, which always proceeds in roughly the same order, from abstract to concrete, irrespective of domain. As he puts it, "if *all* of Hegel's sciences proceed in this manner, which she herself seems to concede, then it seems quite unlikely that Hegel does not have a more general reason for proceeding from abstract to concrete." Here I want to make two brief suggestions for how to explain the order in which Hegel consistently proceeds. First, the cultural diagnosis that I offer to explain the starting point in the *Philosophy of Right* could easily be extended to the other sciences, since Hegel thinks that abstraction is a mark of his age and that this makes the most abstract concept or principle also the one that is most readily available to reflective thought. Second, I think Hegel has an enduring concern for completion. He wants to demonstrate that every principle in ethical life requires ethical life for its actuality, and he thinks

that the only way to do this is to begin with that which appears basic, simple, and minimal – the most abstract – and show that *even it* presupposes ethical life.[3]

Despite Hegel's concern for completion, his method is, according to my reading, very modest in its ambitions, maybe too modest for some people's philosophical tastes. As Alznauer points out, although the fact that a concept can be actualized under given circumstances is necessary for demonstrating its rationality, it is not sufficient for doing so. It might be actualizable given current conditions, but prove to be defective at a later stage of its development. This suggests that it is not possible for a philosophical account oriented toward that which is actual/rational to definitively demonstrate that its object is indeed actual/rational. To repeat, a principle might be actualizable *provisionally*, under present circumstances, without this thereby implying that it is "actual" in the sense of an adequate realization of its concept. Some things work for a while and reveal their deeper flaws only down the road.

Let me now turn to Moyar's objection to my gloss on the *Doppelsatz*. He thinks that, while it is relatively uncontroversial to say that "actuality" is already a normative standard, I go too far in claiming that the *Doppelsatz* is trivially true because actuality and rationality are identical by definition. As he puts it, "Hegelian identity typically implies difference as well, so even the simplest seeming identity claims will turn out not be simple at all." According to Moyar, even though to identify something as actual is to identify it as rational, actuality contains more than rationality, because it also contains what Hegel calls externality, "the contingent and messy appearances." This suggests that it makes good sense to measure actuality by the standard of rationality, to consider some instances of actuality more rational than others, and hence to presuppose that there is a difference between the two. He also notes that this would make the version from the lectures, that "the rational becomes actual and the rational becomes actual," coherent.

Moyar invokes Hegel's reference to Plato to support this distinction, arguing that Hegel could not be crediting Plato with the appreciation of something that is trivially true. But I read the Plato reference differently. Hegel is not crediting Plato with the appreciation of the

[3] Here I appeal to Horstmann's explanation for Hegel's procedure in the *Phenomenology of Spirit*. He argues that Hegel wants to demonstrate the "primacy of the maximally complex over the elementary simple" by revealing that it is presupposed by even in the *most* simple, hence the most abstract. See Rolf-Peter Horstmann, "The *Phenomenology of Spirit* as a 'Transcendentalistic' Argument for a Monistic Ontology," in *Hegel's Phenomenology of Spirit: A Critical Guide,* ed. Dean Moyar and Michael Quante (Cambridge: Cambridge University Press, 2011), 42 – 62.

Doppelsatz per se, but with its implementation in his political philosophy, specifically in the *Republic*. In other words, it is not that Plato discovered that the actual is rational and vice versa, but that he discovered that his own ethical world is rational/actual, hence orienting himself toward that which is "immanent" and "eternal" in the present context. For Hegel this amounts in Plato's case to an acknowledgement that a newly emerged principle ("free infinite personality") needs to be given its fair due, which Plato attempts to do by granting it to the "philosopher-king," even if he does not want to make it available to other citizens. So Hegel's point is that Plato is not constructing an ideal state from his own individual cognitive resources, but employing principles that are already at work, even if not yet fully, in his social context.

I find Moyar's way of differentiating actuality and rationality helpful because it emphasizes that, in order for something to count as actual, it will have to incorporate external elements as well. This I fully accept. But it is also important to avoid confusing "actuality" in the strict sense with the "ethical world" in all its contingent richness. And while the ethical world can be said to undergo a process of becoming rational, I maintain that it does not make sense to say that *actuality* is undergoing such a process.

What I am trying to resist is seeing Hegel's method in the *Philosophy of Right* as one application. I do not want to say that the *Logic* tells us what the idea of the good consists in, and hence what our standard of rationality ought to be, which we can subsequently use in the evaluation of ethical actuality, considering some actualities more rational than others. Rather, it is a task of *finding* the appropriate standard in your object, in the ethical world that already exists. This is consistent with Hegel's insistence that we should give up thinking of the concept as a measure for something that is external to it, but rather recognize the concept as the animating principle that is objectively manifest. In other words, Hegel's method in the *Philosophy of Right* involves showing that the ethical world is not merely external, but has an actual core, even if it retains purely external aspects. And this will involve seeking a more determinate measure of the good, appropriate to the social world, than the one that the *Logic* itself can provide.

Moyar asks me how exactly I understand the relationship between the *Logic* and the *Philosophy of Right*, specifically, whether I want to say that the *Logic*, too, is ultimately grounded in history and hence the product of a developmental process, much like ethical life itself, or whether it is just that the account of the Idea in the *Logic* is too indeterminate to be of

much use to the *Philosophy of Right*. While I am not prepared to make the former claim, I am prepared to say that Hegel's method cannot consist in determining, in an a priori way, the purely "logical" idea of the good and then moving on to *apply* this standard to actual ethical life, which lies outside it. This way of proceeding would in fact betray a misunderstanding of the "idea" as it is laid out in the *Logic.*

Critique, Religion, and Conscience

Last, I want to turn to another set of questions, beginning with Moyar's challenge to my distinction between internal and immanent criticism, which I owe to Rahel Jaeggi.[4] In the book I argue that there is a difference between criticizing a given practice for failing to measure up to the standard espoused by its participants (this would be internal criticism) and criticizing a practice *and* its standard for leading to unavoidable contradictions (this would be immanent criticism). While internal criticism is indeed a common and familiar form of social criticism, the worry is that it is not radical enough, because it keeps us confined to the standards that are already widely shared. So immanent criticism is supposed to explain how social participants are able to engage in more radical forms of criticism that put those very standards, and not just their enactment, into question. I employ this distinction in order to suggest that there is no reason to banish immanent criticism from ethical life, even if Hegel does not discuss it in this context. My two main examples of contradictions which would warrant its exercise are Hegel's own conception of poverty as a necessary consequence of civil society, and racial segregation of schools in the United States. In short, immanent critique is supposed to be a social process of normative transformation that responds to such contradictions, even when those engaging in it do not conceive of the relevant problems as contradictions.

Moyar asks me whether immanent critique is an effective or desirable model, if it implies that such fundamental norms like the right of personhood and the freedom of moral subjectivity – which "have a bedrock status in the modern world" – might be put into question. Although I agree completely that there are some norms that Hegel does not think could ever come up for review, once they have historically developed, everything hinges on how we identify the relevant norm, or better yet, whether we are willing to distinguish the norm in abstraction from its

[4] Rahel Jaeggi, *Critique of Forms of Life*, trans. Ciaran Cronin (Cambridge, MA: Harvard University Press, 2018), 190 – 208.

concrete interpretations. Take for example the norm of equality. While it is true that a model which tells us that equality as such could become legitimately abandoned would not be very attractive, it would be reasonable to suspect that we have not yet arrived at an incontestable interpretation of what equality in the social world concretely requires. So maybe the picture is more complicated than I suggested. There is not just the norm and its practical enactment, but the norm, its interpretation, and its practical enactment. I think that this set of distinctions can illuminate the way that Hegel understands immanent critique in the *Phenomenology*, in which freedom in some form or another remains a constant aspiration from one configuration to the next, though its interpretation becomes dramatically revised in the process.

Moyar also questions whether desegregation is indeed a good example of immanent criticism and I am grateful to him for giving me this opportunity to reconsider my use of it. What interested me about this example was the conceptual claim in the *Brown v. Board of Education* ruling, that "separate is *inherently* unequal," which was made on the basis of the experience of segregation, including empirical evidence from psychological studies. I invoke it to illustrate how experience can challenge the compatibility of clusters of commitments. But as Moyar points out, there are reasons to think that desegregation might be better understood as an instance of internal criticism, since it involves demanding that the ideal of equality be more consistently enacted (and charging the defenders of "separate but equal" with a failure to recognize the value of equality).[5] This suggests that maybe immanent and internal criticism are only neatly distinguishable as models. Once we are dealing with specific cases, it becomes harder to tell which of the two is at work.

I want to address Moyar's welcome point that a reason against thinking that desegregation is a good example of immanent criticism is that the institution the ruling challenged was never extinguished. I define immanent critique as a social process. Through *Brown v. Board of Education* segregation may have lost its legal basis, but an actual process of desegregation never took place, at least not on a national scale. Many school districts, especially in the North, remain segregated to this day, despite the pressure exerted by activists of the Civil Rights movement. One way this has been captured is by saying that *Brown v. Board of*

[5] I should note that immanent criticism does not presuppose that those whose practices are being criticized avow all of the relevant commitments. It is enough that *Brown v. Board of Education* exposed the contradiction at the heart of a society that is committed to equality and to segregation, even if those people who supported segregation were not "good-faith agents who honestly believed that equality was possible with segregation," as Moyar puts it.

Education eliminated "de jure" segregation, not "de facto" segregation. But there have been those who have argued that this very distinction between "de jure" and "de facto" is highly misleading, since those who fought in favor of keeping schools segregated found other legal means for doing so, for example by redrawing school districts along racial lines.[6]

Moyar raises three further questions. First, he asks me why I use only the first part of the chapter on "Bildung" in the *Phenomenology*, when the later parts, specifically the conflict between Enlightenment and faith, would have been useful for my purposes. I decided to focus on the account of the emergence of "Bildung" as a configuration of spirit, prior to its evolution into the Enlightenment, because I wanted to foreground the structure of reflection that "Bildung" introduces and connect it to work and labor. But this structure is visible in many different contexts. So even though a focus on "Bildung" allowed me to thematize it as such, it could have been similarly well illustrated by subsequent developments.

Moyar suggests that there is a particular reason to think that the conflict between Enlightenment and faith ought to have been relevant to my project. The Enlightenment seeks to disabuse faith of its illusions by revealing that this block of wood is just a block of wood, this piece of bread-dough just a piece of bread-dough (PhG §552). In Moyar's words, "Hegel criticizes this Enlightenment posture on the grounds that religion, and in particular religious practice, is not the kind of attitude about which a people can be deceived. The meaning that religion finds in its worship is immune to the attacks based on a theoretical stance, an appeal to natural science or to historical evidence." Hegel's way of putting it is to say that faithful people engage in double-perception: though they are fully aware that the objects they use in their religious rituals are just ordinary objects, they manage to simultaneously see them *as* embodiments of the divine (PhG §572). So telling them that these are just wood or dough is not informing them of something of which they were not already aware.

Moyar goes on to ask me whether Hegel, as I read him, wants to save the truth of religion from excessive reflection, and whether this extends to religious traditions beyond Protestant Christianity, which is supposed to have overcome the contradiction contained in the above-mentioned double-perception. This is a great question, to which I cannot do justice here. What makes it especially tricky is that religion is both at once – a practice in which people for the most

[6] See Jeanne Theoharis, *A More Beautiful and Terrible History: The Uses and Misuses of Civil Rights History* (Boston: Beacon Press, 2018) for a historical treatment of these issues.

part unreflectively participate, and itself a form of reflection, one manifestation of what Hegel calls "Absolute Spirit." I suspect that Hegel would say that religion permits a kind of reflection on ethical life, a reflection akin to that permitted by art. Participating in religious rituals and frequenting a museum are both distinctly *reflective* practices, because they usually require assuming a stance different from that of daily life. But Hegel is of course worried that religion will come to interfere with daily life, specifically with the patriotism manifest in what he describes as a free obedience of the law.

In a lengthy remark to PR §270, Hegel raises a number of concerns about the religious attitude in the context of ethical life. One is that it will dismiss all mundane forms of patriotism as "worldly matters" that pale in comparison to its serious devotion. Another is that it will conflict with the dictates of the state, holding itself to be an independent (and infallible) source of authority in worldly matters. This could lead to religious fanaticism, an obstinate refusal to listen to anyone except the religious leader, or one's supposed "heart." So Hegel is clearly worried about the role of religion in ethical life. But he admits that these are all perversions of the religious attitude, which is in principle compatible with that required for social participation. He even suggests that, because religion can help integrate individuals into the state, the state ought to demand that its citizens belong to some religious community or another, though it cannot dictate which one.

What I find most interesting about his discussion of religion is that Hegel considers the rise in religiosity to be symptomatic of a social problem. He notes that "people recommend and resort to religion above all in times of public distress, disruption, and oppression, and that they are referred to it for consolation in the face of wrong and for hope as a compensation for loss" (PR §270R). The religious attitude is here being identified as a *symptom* of public distress, disruption, and oppression. In this respect it seems to be similar to *conscience*, which Hegel also describes as indicative that "the existing world of freedom has become unfaithful to the better will" (PR §138R). His idea is that explicit appeals to one's own conscience are a bad sign for the time in which they are made.

This brings me to the third of Moyar's questions, with which I want to conclude. Moyar considers my dismissal of conscience as "not a full-fledged perspective" to be a missed opportunity. According to Moyar, "morality and conscience could play a greater role in uniting ordinary practice and philosophical knowledge than [I accord] them, and could in fact be enlisted

to support key elements of [my] reading." What he suggests is that the morality chapter offers a rather abstract conception of the good as the unity of welfare and right, which already gives us an "all-inclusive concept of freedom," but does not yet offer determinate guidance for action. Conscience is supposed to provide this guidance, because it actualizes the good so conceived.

I suspect that my disagreement with Moyar's conception of conscience does not run very deep, that it might even be primarily terminology, as he himself suggests.[7] Moyar claims that conscience actualizes the good, whereas I would say that it is the ethical disposition (which Hegel identifies as *true* conscience) that actualizes the good. And the ethical disposition, as I understand it, consists in the convictions one has about what is good, which one expresses in habitual conduct. So I also want to emphasize the role of individual agency in *actualizing* the good by pursuing what one takes to be good, what one is convinced is good, which is likewise a process of *determining* the good by implementing it in specific circumstances. But I am reluctant to call this point of view that of "conscience," mainly because Hegel suggests that conscience in any other guise is at best the symptom of a problem that it lacks the resources to resolve.

Bibliography

Horstmann, Rolf-Peter. "The *Phenomenology of Spirit* as a 'Transcendentalistic' Argument for a Monistic Ontology." In *Hegel's Phenomenology of Spirit: A Critical Guide,* edited by Dean Moyar and Michael Quante, 42 – 62. Cambridge: Cambridge University Press, 2011.

Jaeggi, Rahel. *Critique of Forms of Life.* Translated by Ciaran Cronin. Cambridge, MA: Harvard University Press, 2018.

Moyar, Dean. *Hegel's Conscience*. Oxford: Oxford University Press, 2011.

[7] There are some respects in which I would depart from his conception. For example, Moyar claims that conscience "cannot be an original source of norms, but it can be a source for transforming, through proceses of negation, the existing norms. It is an activity of liberation rather than the basis for a construction from the ground up of a society's ethical norms." Dean Moyar, *Hegel's Conscience* (Oxford: Oxford University Press, 2011), 30. So he ascribes more critical potential to conscience than I would. But I am open to his suggestion that Hegel's identification of conscience and absolute knowing, which I did not explore in the book, could be fruitful for understanding the relationship of philosophy to the ordinary point of view.

Theoharis, Jeanne. *A More Beautiful and Terrible History: The Uses and Misuses of Civil Rights History*. Boston: Beacon Press, 2018.

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 Colin McLear

In her book *Kant on Reflection and Virtue*, Melissa Merritt presents an extended defense of Kant's "reflective ideal" against the objection that it is "precious, hyper-deliberate and repugnantly moralistic" (2).[1] This defense is in part constituted by the articulation of a theory of cognitive virtue, which Merritt attributes to Kant. Merritt hopes that Kant's statement that we have a "duty to reflect" (e.g. A261/B317) may be mitigated within the context of such a virtue theory.[2] The requirement to reflect is not, as the "caricature" has it, that one must constantly step back and deliberate as to whether something is worth doing or accepting as true. Rather, "the requirement is to be met by putting one's cognitive capacities to use in the right way, or in the right spirit: reflection can be adverbial, and it is not essentially episodic…what drives the development of cognitive and moral character is an essentially outward-directed interest in knowing" (206), which Merritt identifies with Kant's conception of a "healthy human understanding." This interpretation, she argues, allows the Kantian reflective ideal to avoid being overly demanding, and it helps explain how some kinds of cognitive activity that are not deliberate (at least as that term is typically understood) might nevertheless be justified or "cognitively excellent" (205).

The book is clear and, rather unusually for Kant scholarship, a lively read. Much of Merritt's discussion is plausible and compelling. Just as much of Kant's moral theory may be fruitfully construed within a broader virtue-theoretic framework, so too do virtue-theoretic considerations enliven a treatment of his epistemology and theory of cognition. But despite these sympathies, and since I am here in the role of critic, I will try to bring out some aspects of

[1] All parenthetical page references to Merritt's work will be to Merritt, *Kant on Reflection and Virtue*. unless otherwise noted.

[2] All references to Kant are from the *Akademie Ausgabe*, with the first *Critique* cited by the standard A/B edition pagination, and the other works by volume and page. Where available, translations generally follow the Cambridge editions of the Works of Immanuel Kant, general editors Paul Guyer and Allen Wood.

Merritt's discussion that could use clarification or further discussion. I examine three elements of Merritt's position: (i) the view of reflection as divided into two basic senses, which occupies the lion's share of my comments; (ii) the conception of reason as fundamentally cognitive; (iii) the resolution of the demandingness problem.

1. Reflection-C & Reflection-N

Merritt construes Kant's conception of "reflection" (*Überlegung*, *Reflexion*, *reflexio*) as having two distinct meanings or sets of uses. The first is a constitutive sense—"reflection-c"—which is glossed as the (typically) "tacit handle on oneself as the source of a point of view" (18). This is in contrast to the normative sense—reflection-n—which is glossed in various ways. For example, as "caring about how one's point of view is constituted" (18), as the "manner in which one puts one's cognitive capacities to use *in concreto*" (77), as a "normative requirement on sound judgment" (81), and as "making use of one's cognitive capacities in the right way" (206). Merritt employs several distinct but interrelated arguments in the course of advocating for the distinction between the two forms of reflection. First, she argues that this distinction is needed to resolve a putative puzzle regarding Kant's conception of the difference between affect and passion. I did not find this a particulary convincing entryway into the view, so I'll put it to the side here.[3] Second, she argues that it best connects several distinct but central uses Kant makes of the term "reflection". Third, she argues that it makes best sense of the manner in which we are "required" to reflect on judgment. However, in each of these two cases I think the arguments face some difficult challenges.

1.1 Unifying the Texts

When one canvasses the various uses Kant makes of "reflection" and its cognates in even just the critical period, it certainly seems plausible that he means more than one thing by the term. Indeed, one might despair of finding some common thread that unites all of the disparate uses he makes of the notion. Here are five ways that Kant's critical writings use the term "reflection" (*Überlegung*, *Reflexion*, *reflexio*), all but one of which Merritt discusses.

[3] For comments on the problems raised by the account of affect and passion see Russell, this issue.

A. The activity characteristic of the discursive intellect generally:

> the intellect intuits nothing but only reflects (4:288)

B. Activity connected with pure apperception:

> consciousness of oneself (*apperceptio*) can be divided into that of reflection and that of apprehension. The first is a consciousness of understanding, pure apperception; the second a consciousness of inner sense, empirical apperception. (7:135, note)

C. A mental operation by which concepts are generated:

> To reflect (to consider) [*Reflectiren* (Überlegen)], however, is to compare and to hold together given representations either with others or with one's faculty of cognition, in relation to a concept thereby made possible. (20:211)

> [the] inner activity, (spontaneity), by means of which a concept (a thought) becomes possible, [is] reflection [*Reflexion*] (7:135, note)

D. A requirement of judgment:

> we cannot and may not judge about anything without reflecting [*überlegen*] (*Logik Jäsche* 9:76; cf. Amphiboly, A261/B317)

E. A capacity we share with non-rational animals:

> Reflecting [*Das Reflectiren*] (which goes on even in animals, although only instinctively, namely not in relation to a concept which is thereby to be attained but rather in relation to some inclination which is thereby to be determined) (20:211)

If Merritt is correct that Kant's various uses of "reflect" can be sorted into one of two fundamental kinds—viz. reflection-c or reflection-n, then this would at least show the usefulness of reading Kant in this way, whether or not it is what he intended. In Merritt's view (A)-(C) belong under reflection-c while (D) is indicative of reflection-n. She does not address the issue of (E), reflection in non-rational beings, and it is not clear how it can be incorporated into her view.

Below I first raise two sets of considerations that count against Merritt's unification of (A)-(C) via her account of reflection-c. I then go on to discuss some concerns with the account of reflection-n.

1.1.1 Comparison & Reflection

Anglophone interpreters have by and large been attracted to a reading of Kant's various uses of "reflection" as united by the activity of what Kant calls "comparison" (*Vergleichung*).[4] Kant appeals to reflection in the comparison of representations to generate cognition at least as early as the *Inaugural Dissertation* (e.g. 2:394), but he makes clearest use of the notion of comparison in two key passages from the Amphiboly section of the first *Critique*.

> The action through which I make the comparison of representations in general with the cognitive power in which they are situated, and through which I distinguish whether they are to be compared to one another as belonging to the pure understanding or to pure intuition, I call *transcendental reflection*. (A261/B317)

> whether the things are identical or different, in agreement or in opposition, etc., cannot immediately be made out from the concepts themselves through mere comparison (*comparatio*), but rather only through the distinction of the kind of cognition to which they belong, by means of a transcendental reflection (*reflexio*). To be sure, one could therefore say that *logical reflection* is a mere comparison…(A262/B318)

[4] Longuenesse *Kant and the Capacity to Judge*, ch. 5. takes this tack in her work, though as Merritt notes (Merritt, *Kant on Reflection and Virtue*, 35 note 37), it is unclear whether Longuenesse intends to support her reading primarily if not solely on the textual basis of A261-2/B317-18. Houston Smit, in "The Role of Reflection in Kant's Critique of Pure Reason," also appears to endorse a version of the comparison model, arguing that reflection is the representation of relations among one's representations (211), either with one another for the purpose of representing a form of judgment ("logical" reflection), or a form of cognition ("transcendental" reflection, 213).

In these passages Kant characterizes two notions of "reflection"—viz. logical and transcendental—in terms of the activity of comparing, respectively, a representation either with another representation or with the faculty from which it stems. Kant also consistently links reflection to the comparison of representations for the purpose of generating concepts in the logic and the metaphysics lectures, *reflexionen*, and in various other passages in the first *Critique* (see, e.g., *Logik Jäsche* 9:94; cf. R409, 15:165-6 (1772–1779?); *Metaphysik* L_1, 28:233-4 (1777-80); A86/B118; see also B1, A66/B91; *On a Discovery*, 8:222-3; see also the logic texts cited by Merritt, 51-2). Finally, Kant explicitly construes reflection in terms of the activity of comparison in the First Introduction to the third *Critique*:

> To reflect (to consider) [*Reflectiren* (Überlegen)], however, is to compare and to hold together given representations either with others or with one's faculty of cognition, in relation to a concept thereby made possible. (20:211)

The connection of reflection with comparison seems further supported by the fact that Kant indicates that there are rational (specifically, discursive) and non-rational forms of reflection in the *Critique of Judgment*, quoted above, in (E), where he attributes reflection to both (non-rational) animals and humans. Kant there glosses reflection as either that by which one attains a concept, or in the animal case, as that by which some inclination is "determined" [*bestimmt*]. Kant also elsewhere allows that animals may compare representations in order to explain their ability to discriminate similarities and differences amongst things (e.g. *Logik Jäsche* 9:64; *Logik Wien* 24:845-6 (c.1780/81)).[5]

Merritt is concerned to deny that Kant identifies "reflection" with the activity of comparison (35-6), construing it as rather only "involving" comparison (36). But given what Kant says above, one might wonder why it isn't simplest to construe reflection as a kind of comparative activity. Reflection is the activity whereby one attends to one's representations for the purpose of comparison, and where the exact relata of this comparative activity determine the

[5] Kant does at times deny that animals are capable of reflection (e.g. *Anthropologie Friedländer* 25:474 (1775/76); *Anthropologie Mrongovious* 25:1319 (1784/85)), but the context makes clear that what Kant denies is the capacity of an animal to "reflect" in the sense of making a *reflective judgment* and thus subsuming an object (or a concept) under a concept.

exact nature of the reflection involved.[6] Thus, in the case of logical reflection one is comparing one's representations for the purpose of generating a concept, in the case of transcendental reflection one compares one's representations with one's cognitive powers in order to determine their faculty of origin. In the case of non-rational reflection, an animal compares representations for the purpose of generating a desire or inclination to act in one way or another.

1.1.2 Reflection-C as (Pure) Apperception

Merritt allows that comparison is a "mode" of reflection, but specifically that it is a mode of reflection-c. If this were true then there would have to be a connection between comparative activity and thinking, since reflection-c is supposed to be present in any thinking whatsoever. Why think there is such a connection? Merritt relies heavily on Kant's *Anthropology* note that claims that

> consciousness of oneself (*apperceptio*) can be divided into that of reflection [*Reflexion*] and that of apprehension. The first is a consciousness of understanding, pure apperception; the second a consciousness of inner sense, empirical apperception. (7:135, note)

According to Merritt, Kant *identifies* reflection and pure apperception in this passage. However, she concedes that this "remark does not unambiguously entail an identity between reflection and pure apperception" (27) but argues that the kind of reflective comparison of representations at work in generating concepts (i.e. "logical reflection") requires "the possibility of recognizing that *one's own thinking* is the source" of the unity of a concept (28; emphasis in original). Merritt takes this to mean that reflection is either identical with pure apperception or is merely "notionally" distinct from it.

If the text is ambiguous then why make the strong claim about identity? Recall that, among the various uses of reflection mentioned above were:

[6] Note that this proffered view is meant to be distinct from the "received view" criticized in Merritt, "Varieties of Reflection in Kant's Logic." It does not take the controversial *Logik Jäsche* passage at §§5–6 9:93–5 as central, nor does it construe reflection as always aimed at concept generation, which would, in any case, fail to explain the sense in which non-rational animals might reflect.

A. The activity characteristic of the discursive intellect generally
B. Activity connected with pure apperception
C. A condition of the possibility of concept generation

Merritt's view is that (A)-(C) are all versions of reflection-c, the (typically) "tacit handle on oneself as the source of a point of view" (18), which is always present in all thinking. (A) and (C) plausibly are captured by the comparative activity involved in concept generation that Kant discusses in the Amphiboly and in various logic texts. (A) is also plausibly connected to concept generation, since Kant's remark in the *Prolegomena* concerning the fact that the understanding does not intuit, "but only reflects," is presumably a remark about the understanding as a discursive faculty for cognition. Note, however, that Merritt's account is not alone here. The comparative account can also plausibly connect (A) and (C) in a similar manner.

However, it *is* a problem for Merritt that (A) and (C) are episodic, and thus certainly not "constitutive" in the sense of reflection-c, that is, as "always going on, by sheer default, inasmuch as one manages to think at all" (28). I take it that Merritt justifies her interpretation by means of a connection of (A) and (C) with the pure apperception of (B). We can capture this in the following inference:

1. Reflection is identical with pure apperception
2. Pure apperception is a constitutive condition for all thinking, and for all synthetic activity more generally
3. ∴ Reflection is a constitutive condition for all thinking, etc…

That's a valid argument, but I think there are good reasons for not accepting (at least without further defense) either premise (1) or (2).

Concerning the first premise, it seems to me that there is a straightforward alternative to the identification (or mere "notional" difference) of reflection with pure apperception that still accommodates *Anthropology* 7:135, which is that reflection is sufficient, though not necessary for (and thus not identical with) pure apperception. It is thus a form of being intellectually conscious of oneself even if one can be so self-conscious without reflecting.

But if reflection is merely sufficient rather than necessary for pure apperception then it is not the case that reflection is identical with pure apperception. Therefore it is not the case that, as Merritt holds, pure apperception is "a reflection that is always going on…inasmuch as one manages to think at all" (28). It is certainly true that reflection, in the senses at issue in the Amphiboly and logic lectures, might be *necessary* for thinking in so far as it is necessary for the grasp or acquisition of concepts by means of which one thinks (the material of thought as it were), but this does not mean that in thinking one also reflects, or that reflection is otherwise always occurring.

Moreover, even if we were to accept the first premise in the above argument, there is a still a problem with the second. Kant famously says that the "I think" of pure apperception must merely *be able to* accompany all thinking (B131). However, premise 2 is a stronger claim. It says that in thinking (or related synthetic activities) one *is* purely apperceiving. Merritt needs the stronger claim to get the result that reflection-c occurs whenever thought occurs. While I am sympathetic to the position that pure apperception is a form of self-consciousness that is present in mental activity even with not explicitly signaled by the "I think," I'd like to hear more about the connection (on its face it is much more obviously Fichtean than Kantian). What justifies the move from a possible apperception to actually apperceiving?

If Merritt fails to make the case that reflection is pure apperception, then she is unable to show that there is a genus of reflection that is constitutive of thinking or intellectual mental activity generally, and so a use of "reflection" that unites the various uses specified above. In particular, reflection as the activity of attending to one's representations for the purpose of comparison, which is the paradigmatic form of reflection that Kant discusses in both his published work and his logic lectures, does not obviously connect with the constitutive sense of reflection advocated by Merritt.

What of the duty to reflect? Whether or not Kant thinks that reflection is something we are constitutively engaged in as thinkers, Merritt may well be correct that he regards reflection as something we ought to be doing. I turn to this issue next.

1.2 The Duty to Reflect

Merritt notes that there is a fair bit of textual evidence that Kant conceives of "reflection," in some sense of that term, as a duty or requirement for judgment. This textual evidence, along with the fact that Kant typically discusses reflection in connection with combating what he calls "prejudice" forms the basis for her position that there is a distinctively normative sense of "reflection"—her "reflection-n"—at work in Kant's critical writings. I'm not convinced that the textual evidence supports her claim, and I think there is another way of understanding the connection between reflection and prejudice that undermines a distinctively normative sense for "reflection." Let's take these points in turn.

Here are some central texts:[7]

From the *Logik Blomberg*:

> To reflect is to compare something with the laws of the understanding. To investigate, however, is actually to reflect mediately. Concerning many things we can quite well cognize without investigation what is true, what false. But reflection, on the other hand, is always necessary for any judgment, and for the distinction of the true from the false, even if it be in general, or in a [particular] cognition, etc., in all cases indispensable. (24:161)

From the *Logik Jäsche*:

> The cause of this deception is to be sought in the fact that subjective grounds are falsely held to be objective, due to a lack of reflection, which must precede all judging. (9:76)

> [E]ven if we can accept some cognitions, e.g., immediately certain propositions, without investigating them, i.e., without examining the conditions of their truth, we still cannot and may not judge concerning anything without reflecting, i.e., without comparing a cognition with the power of cognition from which it is supposed to arise (sensibility or the understanding). (9:76)

From the Amphiboly of KrV:

[7] For a full list see Merritt, 50-1.

> [A]ll judgments, indeed all comparisons, require a reflection, i.e., a distinction of the cognitive power to which the given concepts belong. (A261/B317)

> This transcendental reflection is a duty from which no one can escape if he would judge anything about things a priori. (A263/B319)

Merritt presents Kant's position in these texts as if he were saying that reflection is a requirement of all rational judging subjects as such (29, and note 22). But I think this is not so clear.

First, the requirement to reflect discussed in the Jäsche texts and in the Amphiboly is not a requirement to "reflect" *in all its senses*.[8] Rather, Kant is concerned to show only that one must *trancendentally* reflect, which he glosses as the comparison of one's representations (or cognitions in particular) with their faculty of origin (or "power of cognition"). Transcendental reflection allows us to discern the characteristics objects would have if we had purely intellectual or conceptual cognition of them, which we can then compare with the characteristics of the objects we actually experience.[9] According to Kant, such comparison exhibits four different discrepancies between the characteristics objects would have if cognized purely intellectually or conceptually and those cognized via our actual experience. Ignoring such discrepancy results in what Kant calls "a transcendental amphiboly, i.e., a confusion of the pure object of the understanding with the appearance" (A270/B326) or of the purely "logical" use of a concept with its "real" use in experience.

Second, in the Amphiboly the requirement to reflect is a conditional or hypothetical one. One must transcendentally reflect *if* they would judge anything about things a priori. Why is this a requirement? As the discussion of the Amphiboly makes clear, Kant considers a priori judgment as subject to error—subject to "false pretenses of the pure understanding and illusions arising therefrom" (A268/B324)—which transcendental reflection helps to dispel. The metaphysician, who aims at giving a true account of reality via pure a priori reasoning, must thus engage in transcendental reflection or risk serious error in their theory, error that Kant illustrates

[8] What about the *Blomberg Logik* and related texts? Kant is reported there and in the other lectures Merritt canvasses to talk of reflection as "comparison with laws of the understanding," which seems rather close to transcendental reflection. Since the Amphiboly account is consistent with the *Blomberg Logic* and is a published work, I take it as having precedence.

[9] See Pereboom, "Kant's Amphiboly," 57 for this kind of account.

and criticizes in his examination of Leibniz/Leibnizian philosophical commitments in the latter part of the Amphiboly.

Moreover, Kant seems to construe the requirement to reflect in judgment as stemming from a requirement to reflect on all (logical) acts of comparison (A261/B317).[10] He clearly does not mean here that one cannot actually judge a truth-functional content without engaging in reflection, or at least, this much seems clear from his claim that many judgments are "accepted out of habit" in that "no reflection preceded" them (A260/B316). Why then might transcendental reflection be required for *comparison*, and in a way that relates to judgment? I take Kant's point here to be that without such reflection one always has the potential to fall into amphiboly, i.e. to make errors based on ambiguities concerning the concepts one uses. Since comparison is a condition of concept formation, and (objective) concepts are free of all ambiguity, any act of "mere" logical comparison in the formation of a concept is going to depend on transcendental reflection as to the source of the concept in either the intellect or in sensibility. This does not seem to me to be a new or different sense of "reflection" but rather an explication of why (transcendental) reflection is important.

If the texts themselves do not clearly point to a distinctively normative sense of "reflection" with respect to judgment, what about the fact that Kant links the need to (transcendentally) reflect with combating the three natural sources of prejudice: habit or custom (*Gewohnheit*), inclination (*Neigung*) and imitation (*Nachahmung*). All three prejudicial sources constitute principles for associating representations. They are thus ways of fixing the acceptance of a propositional content due to aspects of one's sensibility or non-rational nature, rather than due to the exercise of one's rational spontaneity or intellectual faculty. If Merritt is correct that Kant's view is that prejudice is a case of unreflective judgment, then insofar as one should avoid prejudicial judgment, one would have a duty to reflect.

It is crucial to Merritt's argument that a distinctive "normative" sense of reflection would be one according to which reflection is a condition of judging "well," though one could still count as judging in the relevant sense without such reflection (29).[11] However, there are some

[10] It is this section that is quoted almost verbatim in the *Jäsche Logik*, and which is largely the same in the other lectures as well.

[11] This is in keeping with the idea that something may be targeted for normative assessment even if it fails to meet the requisite norm. For discussion see Tolley, "Kant on the Nature of Logical Laws"; "Kant and the Normativity of Logic."

reasons to doubt whether this is true. That is, there are reasons to think that Kant's view is that one altogether *fails* to judge in the relevant sense when one does not reflect. Let me explain.

There is a consistent ambiguity in Kant's use of the term 'judgment' (*Urteil*) (e.g. 7:146). He sometimes uses it to denote the act of contentful, truth-functional representation (what we might now describe as the grasp of a propositional content—call this "judgment$_C$"), and other times uses it to denote the *acceptance* of the truth of such a content—call this "judgment$_A$."[12] Occasionally Kant will signal the distinction between content and acceptance by using the transitive verb '*beurteilen*' and its nominative '*Beurteilung*' for denoting the act of assertion or acceptance (e.g. CPrR 5:57–8), and the nominative *Urteil* for denoting the content asserted. But Kant does not always do this. Nevertheless, the ambiguity here is often harmless, since it is typically the case that the grasp of a content involves some stance towards its truth, even if only to take its truth as "problematic" in the sense of being assertible though not asserted (see A74/B99-100).

Without reflection, one rationally "judges," at most, only in the sense of grasping a truth-functional content—i.e. judgment$_C$. In order to *accept* a judgment (i.e. judge$_A$) in a way that counts as an exercise of one's *rational* capacities, and thus one's spontaneity, one must fix one's acceptance of a content by virtue of epistemic and logical laws and not via associative mechanisms tied to sensibility, such as custom, inclination, or imitation. Content accepted via such mechanisms is merely the "mixed effect" (*vermischten Wirkung*) of sensible forces on the intellect, and Kant says that one "errs if one takes this mixed effect to be a judgment of the understanding" (R2244 16:283 (1773–78?); cf. A294/B350, *Logic Jäsche* 9:53-4, *Logik Wien* 24:824-5 (1780-2), R2142 16:250 (1776-1781)).

Hence, in the case of *irrational* acceptance, one is not really *acting* at all. In cases of prejudice, or more broadly, any case of "persuasion" (*Überredung*), the content (the judgment$_C$) is *being fixed* in one's mind by sensible impulse rather than its being the case that one's own intellectual activity (i.e. one's reason) does the *fixing* (the acceptance) of the content.

[12] For discussion of Kant's view of doxastic attitudes in general and notion of acceptance in particular see Chignell, "Belief in Kant." One might worry that Kant does not adequately distinguish between predication and the acceptance or endorsement of the truth of what is thereby predicated. For such worries about figures in the Early Modern period see, e.g., Geach, "Ascriptivism"; Hylton, "The Nature of the Proposition and the Revolt Against Idealism"; Buroker, "Judgment and Predication in the Port-Royal Logic"; Owen, "Locke and Hume on Belief, Judgment and Assent." For discussion of Kant see McLear, "Kant on Perceptual Content," 106–12.

If I am right about the plausibility of an ambiguity in Kant's use of "judgment" then there is a problem when Merritt claims that good or "sound" judgment requires reflection, but that the very same judgment could occur, poorly, without such reflection. For example she says,

> Kant's considered view is that one exercises cognitive agency badly in prejudice, but not that one fails to exercise cognitive agency altogether. Taking things to be a certain way is an exercise of cognitive agency, even if we don't deliberately reflect on the soundness of those views and the principles on which they rest. (44)

Here I think Merritt slides from judgment$_C$ to judgment$_A$ in her use of the notion of "taking things to be a certain way." I agree with Merritt that grasp of a truth-conditional content—a proposition—via the cognitive process of ordering representations in one consciousness is a kind of cognitive activity, and at least to that extent an exercise of the subject's "agency." This is the generation of a judgment$_C$. But, in contrast to Merritt, I take Kant's considered view to be that one indeed *fails* to exercise one's cognitive agency when one *accepts* a judgment$_C$ due to sensible rather than rational conditions. As I take Kant's view, there is no stable target of normative assessment across cases where a judgment$_C$ is sensibly rather than rationally accepted. Failure to meet the conditions of rational acceptance means that one does not judge$_A$ rather than, as Merritt has it, that one simply judges$_A$ poorly.

Thus I think neither the textual evidence nor the philosophical position relating reflection to prejudice conclusively supports Merritt's position that there is a distinctive normative sense of "reflect" at work in Kant's critical philosophy. In what follows I raise some questions for Merritt's treatment of reason, and ultimately the sense in which the reflective ideal is overly demanding.

2. Reason

Merritt construes reason as "a single cognitive capacity admitting of distinct theoretical and practical employments" (113). This is in service of her argument that there is a general form of cognitive normativity, which her account of healthy human understanding captures, and which admits of further specification into theoretical and practical uses. The general claim about reason

is controversial in at least two respects. First, it construes the practical use of reason as *cognitive*. Second, it construes reason as a single unified power or faculty. I want to briefly raise a few questions concerning both of these claims. I then raise a third issue that seems important even if we accept Merritt's interpretation of reason.

First, why think that reason is fundamentally a cognitive faculty? In order to answer this we need to know more about what it means to say that reason is "cognitive." Since it isn't clear what Kant means by cognition (figuring this out has now become its own cottage industry), and Merritt never explicitly articulates her conception of cognition, I'm speculating somewhat. But I take it that Merritt construes the cognitive as aiming at knowledge, and thus of reason as a unitary faculty of cognition as a faculty for acquiring knowledge, specifically, knowledge of objects. Knowledge of an object, in turn, requires that the object be "something that obtains independently of any particular effort to come to cognitive terms with it; 'the good' by Kant's lights, is the object of practical reason" (116, note 3). So theoretical reason is fundamentally a capacity for knowledge of phenomena, while practical reason is fundamentally a capacity for knowledge of the good, but both hold in common the fact that they are capacities for knowledge of an object in the above sense.

I think there are at least three problems with this interpretive approach.First, while Kant often talks of theoretical knowledge, theoretical and practical cognition, practical belief, and practical wisdom (*Weltweisheit*), to my knowledge he *never*, in his entire written corpus, uses the phrase "practical knowledge" (*praktische/s Wissen*). Second, while it seems true that the *intellect*, understood as the "higher" faculties of understanding, judgment, and reason, is a faculty of or for knowledge, it isn't clear to me either that the intellect is *fundamentally* a faculty for knowing, or that *reason*, in the sense at issue in the first two *Critiques*, is such a faculty. Moreover, in the first *Critique* Kant characterizes reason as a faculty for making inferences, and more broadly, the faculty by which rational beings achieve *understanding* via the grasp of the explanatory relations holding amongst various discrete bits of knowledge, rather than knowledge per se. Third, virtually all of the evidence presented by Merritt for the claim that reason is a single cognitive capacity with different "modes of knowing" (115) concerns Kant's discussion of *cognition*. But cognition is not knowledge (e.g. there can be false cognition), and so while we should all readily agree that Kant allows for practical *cognition*, this is not the same as practical knowledge, nor does the fact that theoretical and practical cognition are species of cognition entail that the

theoretical and practical uses of reason are both modes of *knowing* in a univocal sense of that term.

Concerning the second issue, of reason as a single unified faculty, while Merritt discusses various texts in which Kant speaks of reason as a single cognitive faculty (e.g. G 4:391; KpV 5:121), or of practical and theoretical reason as derived from a "common principle" (Axx), as Merritt further points out (e.g. 120), Kant neither explicitly states what such a common principle or unifying ground might be and also consistently treats the unity of reason as a problem in search of a solution. For example, Kant speaks in the *Groundwork* of the task of uncovering the "common principle" that unifies practical and speculative reason, a task that he has to postpone until he can give a "Critique of Pure Practical Reason" (G 4:391). But then, in the second *Critique*, he hopes merely of "being able *some day* to attain insight into the unity of the whole rational faculty (theoretical as well as practical) and to derive everything from one principle" (5:91; my emphasis).[13] If the unity of theoretical and practical reason lies in their being basic modes of knowing, as Merritt suggests, then it is hard to see why Kant kept postponing saying so. Merritt rightly points out that one version of the "problem of the unity of reason" concerns the unification of the principles of nature with those of freedom, but this is not obviously what concerns Kant in the passages from the *Groundwork* or second *Critique* when he discusses the task of unifying the disparate uses of reason under one principle. So if we admit that Kant thought there was a task or problem requiring explanation—viz. how are the very different uses of *theoretical* and *practical* reason to be both construed as uses of a single faculty of *reason*—then it is unclear why we should accept Merritt's view of reason as a faculty for knowledge.

Finally, even if one accepts that reason is fundamentally a cognitive power along roughly the lines that Merritt suggests, this does not entail that the two fundamental modes of knowing—viz. theoretical and practical—are *harmonious* with one another. Indeed, Kant briefly discusses the possibility of a "conflict of reason with itself" that would arise if theoretical and practical reason "were arranged merely side by side" (KpV 5:121). Perhaps Merritt intends to rule out this potential problem by means of the argument that (i) reason is a faculty of knowledge and (ii) knowledge is always consistent with knowledge (to use Engstrom's phrase).[14] But it would be

[13] For discussion of these sorts of problems see Kleingeld, "Kant on the Unity of Theoretical and Practical Reason"; Williams, "Kant's Account of Reason."
[14] See Engstrom, *The Form of Practical Knowledge*, 14.

good to hear more about why (ii) is true, especially since Kant seems to worry in the *Critique of Practical Reason* that it is not.

3. The Demands of Reflection

I want to close with a few questions surrounding the problem of the demandingness of Kant's reflective ideal and Merritt's solution to this problem. Merritt worries that if we take Kant seriously about the duty to reflect, this demand will be, if even plausible, far too onerous when taken as a demand to engage in discrete acts of reflection every time we judge. Merritt's solution is to construe the demand to reflect as a demand to be a certain kind of cognitive agent.

> the requirement to reflect-n properly lodges at the level of character, rather than piecemeal over individual acts of judgment: the requirement is to take the appropriate interest in one's cognitive agency, where this interest governs cognitive activity globally. (81)

We can see one clear advantage of Merritt's interpretation—the demand to reflect need not be construed as a demand to engage in an episodic activity of "stepping back" on the occasion of each and every judgment. We do not need to "step back" and deliberate every time we make a judgment or form a belief. But here's one worry: haven't we just traded one overly demanding conception for another? Merritt's interpretation saves the reflective agent from the task of near-constant deliberation but leaves such an agent with the perhaps equally demanding task of improving their cognitive virtue. This seems to be both as near-constant as the caricatured version of over reflection itself and just as difficult to achieve. For example, Merritt claims that

> a central question that the reflective person bears in mind, and continually returns to, is: *What am I paying attention to, and why?* This is an essential part of what it is to take an interest in one's own cognitive agency (107; emphasis in original)

But asking this question, and having the right set of cognitively virtuous dispositions to understand what counts as a good answer (or set of answers) to this question sounds pretty

demanding! Indeed, it sounds, if not more, at least *as* demanding as the "precious, hyper-deliberate and repugnantly moralistic" caricature of the reflective ideal that Merritt seeks to replace.

Here's another worry: since no actual person is perfectly cognitively virtuous, one can always raise the question, "Did I make this judgment in a cognitively virtuous manner?" (i.e. did it come about due to rational as opposed to merely natural causes). In asking such a question one's rationality is open to doubt. It seems that the curative here would be engaging in an episode of reflective deliberation on the grounds of one's judgment and the process(es) by which one acquired the belief. Now, to the extent that a cognitive agent is at all virtuous they should admit that they are not *perfectly* virtuous. So why isn't it the case that every exercise of their cognitive skill isn't also one for which they need to ask the question, "was this a virtuous exercise?" and thus one that requires (the demanding, precious, repugnantly moralistic, etc.) overt deliberation? But then we're back to the old "caricature" of the overly-demanding requirement to reflect.

Bibliography

Buroker, Jill Vance. "Judgment and Predication in the Port-Royal Logic." In *The Great Arnauld and Some of His Philosophical Correspondents*, edited by Elmar J. Kramer, 3–27. Toronto: University of Toronto Press, 1994.

Chignell, Andrew. "Belief in Kant." *The Philosophical Review* 116, no. 3 (2007): 323–60.

Engstrom, Stephen. *The Form of Practical Knowledge*. Cambridge, MA: Harvard University Press, 2009.

Geach, Peter T. "Ascriptivism." *The Philosophical Review* 69, no. 2 (1960): 221–25.

Hylton, Peter. "The Nature of the Proposition and the Revolt Against Idealism." In *Philosophy in History*, 375–97. New York, 1984.

Kleingeld, Pauline. "Kant on the Unity of Theoretical and Practical Reason." *The Review of Metaphysics* 52, no. 2 (1998): 311–39. doi:10.2307/20131142.

Longuenesse, Béatrice. *Kant and the Capacity to Judge*. Princeton: Princeton University Press, 1998.

McLear, Colin. “Kant on Perceptual Content.” *Mind* 125, no. 497 (2016): 95–144.

Merritt, Melissa McBay. *Kant on Reflection and Virtue*. Cambridge: Cambridge University Press, 2018.

———. “Varieties of Reflection in Kant’s Logic.” *British Journal for the History of Philosophy* 23, no. 3 (2015): 478–501.

Owen, David. “Locke and Hume on Belief, Judgment and Assent.” *Topoi* 22, no. 1 (2003): 15–28.

Pereboom, Derk. “Kant’s Amphiboly.” *Archiv für Geschichte der Philosophie* 73, no. 1 (1991): 50–70.

Smit, Houston. “The Role of Reflection in Kant’s Critique of Pure Reason.” *Pacific Philosophical Quarterly* 80, no. 2 (1999): 203–23.

Tolley, Clinton. “Kant and the Normativity of Logic.” *Recht Und Frieden in Der Philosophie Kants: Akten Des X. Internationalen Kant-Kongresses* 5 (2008): 215–27.

———. “Kant on the Nature of Logical Laws.” *Philosophical Topics* 34, no. 1 & 2 (2006): 371–407.

Williams, Garrath. “Kant’s Account of Reason.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University, 2017. https://plato.stanford.edu/archives/win2017/entries/kant-reason.

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.

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.

Kant on Reflection and Virtue
Response to Colin McLear, Francey Russell, and David Sussman

Melissa Merritt

I want to thank the Society for German Idealism and Romanticism, and especially Keren Gorodeisky, for organising this series of reviews. I would also like to thank Colin McLear, Francey Russell, and David Sussman for taking the time to read my book and respond with such great care. Your comments were extremely thoughtful, and pushed me to think harder about some of my claims in the book, and also about where I want to go from here.

1.

One of the puzzles that drove my research early on was the enormous complexity in the textual record on reflection. And yet the secondary literature on reflection was somewhat over-focused, I thought, on a couple of paragraphs in the *Jäsche Logic*, which was not a text that Kant wrote or prepared for publication himself. The *Jäsche Logic* passage was about three mental acts in the formation of concepts, one of which is called reflection. Before long, commentators were taking the *judgment of reflection* at issue in the third *Critique* simply *to be* the reflection mentioned in the *Jäsche* paragraph on concept formation.[1] This seemed like a mistake to me (but not one that I felt able to correct in my book). Another important text on reflection is found in the first *Critique*, in the Amphiboly chapter. It discusses reflection in ways that I initially connected with Kant's conception of enlightenment, and eventually came to see as part of the "humanist" side of Kant's logic[2] — specifically the "applied," rather than pure, general logic that deals with the messiness of the human epistemic condition, such as our susceptibility to prejudice, as well as with other topics that interested me, like the direction of our attention in experience. Kant develops these concerns in his later writings, chiefly in the

[1] Henry Allison, *Kant's Theory of Taste* (Cambridge: Cambridge University Press, 2001), is representative of this tendency.
[2] As Huaping Lu-Adler observes in *Kant and the Science of Logic* (New York: Oxford University Press, 2018), Kant is taking up a humanist tradition as he develops applied logic.

Anthropology. While I wanted to be able to find my way, on principled systematic terms, through the complex textual record on reflection, ultimately I realised that my chief interest was to understand what Kant's talk about a requirement to reflect means for us at the ground level.

To sort out the complexity of the textual record, I drew a distinction: namely, between senses of reflection that seem to concern what is constitutively involved in thinking as such, on the one hand, from a normative sense of reflection that Kant characterises, variously, as a requirement or even a duty. Colin McLear is not convinced by this fundamental distinction that I propose. He also notes that the record is even more complex than I acknowledge, since there is a passage in the First Introduction to the third *Critique*, where Kant claims that reflection "goes on even in animals, though only instinctively" (20:211). I set that passage entirely aside: I don't know how Kant can talk with such confidence about the mental lives of non-rational animals — and, at any rate, I am concerned with reflection as an expression of rational mindedness.

Some of the earlier literature on reflection identified reflection with *comparison*, based partly on remarks from the *Jäsche* passage on concept formation.[3] I believe that McLear endorses this identification when he points to places where Kant seems to explain reflection as a kind of comparison: e.g., in transcendental reflection, one compares a given representation with the cognitive capacity from which it should arise; and in logical reflection, one compares representations for the purpose of generating a concept. However, I think that comparison is simply too indeterminate a notion to give us a handle on what reflection is. As I argue in the book, if one invokes the notion of comparison, one must also invoke — at the very least — a standard of comparison: x and y are compared in length, or wisdom, or hue, and so on. And then we have to consider the point of making the comparison, for which we must invoke some *end* or governing principle. So, I just do not think it will do, in the face of such a complex textual record, to simply say that reflection is comparison. Comparison is involved in reflection, as I said in the book; but they are not the same thing.

McLear discusses Kant's notion of transcendental reflection from the Amphiboly chapter. I set that aside in the book, although I did discuss it in a 2015 paper (part of which appears in the book).[4] So I am not going to take up that aspect of his comments at much length.

[3] See, e.g., Béatrice Longuenesse, *Kant and the Capacity to Judge* (Princeton, NJ: Princeton University Press, 1998).

[4] Melissa Merritt, "Varieties of Reflection in Kant's Logic," *British Journal for the History of Philosophy* 23, no. 3 (2015): 478-501. I am not going to try to defend my reading of *transcendental* reflection from that paper

However, let me briefly say that one of my goals in that paper and the book is to direct attention to the overlooked aspect of Kant's logic: namely his conception of *applied*, rather than pure, general logic. As I mentioned at the outset, Kant directly nods to applied logic, and its dealings with our susceptibility to prejudice, at the start of the Amphiboly. That is what is on his mind when he talks about transcendental reflection. According to applied logic, prejudices are fundamentally expressions of a kind of *epistemic self-conceit*. The point comes out nicely in the Blomberg logic (one of the earlier records of Kant's logic lectures, where his interest in the logical humanist tradition is developed at particular length): "a prejudice is indeed nothing other than the mere desire to want to judge, but without the proper acuity or reflection" (24:187).[5] Now I do think that McLear may be overlooking the significance of this sort of point from applied logic; so I will try to sketch where that occurs and what its implications are.

Kant with great consistency names three sources of prejudice: *custom*, *inclination*, and *imitation*.[6] McLear is right to point out that they "constitute principles for associating representations." In this context, McLear suggests that an ambiguity in Kant's notion of *judgment* makes mischief when it comes to Kant's applied-logic claim that "all judgments require reflection" (A261/B317). McLear suggests that Kant sometimes takes *judgment* "to denote the act of contentful, truth-functional representation (what we might now describe as the grasp of a propositional content)" and yet "at other times uses it to denote the *acceptance* of the truth of such a content." I haven't worked through the textual record with this matter in mind, but for the sake of argument I will accept this observation. Now, McLear suggests that the sense in which one "judges" without reflection — *as when Kant explains prejudice as a judgment without reflection*[7] — is at most the first sense just distinguished, i.e. grasping a truth-functional content. The sense of judgment that involves endorsement or acceptance of that content can occur only "in virtue of epistemic and logical laws and not via associative mechanisms tied to sensibility, such as custom, inclination, or imitation." Thus McLear says that the prejudice is not really judgment, because it occurs in us owing to the force of these associative mechanisms.

---

here: first, I am not entirely sure that I remain satisfied with it; and at any rate, it does not figure in *Kant on Reflection in Virtue*.

[5] Kant's works are cited, as is customary, by volume and page of the German Academy of Sciences edition. The exception is the *Critique of Pure Reason*, which is customarily cited by the pagination of the first and second edition of the text, abbreviated A/B. The following other abbreviations are used: Anth=*Anthropology from a Pragmatic Point of View* (Ak. 7); MS=*Metaphysics of Morals* (Ak. 6); Rel=*Religion within the Boundaries of Reason Alone* (Ak. 6).

[6] See *Kant on Reflection and Virtue* (29) and note 24 for references to Kant's works.

[7] See *Kant on Reflection and Virtue* (29), and note 23 for further references to Kant's works.

But if this were correct, it would be hard to see how we could make sense of Kant's view that prejudice is *culpable*: we are responsible for it; and we ought to free ourselves from it. Part of what is going on here is, I think, a tendency to confuse the sources of prejudice from prejudice itself. Kant is *quite clear* that there is nothing culpable about custom, inclination, and imitation: these are just facts of human psychology, and nothing we could coherently aspire to free ourselves from. But prejudice is culpable, and this is why Kant's applied logic — and the discussion of reflection in this context — takes on a moral, even moralistic, cast at times. And what this shows, I think, is that we need my distinction between reflection-c and reflection-n to make sense of what is going on. Prejudice has to involve assent: that's why it is culpable. How do we understand this? I think Kant's apperception principle is a good guide: "The *I think* must be *able* to accompany all of my representations" (B131). Contrary to what McLear suggests, I do not misunderstand the modality of this. I do *not* think that Kant thinks that the "I think" is always actually accompanying all of my representations. The apperception principle says that a rational being is necessarily — constitutively, in virtue of its being the sort of mind that it is — able to consider its thoughts as its own. That sort of deliberate activity, whether undertaken on the occasion of a given judgment or undertaken in the adverbial mode I suggest later in the book, broadly falls under the heading of the normative conception of reflection. When you are thinking but not self-consciously accompanying these thoughts with the "I think," you are still moving from one thought to another in ways that show some sensitivity — even if imperfect — to cognitive norms like coherence. You aren't suffering random, unconnected states of consciousness; you are *thinking* one thing, which bears on how you can think other things, and so on. That involves having some kind of grip on yourself as the source of these thoughts as you think them, but it doesn't require that you are *doing* anything directly occupied with your own agency as a thinker. This grip on yourself, which comes online once you have come into the use of your reason, transforms what can figure as a mental state, and what it is like to think them. And if you don't have that at least tacit grip on yourself as a thinker, then I don't see how you can be thinking at all.

What then about taking things to be a certain way on the basis of prejudice? This is still directed by the shape of one's own mindedness. Prejudices are not random thoughts that are just induced in one by mechanical forces. They are the result of one's *taking* things to be a certain way — whether through custom, or the influence of prestige, or through inclination (habitual desire), and so on. Therefore, McLear overstates things when he claims that prejudice

and other sorts of judging of the first, merely content-presenting, variety is a *failure* to exercise one's cognitive agency altogether. This view does not make sense of the fact that we are epistemically responsible for prejudice, even though we are not meeting any number of epistemic standards or ideals when we allow our thought to be so driven. They are, indeed, an expression of our cognitive agency, but a corrupted one. Applied logic, which is concerned with this problem, is a therapeutic guide to the fallen human (epistemic) condition. And the sort of problems it takes up require the distinction between reflection-c and reflection-n.[8]

Finally, McLear raises doubts about my view that Kant takes reason to be a cognitive capacity through and through (and not just in its theoretical exercise); this claim is important for the transition from the basic cognitive virtue (which I associate with "healthy human understanding") to Kant's conception of moral virtue in the second half of the book. In Chapter 4 of the book, I present the textual evidence — which I do regard as overwhelming — that Kant takes reason to be a cognitive capacity in its practical exercise. McLear replies that Kant never uses the phrase "praktisches Wissen," although he acknowledges that there are plenty of mentions of practical cognition (*Erkenntnis*) and even some of practical wisdom (*Weisheit*). But nothing should ride on the former, he contends, since "cognition is not knowledge (e.g. there can be false cognition)." That is fine by my lights. Curiously, McLear does not seem to see *Weisheit* — wisdom — to be a kind of knowledge. I don't know how else to understand what "wisdom" could be, if not *some* kind of knowledge. Now, it may be that the non-wise person will have practical judgments of McLear's merely content-representing kind; however, I would say that there is normally tacit endorsement of what is thus represented, which is why passion, like prejudice, is culpable. Of course, we ought to be seeing things, and to be motivated, in the manner of someone who is wise, or has practical knowledge of the good. Kant, moreover, indicates in the Introduction to the *Metaphysics of Morals'* Doctrine of Virtue that he takes "*wisdom* in the strict sense, namely practical wisdom" — wisdom properly understood — to be nothing other than virtue itself (6:405).[9]

2.

[8] One could say that reflection-c is the capacity that reflection-n exercises. (This seems to be what Sussman suggests in a footnote to his comments.) That would be fine, as long as we clarify that reflection-c transforms what it is to have mental states, and what they can be like. In the book (Chapter 3, pp. 107-110) I suggested that the distinction between them may ultimately be notional in the sense that the coming online of the one will always bring in its tow the coming online of the other; but the distinction between the two is required in the *account* of the requirement to reflect.

[9] As he says this, he nods to the Stoic tradition, a point that will come up again in these remarks.

As it happens, that remark about virtue takes place in a stretch of the Doctrine of Virtue that I've more recently come to think of as "the Stoic interlude": it runs several pages (MS 6:405-409), includes a sympathetic allusion to the so-called Stoic paradoxes (only the sage is healthy, rich, free, and so on; non-sages are fools, who cannot genuinely be any of those things), discusses affect and passion at some length, and concludes by endorsing the Stoic principle of apathy as necessary for virtue. This brings me to Francey Russell's acute remarks about the affect-passion distinction. I think she is quite right in the problems she points to in my account of affect in particular; and my views have evolved somewhat as I have begun to work on a new project on Kant and Stoicism. So, I will begin by recapping what is quite correct in Russell's comments on this issue, then I will sketch how I now see the affect-passion distinction and its relation to reflection. Then, I will turn to some of the aspects of Russell's comments that are more directly related to the topic of moral virtue.

Maybe I should begin by expressing some reservations I now have about the fact that I effectively began the book with the discussion of affect and passion. I was trying to change the orientation of existing discussion on reflection from its narrow focus on logical reflection, i.e. the reflection that is involved in concept formation and which is a topic for pure general logic. So, part of my agenda was trying to get people to see that there is another notion of reflection that figures in *applied* general logic: i.e. reflection-n. But I was also trying to make the case that the topic of reflection is not just a concern for logic, whether pure or applied; that it has relevance for practical philosophy as well. Moreover, the way Kant distinguishes between affect and passion does — and I still think this — support my view that a distinction between constitutive and normative requirements to reflect is needed. But I now see that Kant's discussion of affect and passion draws quite directly from the psychological monism of the Socratic and Stoic tradition, and in its details, particularly from Seneca. My failure to appreciate this clearly when I wrote *Kant on Reflection and Virtue* largely accounts for my error in interpreting his notion of affect.

Let's start with my error. Now, Kant says that passion (*Leidenschaft*) involves reflection; indeed, he appears to distinguish affect from passion precisely on that ground. (I am going to use German terms here to avoid terminological confusion.) Since Kant speaks of *Leidenschaft* as a culpable disease — "a cancerous sore of practical reason" (Anth 7:266) and something that one brings on oneself — it was clear that it could not be meeting the normative standard of reflection. At the same time, *Leidenschaft* quite obviously involves taking a practical point of view on what matters, what is worth doing; and so, I took it to involve reflection in the constitutive sense. It is the expression of an evaluative point of view — only,

it turns out, a confused one. I still stand by that much of my story. The problem comes with affect (*Affekt*), which — with this set-up — then seems to be something that lacks evaluative or practical point of view altogether, since Kant characterises it as either making reflection impossible, or else just very difficult. He is vague. Rather than seeing the nuance here, I ran with the idea that affect is incompatible with reflection. But since *Leidenschaft* already lacks reflection in the normative sense, I took it that *Affekt* must lack reflection in the constitutive sense. I bit that bullet. Yet if that were right, affect would be something very extreme indeed. An example might be the sort of rage that leaves you shaking and disoriented afterwards. (I have experienced this once or twice; I hope you have not.) Russell suggests that *Affekt*, so understood, would only be at the extreme end — "something more like a trauma-level emotional experience." I think that is right; and moreover, such an extreme interpretation of affect does not fit many of the *examples* Kant gives of affect in the *Anthropology*, as Russell indicates. It might perhaps fit extreme levels anxiety and fear; but it certainly cannot accommodate, for example, Kant's inclusion of admiration (*Bewunderung*) as an affect. Admiration is the mode of our enjoyment of natural sublimity by Kant's lights, and it involves a kind of astonishment in finding a wisdom in the order of nature that one did not expect to find (*Anthropology* 7:261): this is not obviously viable in someone who has *ex hypothesi* momentarily lost all grip on herself as a thinker and an agent.

Moreover, the extreme interpretation of *Affekt* as lacking reflection-c would be incompatible with psychological monism. Affects would be something that blows in from outside rational mindedness. The way that Russell puts this is to say that, on this reading, "much of our emotional lives would be oddly unexperienceable." The way I was inclined to think of this in *Kant on Reflection and Virtue* was through an idea of *simplicity*: there's just searing feeling that, at least for a time, overwhelms any capacity one might have to think even in terms of good and bad. There's just some variation on extremely intense pleasure, or extremely intense pain. The problem here is to understand what Kant means when he says that they "more or less" lack reflection. As I said, I think the *examples* that Kant picks make clear that he can't take this to mean that a sufferer of *Affekt* entirely lacks a point of view on how things are. If I am afraid, I see something as threatening, and so on. How should we understand *Affekt* then? I think this is a complex topic — and I think Seneca's version of the Stoic theory of emotions or passions (*pathē*) is directly on Kant's mind. I can't go into those details here, but I don't think there can be any loss of what I am suggesting is constitutive of rational mindedness, the typically tacit grip on oneself as the source of a point of view. What I can do here is at least point to the fact that Kant takes *Affekt* to be an expression of the faculty of

feeling, not the faculty of desire; and thus *Affekt* is not, simply as such, an engagement of one's agency.

Passions (*Leidenschaften*), on the other hand, are expressions of rational mindedness, and involve taking a practical point of view on what matters. They involve maxims. Glossing this, Russell suggests that *Leidenschaften* involves "the *ersatz exercise*" of the capacity to reflect, "that is a habitual and perverse misuse" of it. I think this is mostly right. What I think it might miss is that *Leidenschaften* involve substantive evaluative confusion, perhaps along lines akin to the Stoic view of the mental infirmity or illness underlying the *pathē*. Cicero presents the Stoic view of this issue, explaining that the mental infirmity at issue is '"a vigorous opining that some object is worthy of pursuit which is in fact not worthy of pursuit, that opinion being deeply attached and rooted in the mind"' (*Tusculan Disputations* 4.26).[10] The *pathē* are expressions of assent to evaluative propositions — and in that sense are analysed as a kind of judgment. These judgments, according to the Stoics, are all false. They involve an evaluative confusion that is endemic to human mindedness. Substantively, they involve taking to be genuinely good what is not really good but may be preferable and worthy of selection for its value in making things go well for one. The root of the problem is an epistemic failure, namely our propensity to *opine*. Citing the old Stoic sources, he says: "'opining' is when a person judges that he knows something which he does not in fact know" (*Tusculan Disputations* 4.26). So opining is the expression of a default presumption of taking oneself to understand, or know, when in fact one is in no such position.

Now, this should ring a bell: for this is exactly what is *generally* at issue in prejudice by Kant's lights — namely, a default presumption that one is a knower, that one understands, when in fact one is in no such position. We might ask why this should be a *default*. I think the answer lies in the constitutive notion of rationality: that to be awake and thinking is to be taking a point of view on how things are and what is worth doing. So the default is that one takes oneself to see how things are and knows what is worth doing. This brings me to Russell's remark about why passions are difficult to correct: she suggests that this is because "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 rationalisation." I think it misconstrues things to say the passionate person is rightly oriented

[10] For translation and commentary, see Margaret Graver, *Cicero on the Emotions: Tusculan Disputations 3 and 4* (Chicago: University of Chicago Press, 2002).

on these grounds; it is just an expression of default mindedness — of opining, if you will. But Russell is quite correct to note that the passionate person engages in (some sort of) reflection, and rationalisation.

On the topic of virtue, Russell raises one overarching issue, which — as I see it — has to do with the idea that virtue, as Kant evidently conceives of it, is a kind of struggle: "the moral disposition in conflict," as he says in the *Critique of Practical Reason* (5:84). As Russell rightly points out, this might be read either as a comment about an individual's moral life, or as a comment about our creaturely existence as human beings. As I see it, the second point is broadly at issue in Kant's relation to the Stoic tradition. What Kant disputes in Stoic ethics is principally the idea that its sage is supposed to have transcended human nature: this is arguably how the impossibility of backsliding, once one has attained sagehood, is understood. Note that Kant accepts the idea that virtue, once gained, could not be lost: it is there when he endorses the so-called Stoic paradoxes in the Doctrine of Virtue (6:405), and I think it figures in his account of virtue in Part I of *Religion within the Bounds of Reason Alone* (6:47-8). More precisely what Kant says there is that the default human condition is to make the maxim of self-love the governing principle of one's motivational economy, subordinating to it the moral law that is constitutive of reason in its practical cognitive employment. Becoming a good human being requires a "revolution," which Kant characterises as "a single and *unalterable* decision" by which "a human being reverses the supreme ground of his maxims" — subordinating now the maxim of self-love to the moral law (Rel 6:48; my emphasis). But no human being can ever know that this unalterable decision has been made; and to suppose that it has is to be guilty of moral fantasizing and the worst forms of religious enthusiasm. Yet some such revolution must be recognised, Kant argues, as requisite for virtue.

Now, Russell asks: what is virtue? Is it moral perfection? Or does it lie somewhere in the struggle to be good — the struggle undertaken, gladly, from the fallen condition of humankind? Here is another way of thinking of Russell's question: is virtue an ideal of perfection, or is it a post-revolution commitment to morality? It is not sufficient to say, as she does, that virtue is "the attitude in the struggle towards such perfection" — but I do think it is plausible to say, by Kant's lights, that virtue is a joyous attitude proper to the *post-revolution* struggle towards perfection. That qualifier, as I see it, is essential. Of course, the matter of being post-revolution is noumenal: it is not open to experience, in any sense. But crucially, the struggle, it seems to me, is there in any event. If Kant does indeed take the moral law to be constitutive of reason in its practical exercise, then anyone has some dim grasp of moral requirement just as soon as she comes into the use of her reason. This is part of what Kant

conceives under the heading of the original predisposition to the good in human nature. He also thinks that each will by default pursue her own happiness and tend to regard it as objectively or unconditionally *good*: and this is the fundamental evaluative confusion that, as I see it, Kant takes to be endemic to human nature. So anyone should have experienced some sort of struggle between the morality and self-interest. You will experience that conflict whether or not you have undergone a revolution in your soul, because the moral law is constitutive of reason in its practical exercise. So I think *struggle towards* perfection is probably not enough to capture the idea of virtue, unless you add that this is undertaken joyously in a post-revolution frame of mind. At any rate, I am inclined to think that virtue is an ideal by Kant's lights, something which we can only conceive, in pure thought, through the moral law. We can form images of ways in which this ideal might take shape concretely, in people who are situated and minded and temperamentally disposed in perhaps a wide range of ways. But those sketches are, perhaps unavoidably, always somewhat impressionistic.

One final remark about Kant's conception of virtue and ancient sources. I am somewhat uncomfortable with the idea that my account of Kantian virtue is Aristotelian, which is suggested by both Russell and Sussman. Although some great work has been done developing a Kantian account of virtue along Aristotelian lines — perhaps above all by Barbara Herman — as a historical point, I do not think there is much reason to think that Kant ever thought that long or hard about Aristotle, or even that Kant's fundamental systematic commitments align all that well with Aristotle's. As I hope to show in my next project over the next few years, there is ample evidence that Kant thought long and hard about Stoic ethics, particularly through Seneca, and that he endorses the psychological monism of the Stoics, as well as something along the lines of their cognitivism about virtue. Russell rightly points out that my view of virtue bears some resemblance to John McDowell's account in "Virtue and Reason" (1979): McDowell was undoubtedly, and in ways that I probably can't even quite appreciate, an influence. However, I am also inclined to think that McDowell may have mischaracterised his own view of virtue by calling it Aristotelian. By my lights, McDowell's view of virtue belongs more to the Socratic and Stoic monistic tradition — but that is not a claim I could elaborate here.[11]

[11] I haven't said anything here about Russell's interesting proposal that my argument that attention is required for "full blown experience" should imply, more or less by the same token, that "the capacity for *aesthetic* experience is also required for full-blown experience." I am inclined to think that the connection that Russell wants to draw here would require some kind of view about how Kant's conception of experience relates to his distinction between theoretical and practical cognition — which is a difficult topic, and not one I directly took up in the book. But I would like to see how the idea might be developed.

3.

David Sussman focuses on the final parts of my book, about moral virtue. He expresses my view quite accurately, and certainly more elegantly than I do. He also says that the case I make for the essentially cognitive nature of Kant's conception of moral virtue, while having "considerable truth," is nevertheless somewhat overstated. That may be: I was trying to reorient the discussion. So, in reply to Sussman, I want to elaborate on how I understand some of the issues that arise here about the cognitive nature of human moral psychology by Kant's lights. This will be a development of some of what already came up in response to Russell's comments, particularly on the topic of the passions.

Sussman gets the fundamental claim about my view of virtue exactly right, when he says that a virtuous person "is someone who has a vivid kind of appreciation of the central moral value (essentially, of just what a person is), and [is] able to see how this value is at play in the various features of the particular circumstances we are faced with." The way I see it, our grasp of the fact that "this is a person; not a thing" admits of degree — of clarity, of thickness, and of general practical determinacy. The virtuous person may not be able to express, propositionally, the knowledge she has; rather, as Sussman so nicely glosses my view, "the appreciation or attunement [at issue] is cognitive in the sense that having an ear for music might be, or a sense of humour, or a feel for the strength of a position in chess." While Sussman thinks that much of the picture is right, he also thinks that moral virtue "also has distinct volitional elements, at least insofar as it makes sense to still talk of any contrast between the cognitive and the volitional at all."

To elaborate, Sussman points to Kant's conception of virtue as a kind of strength — summed up in his Latin gloss *fortitudo moralis* (MS 6:380) — and he notes that Kant "tells us that virtue is a matter of having a moral resolve that is powerful enough to overcome whatever obstacles that inclination puts in our path." Later on, he acknowledges a point that I stress in the book: namely, that there is nothing *wrong* with the inclinations per se. Kant's exact words about virtue as strength are as follows: "strength is required, in a degree which we can assess only by the magnitude of the obstacles that the human being himself furnishes *through his inclinations*" (MS 6:405; my emphasis). (This passage, too, comes from the Stoic interlude in the Doctrine of Virtue.) Now, Kant does not here claim that inclinations themselves are the obstacles; and Sussman, I think, recognises this when he notes that the obstacles are not the inclinations but rather what we do "in response" to them. Kant's point, as I see it, is that the

obstacles to virtue are the *passions*; and I want to elaborate briefly on this point in the hopes of making more plausible my cognitivist conception of Kantian virtue.

Sussman says that "[t]his feature of virtue" — namely, the business of its strength to overcome obstacles furnished through the inclinations — "doesn't sound very cognitive." He goes on to give examples of clear-eyed weakness of will. Here are two of Sussman's examples: You know full well that eating pizza at this hour will give heartburn; but you do it anyway. You know full well that you should not commit adultery; but you do it anyway. I'm not entirely sure what to say about the pizza case. Doctrinally I am a committed psychological monist, so I would be inclined to explain that case as exhibiting shifting and unstable views about what is good or at least worth going for. But I want to focus on the cases where moral value is more immediately at issue, like the adultery case. Of course, inclinations seem to come to us unbidden; and some may be quite unwelcome. But there are *many ways* of "knowing that adultery is wrong." Suppose you find yourself with an inclination to flirt with someone who is not your spouse; suppose that you have this inclination *a lot*, so that thoughts, fantasies, perhaps even plans of adultery get involved as well. In what sense do you know that "adultery is wrong"? Let us ask what you do in response to these inclinations. Do you gratify them, even if not in deed, but in rumination? You might still know that adultery is something you should not do, but I do not think such knowledge is very robust. But perhaps you stop and look around you, look your spouse in the eye when you talk about what to have for dinner or who will pick up the kids: *Here is the person that I would hurt* by carrying on with this. Maybe then you think, in rich ways, of your past history, the commitments, the life shared. Intuitively, such a person might know that adultery is wrong in a somewhat different way than the first. Such knowledge might not be compatible with the gratification of this inclination, even just by ruminating on it.[12]

[12] I don't think the passage that Sussman quotes from the *Religion*, where Kant alludes to the Apostle Paul from Romans (7:14-15) entails that Kant endorses clear-eyed weakness of will, and in turn a fundamental separation between motivational and cognitive aspects of practical reason. But I grant that the passage is challenging for my reading. Interestingly, what Paul actually says right before Kant begins his quotation is "I do not understand what I do [ὃ γὰρ κατεργάζομαι οὐ γινώσκω]." Paul then says: "what I want, this I do not do, but what I hate this I do [οὐ γὰρ ὃ θέλω τοῦτο πράσσω, ἀλλ' ὃ μισῶ τοῦτο ποιῶ]" (Romans 7:15). Kant renders it, very loosely, thus: "Wollen habe ich wohl, aber das Vollbringen fehlt" (6:29). Kant then elaborates with the following explanation: the good that I will, and have incorporated into my maxim, *would be* "an irresistible incentive objectively or ideally (*in thesi*)" but subjectively — "*in hypothesi*" or under the human default condition — is "the weaker (in comparison with inclination) whenever the maxim is to be followed" (Rel 6:29). If I were a holy will, my grasp of this good would be perfect. But I am not a holy will, which is to say that I am subject to fundamental evaluative confusion that is endemic to human beings. I have other views of what would be good, which acquire considerable force through the accretions of my own assent. They can overpower my understanding, which is weak and indeterminate, of what really is good, of what I can recognise as moral requirement.

Passions (*Leidenschaften*) involve maxims, Kant says. They are an expression of rational mindedness, of taking a point of view on what matters, what is worth doing for what, and so on. Kant explains passion also as a kind of inclination, which he regularly glosses as "habitual desire":[13] they have the force of the habitual accretions of one's own assent, over and over again, to their principle. This is how Kant conceives of passions as self-wrought, but also in their way overpowering: they work like water that digs itself deeper and deeper into its bed, as Kant memorably puts it (Anth 7:252). They are also culpable, and not for some silly reason that it is in itself bad to be habitual about anything, especially desires. They are culpable because of the fundamental evaluative confusion that underlies them. The person who supposedly knows full well that adultery is wrong and yet finds herself carrying along a track leading to it, has something in her that sees this flirtation as good, as to be pursued. What those thoughts might involve will be particular is ways too various to try to elaborate from the armchair. Schematically, it involves assent to a practical principle, a maxim. This kind of direction of one's thought, I am suggesting, could be understood as a kind of confusion about what really matters, and what really is good. Put more simply, we should doubt that she really "knows full well" that adultery is wrong.

I'd like to end with the issue about self-conceit, and delusion about one's own moral worth, that I think both Russell and Sussman, each in different ways, have raised. As Sussman puts it, it is owing to our "radical evil" that "we endlessly rationalise and deceive ourselves about what we are doing and why." That's right; and as I see it, that sort of point is best understood with the kind of cognitivist, monist psychological framework that Kant develops from the Socratic and particularly Stoic tradition. Elaborating, Sussman says: "It is not just that our vision is often blurry; but rather, that we see most clearly when we are in the grip of an illusion." What he means by this, I think, is that mongering for reasons and justifications can just as well feed our self-conceit, and moral fantasies, as anything else. I quite agree. This is one reason why I wanted to show the roots of Kant's conception of virtue in a standard of healthy human understanding; it is also why I urged, with Iris Murdoch somewhat secretly in mind, that appreciating the reality of other persons — really, deeply, thickly getting that *here is a person*, and not a thing — is perhaps what most properly halts the smooth thrum of such self-justifying trains of thought. Of course, when a plausibly conscientious person realises that

[13] The German is *habituelle Begierde*: see e.g. Anth 7:251, 265; MS 6:212; Rel 6:28. For discussion of the point, see *Kant on Reflection and Virtue* (39-40).

her understanding of this fundamental point has been deficient here, now, with regard to this person, then she learns something about herself. Probably something hellish, too, if she is a plausibly conscientious person. Note that Kant uses the same language for his presentation of the three maxims in the *Anthropology* (7:228) as he uses in his presentation of the Delphic command in the Doctrine of Virtue (MS 6:445): both are said to be, effectively, the beginning of human wisdom. Together, Kant is perhaps making a point that moral self-knowledge is possible only through the outward engagement of our thought, where concrete others play a special role. Sussman is right to point out that Kant's account of virtue is Christian: at its centre lies an agapic conception of love, a welcoming being-pleased in the existence of others. This love is an ideal. I also think that, for Kant, it might be the best and truest way of *knowing* the value of persons.

Bibliography

Kant's works are cited by volume and page of the German Academy edition, with the exception of the *Critique of Pure Reason*, which is cited according to the pagination of the first and second editions, abbreviated A/B. (See note 5 for other title abbreviations used.) Quotations are drawn from the translations in the Cambridge Edition of the Works of Immanuel Kant, series editors Paul Guyer and Allen Wood.

Allison, Henry. *Kant's Theory of Taste*. Cambridge: Cambridge University Press, 2001.

Graver, Margaret. *Cicero on the Emotions: Tusculan Disputations 3 and 4*. Translation and commentary. Chicago: University of Chicago Press, 2002.

Longuenesse, Béatrice. *Kant and the Capacity to Judge*. Translated by Charles T. Wolfe. Princeton: Princeton University Press.

Lu-Adler, Huaping. *Kant and the Science of Logic*. New York: Oxford University Press, 2018.

McDowell, John. "Virtue and Reason." *The Monist* 62, no. 3 (1979): 331-350.

Merritt, Melissa. "Varieties of Reflection in Kant's Logic." *British Journal for the History of Philosophy* 23, no. 3: 478-501.