

The Implications of Kant's Empirical Psychology

Las implicaciones de la psicología empírica de Kant

PABLO MUCHNIK*

Emerson College, USA

The papers collected in this section were part of a session organized by the North American Kant Society at the Eastern Division Meeting of the American Philosophical Association in Washington, DC (January, 2016). The session, “New Perspectives in Kant’s Psychology,” marked a rare occurrence: the almost simultaneous publication in 2014 of two important new books on this topic, Corey Dyck’s *Kant and Rational Psychology* (Oxford University Press) and Patrick Frierson’s *Kant’s Empirical Psychology* (Cambridge University Press). At first glance, these books have little in common. While the bulk of Dyck’s project is devoted to interpreting Kant’s Paralogisms in light of 18th century German discussions of rational psychology, Frierson shows how Kant’s conception of human beings as *objects* of empirical investigation is essential to understanding his theory of action, cognition, and moral motivation. If one looks more closely, however, these projects share a fundamental assumption, namely, that in order to understand what motivates and shapes the development of Kant’s transcendental philosophy, one must grapple with the implications of his empirical psychology.

Dyck arrives at this view by placing the kind of psychology that comes under attack in the Paralogisms in its historical context. This line of inquiry has important philosophical consequences: it undermines a long-standing prejudice among Kant commentators, who usually assume that the rational psychology Kant disparages in the first *Critique* was based on knowledge of the soul obtained in complete independence from experience. So construed, rational psychology was a wholly “pure,” a priori intellectual exercise. A more careful examination of the historical record, however, shows that what is often identified as a paradigmatically Cartesian/Leibnizian brand of theorizing is not what Kant had in mind. Instead, Kant’s target in the Paralogisms was a hybrid kind of rationalist approach to the soul, according to which what we know of the soul is derived from our cognition of the self

* Associate Professor at Emerson College (USA). E-mail for contact: pablo_muchnik@emerson.edu .

through inner sense. Seen this way, empirical content pervaded rational psychology from start to finish: the perceptions and thoughts we encounter in inner sense not only serve to provide its first principles, but also to confirm its results. The pervasiveness of empirical content, Dyck argues, complicates the traditional picture we have of the Paralogisms: to debunk the pretensions of rational psychology, it is not enough for Kant to expose its erroneous metaphysical claims; he must, in addition, account for the grounds for its reliance on an empirical intuition of the soul in the first place. This reliance, Kant gradually came to realize, is the fundamental error at the heart of this enterprise and explains the rational psychologist's temptation to confuse "the unity in the synthesis of thoughts for a perceived unity in the subject of these thoughts" (KrV A 402). The march to clearing this unfortunate confusion regarding the limits of our cognition of the soul, therefore, should be understood as Kant's coming to terms with the radical consequences of his doctrine of the unity of apperception. In the *Critique*, Kant calls this unity the "highest point" in the employment of our understanding (KrV B134). To be able to reach it, Kant had to leave behind the "valley" of empirical psychology he had traversed in the company of his fellow Wolffians for many years.

Frierson's project, on the other hand, moves in the opposite direction –downward, so to speak, from the diaphanous "summit" of transcendental philosophy to the thick causal nexus of inner sense. For what is distinctive of Kant's empirical psychology is its commitment to treat the self *only* as a phenomenon: human thoughts, actions, and motives are seen just as any other natural event, subject to a deterministic *scientific* explanation. So construed, the object of empirical psychological study is "the I as it appears to itself, not as it is in itself. Empirical psychology does not study the transcendently free subject or transcendental 'I' of apperception, but the self of which one is aware in inner sense" (Frierson 2014, p. 49). The goal of this inquiry is to treat "the mind/soul as a substance with various causal powers, each of which is governed by a distinctive (set of) causal law(s)" (ibid.), which Kant discovers by tracing mental states back to a handful of irreducible natural predispositions. This process, Frierson explains, starts with introspection, by which the subject observes itself as an object, i.e., "from without," adopting a third-personal perspective on its own inner experiences. On the basis of these observations, the inquirer proceeds to make anthropological generalizations by comparing how things stand with other human beings. There is no room in this empirical account to make evaluative judgments of inner states.¹ This is the job of transcendental philosophy, which investigates the same faculties, but with an eye to finding the a priori conditions for the possibility of their engagement in cognitive, desiderative, and aesthetic forms of experience. Instead of providing causal explanations, transcendental philosophy is concerned with the normative space of reasons and their evaluation. This line of investigation engages the mind in pure, spontaneous, a priori self-examination. The perspective here is strictly first-personal ("from-within"). A central goal of Frierson's book is to examine the consequences that these two styles of philosophizing about the self, from

¹ As Frierson puts it in his paper: "The question whether, say, anger is a 'good' *cause* is misguided; it either is the cause or it is not" (p. 359).

within and from without, have for Kant's theory of action, and in particular, for how to understand the moral law, the feeling of respect, responsibility, weakness of will, and cognitive error.

The intention of the APA Panel was neither to follow the traditional Author-Meets-Critics format nor to present a summary of the books for those who did not have a chance to read them. The idea, instead, was to offer an opportunity for the authors to elaborate and defend controversial aspects of their interpretations. For that purpose, we invited Patricia Kitcher and Jeanine Grenberg, scholars who have made major contributions to this area of study, to reflect on Dyck and Frierson's views, respectively. In what follows, I will discuss only a few aspects of their exchange.

1. The Misadventures of Inner Sense

Dyck's thesis is that the fluid boundaries of inner sense in Kant's pre-critical period, and, specifically, throughout the so-called "silent decade," account for the shifting fate of rational psychology in the Kantian corpus. The story Dyck tells is one of an initial fall, redemption, and final demise – a drama in three acts. According to this story, the extravagant, quasi-mystical deliverances of inner sense in *Dreams of a Spirit-Seer* (1766) are reined in in the *Inaugural Dissertation* (1771), where Kant restricts them to the phenomenal world (act 1). Those restrictions are lifted in the Anthropology Lectures of the early 1770s – a change which led Kant to embrace the prospects of a hybrid rational psychology (built on an empirical cognition of the self) in the Metaphysic Lectures at the end of the decade (act 2). Finally, as a consequence of the doctrine of apperception, which Kant developed a few years earlier (circa 1775, in the so-called Duisberg Nachlass), he came to recognize the unsuitability of empirical cognition for grounding any rational knowledge of the soul. This realization put an end to the aspirations of rational psychology to gain cognition of the soul's substantiality, unity, and personality. The need to offer a principled account of this failure, Dyck concludes, is what motivates Kant's discovery of the Paralogisms (act 3).

Framing things this way has several heuristic advantages over traditional interpretations. First and foremost, it provides a gradualist account of Kant's philosophical development, according to which the Paralogisms are neither an about face nor a sudden rejection of the views Kant espouses in ML1, but the result of a long and arduous process of maturation. Second, it saves Kant from the embarrassment of having held inconsistent views about the self after developing the doctrine of inner sense in the *Inaugural Dissertation* – such inconsistency would be unavoidable if, as Wolfgang Carl assumes, Kant had identified the self or soul with the *res cogitans*, i.e., a completely different substance to

which we have a pure intellectual access.² Finally, placing rational psychology within the Wolffian context to which it belongs, shows Kant's debt to a tradition his Critical philosophy renders eventually obsolete, and is hence a good measure of Kant's originality in working out the tensions within a view he himself had endorsed. Transcendental philosophy, interpreted in this way, is the result of a therapy Kant had to apply first upon himself.

Kitcher finds much to recommend in Dyck's narrative. She objects, however, to "two details" about the dilemma Kant was facing in the first and second developmental stages; building on this criticism, she develops an alternative account of why Kant changed his mind with respect to the scope of inner sense. As Kitcher reads it, the extravagances of inner sense associated with Swedenborg's metaphysics do not entail that Kant rejected a more standard Lockean variety of internal sense –a possibility which lessens the importance of the role of the *Inaugural Dissertation* in Dyck's story. Similarly, Kitcher questions Dyck's readiness to believe that the tensions regarding the purview of inner sense led Kant to resort to some kind of intellectual intuition to resolve them. She proposes, instead, to apply a "causal theory of representation/cognition," according to which "inner sense would provide cognition of the actions of the mind that coordinate sensations, through providing intuitions of *images* that invariably exhibit temporal relations" (p. 349). On this reading, cognizing temporal arrays allows one to know the mental actions that produced them and gave them the form they have, and Dyck's appeal to intellectual intuition is thereby rendered idle.

Kitcher draws on these objections to offer an alternative diagnosis of the error of rationalist psychologists. For, she reckons, if what is cognized through inner sense is neither merely sensory nor merely empirical, the contributions of empirical psychology will be poorly suited to provide cognition of an identical cognitive subject through time: "Even if the cognition of a thought or perception entitles the subject to infer that the state must inhere in some substance, there can be no inference to a common subject for different states" (p. 350). A similar problem occurs if, instead, the deliverances of inner sense included not merely sensory materials but also the form of time: the changing perceptions do not warrant an inference to an underlying subject at their basis. The combination of these problems, Kitcher concludes, explains why Kant was compelled to look for an alternative solution. Instead of trying to find the continuity of the subject through inner sense, a project that proves itself to be futile, Kant changed tacks and sought it now in the requirements of thinking –in the act of combining representations in *one* consciousness. Once Kant understood the full force of the unity of apperception, the old dream of rationalist psychology of finding the I of the "I think" via the empirical content of inner sense could no longer be pursued. According to Kitcher's analysis, therefore, the key to the Paralogisms is not to be found in the Metaphysics Lectures of the late 1770s as Dyck

² Wolfgang Carl, *Der schweigende Kant: Die Entwürfe zu einer Deduktion der Kategorien von 1781*, Abhandlungen der Akademie der Wissenschaften zu Göttingen, Philologisch-Historische Klasse, Dritte Folge Series 182. Göttingen: Vandenhoeck and Ruprecht, 1989, 91-2.

supposes, but is already implicit in the incapacity of inner sense to yield knowledge of a subject that underlies and withstands temporal fluctuations.

No matter which interpretation the reader might find most compelling, these authors agree on the fundamental issue: in the scathing critique of rational psychology that Kant develops in the Paralogisms, they see signs of a long inner philosophical struggle – the overcoming of an illusion less to do with the purity of a thinking substance than with the doomed project of extracting knowledge of the soul from the impurities of inner sense.

2. The Phenomenology of Respect and the Opacity of Reason

Frierson's paper responds to the challenge that Grenberg's most recent book poses to his general interpretative framework –and particularly, to his neat division of labor between empirical psychology and transcendental critique. In her book, *Kant's Defense of Common Moral Experience: A Phenomenological Account* (Cambridge University Press, 2013), Grenberg provides a sustained argument for the importance that the first-personal encounter with moral demands has for our understanding of Kant's practical philosophy. Central to her account is the third chapter of the *Critique of Practical Reason*, "On the Incentives of Pure Practical Reason," where Kant discusses the feeling of respect and its role in moral motivation.

In *Kant's Empirical Psychology*, Frierson gave an empirical-psychological reading of this chapter. He supported his case by appealing to Paton's famous distinction between two different perspectives with respect to moral motivation, according to which "from one point of view, [respect] is the *cause* of our actions, but from another the moral law is its *ground*" (Paton 1947, 67). On the basis of this perspectival account, Frierson carved out a space for two parallel lines of inquiry with respect to the *same* action. While empirical psychology investigates what sorts of feelings the moral law produces when it *causes* moral action, transcendental philosophy concerns itself with the intelligible "ground from which the moral law provides an incentive" (5:73). Read this way, the purpose of the Incentives chapter is to describe "what moral motivation actually looks like, when it's effective in a particular person (in a particular case)" (p. 357). Whether the moral law (a purely intellectual ground) causes one's action or not "depends upon the underlying structure of one's instincts, inclinations, and character, which structure is in turn determined by various natural predispositions and causal influences on our development (such as education)" (p. 356). The feeling of respect, therefore, is interpreted in light of Kant's general theory of action, according to which feelings necessarily mediate between a particular cognition and the desire and action this cognition produces. In Frierson's view, Kant's theory of action ranges over moral and prudential behavior alike, and hence applies to cognitions as different as "say, the smell of a mango or the consciousness of the moral law" (p. 356).

One consequence of the empirical psychological treatment of respect is that it puts it on the same footing as any other feeling. Such approach, Frierson admits, faces a “pressing textual problem” (p. 358). Kant puts it thus:

«What is essential to any moral worth of actions is *that the moral law determines the will immediately*. If the determination of the will takes place *conformably* with the moral law but only by means of a feeling, of whatever kind, that has to be presupposed in order for the law to become a sufficient determining ground of the will (...), then the action will contain *legality* indeed but not *morality*». (5: 71)

The mediation of feeling, without which empirical psychology could not explain human behavior, seems to strip moral action of its characteristic “moral worth.” At most, Kant seems to be saying, empirical psychology could account for the legality of our conduct, but is doomed to misrepresent its morality, since the interjection of feeling turns duty into a system of hypothetical imperatives. Furthermore, since Kant believes that “it is impossible to see a priori which representation will be accompanied with *pleasure* and which with *displeasure*” (KpV 5:58), it is hard to understand how empirical psychology can ever capture the most distinctive feature of Kantian respect, “the only [feeling] we can cognize completely a priori and the necessity of which we can have insight into” (KpV 5:73).

Frierson's general strategy to overcome the first difficulty is to distinguish between an agent's empirical character and the intelligible character we posit as its ground: “The ‘empirical character’ is what empirical psychology investigates through the systematic employment of a theoretical (‘third personal,’ even if introspective) standpoint on human action, while practical philosophy... emerges from the systematization of the practical standpoint, in which *moral* choice can never be determined by –that is, based on– feeling” (p. 358). The immediacy of moral motivation can be secured in this way, since it refers to the noumenal happenings at the level of our intelligible character, not to the psychological levers we use to explain moral action as an empirical event.

Frierson's solution, however, does not address the question of the apriority of the feeling of respect.³ More importantly, it introduces a new set of difficulties. First, it makes the study of morality inaccessible to us: “For, how a law can be of itself and immediately a determining ground of the will (though this is essential in all morality) is for human reason an insoluble problem and identical with that of how a free will is possible” (KpV 5: 72). We run here into the “extreme boundary of all practical philosophy” (G 4:455), and, as Dieter Henrich has argued, Kant eventually abandons the project of finding a purely rational justification of morality and embraces the doctrine of the “*Faktum der Vernunft*.”⁴ Even if we were to accept the Kantian view that the moral law presents itself as a “fact” in the process of moral deliberation, this fact is nonetheless made known, to finite creatures

³ Filling this gap occupies much of “Towards a transcendental Critique of Feeling,” Frierson's reply to Grenberg. I will leave the evaluation of his reformulation for another time.

⁴ Dieter Henrich, “The Concept of Moral Insight and Kant's Doctrine of the Fact of Reason,” in *The Unity of Reason: Essays on Kant's Practical Philosophy* (Richard Velkley ed.). Cambridge, MA: Harvard University Press, 1994, pp. 55-88.

like us, through the effects morality exerts on our sensibility. Frierson's transcendental philosophy, confined as it is to the sheer space of reasons and the conditions for the possibility of their efficacy, hobbles without a moral aesthetic to support it. Second, as Grenberg points out in her comments, the third-personal perspective of empirical psychology is ill suited to tackle this aesthetic dimension of morality, since it seems to leave "no meaningful room for the exploration of feeling first-personally" (p. 374).⁵ To the extent that, unlike sensations, feelings are intrinsically subjective, to study them "from without" neglects their most important aspect: cognitions here are not directed to an object but refer "merely to the subject" (KU 5:207). The neglect of this first personal dimension of moral feeling, Grenberg argues, is not accidental: it "follows almost tautologically from [Frierson's] initial cutting up of the transcendental pie" (p. 374). By restricting the first personal approach to disembodied reasons, Frierson faces the dilemma of either having to detach the agent from her own lived experience or force her to see her experience theoretically, from the spectator's point of view. Grenberg takes this predicament as a symptom of an intellectualist reading of Kantian ethics. She calls it "Patrick's 'Paton Problem'" or, alternatively, "his 'Korsgaard Krankheit'" – "[t]he problem, or disease, is this: "in a misguided effort to protect pure reason from the undue influence of sensibility, the genuine role for feeling in Kant's practical transcendental philosophy is rejected" (p. 375).

The two difficulties are, of course, intertwined: since Kant believes that it is impossible to show a priori "the ground from which the moral law in itself supplies an incentive" (KpV 5:72), he comes to see that the only way for us to understand how pure reason is itself practical is by turning the gaze inwardly, toward the subject, and examining the effects the moral law must have "in the mind insofar as it is an incentive" (ibid.). It is by attending to the emotional traces which the incorporation of the moral incentive *must* produce in a finite rationality like ours that we first discover the authority of the moral law, the categorical nature of moral demands. These demands are made known by *being felt*: they must be accessed via sensibility, through a glass darkly, since human beings lack intellectual intuition and hence cannot possibly grasp how the moral law could determine their will directly. Although, as Grenberg indicates, the feeling of respect does not play any "justificatory role in grounding moral reasons" (this would turn Kant into a moral sentimentalist), respect does play a crucial role nonetheless – an "epistemically *enabling* role" (p. 376). As she puts it: although the fact that I feel respect "is not the reason I am morally obligated [...], 'knowing' my moral obligations authentically requires that I both *experience* and *attend* to my experience of the feeling of respect. That is: respect plays an epistemically *enabling* role in *getting* me to those reasons that justify moral obligations" (p. 376).

⁵ Although he concedes that Grenberg's "is a reasonable view to come away from my paper with" (p. 383), Frierson considers the charge a misunderstanding. He appeals to a footnote (number 1) in his original essay and a recent paper to make his case (See Patrick Frierson, "Affective Normativity," in *Kant on Emotion and Value*, pp. 166-90).

Neither empirical psychology nor transcendental philosophy as Frierson conceived them in his book can do justice to this subjective dimension of our moral experience: while the former examines feelings “from without,” the latter is circumscribed to the space of reasons. Whereas the transcendental approach to moral feeling misses its aesthetic dimension, the third-personal approach of empirical psychology is unable to capture the subjective side of moral motivation.

We need, it would seem, another way –a *via tertia*, a perspective that avoids the shortcomings of Paton's dichotomy. Frierson calls it “phenomenology”: “As in transcendental philosophy, the focus [of phenomenology] is on the self *as subject*. But as in empirical psychology, we are *receptive* to what is *given*, rather than actively taking something as a reason” (p.). Neither pure activity (though itself containing a moment of agency) nor pure passivity (though unmistakably sensible), phenomenology is, in Grenberg's apt formulation, “an activity of being receptive to what is present” (Grenberg 2014, 185). As such, this activity straddles the gulf between our empirical and intelligible characters, and thus supersedes Paton's old dualisms, for it is predicated on a kind of moral schematism without which the moral law would be empty and our moral feelings would be blind.⁶ Once Kant recognized the implications of the inscrutability of human freedom, phenomenological self-knowledge becomes, as he puts in the *Metaphysics of Morals*, “the first command of all duties to oneself” and “the beginning of all human wisdom” (MS 6: 441).

Complying with this command is extremely difficult, since “only the descent into the hell of self-cognition can pave the way to godliness” (MS 4:441). As Kant explains in the Incentives chapter, phenomenological self-cognition is hellish because it requires the infringement of our self-love and the humiliation of our self-conceit (KpV 5:73). Only after our painful submission to the demands of morality can the pleasure of self-approbation supervene (KpV 5:75). We are tempted here to find a shortcut, a way to sidestep the strenuous discipline that our hedonistic tendencies must endure. We are enticed to confuse “what one *does*” with “what one *feels*” (KpV 5:116), “an illusion that even the most practiced cannot altogether avoid” (ibid.). Unlike other transcendental illusions that afflict speculative reason and are the result of a non-culpable error on the part of the cognitive subject, Kant believes that the confusion in question here is a sign of a perverse moral disposition by which we trick ourselves into mistaking the pleasure of self-approbation (the last moment in the experience of moral self-affectation) with empirically generated pleasures (the passive effect that cognizing external object has on our sensibility). While the former is the subjective counterpart of the moral law, the latter can at most secure the veneer of morality –the *legality* of our actions. Kant calls this confusion the “error of subreption [*Fehler des Erschleichens*]” (KpV 5:116), and associates it with a “vice” (“*vitium subreptionis*”) because such an error is indicative of a self-inflicted tendency to invert the order of priorities between the ethical incentives affecting the human will. This tendency is no other than “the propensity to evil in human nature” (R 6:36)

⁶ I develop this view in “The Heart as Locus of Moral Struggle in *Religion*,” in *Kant on Emotion and Value* (Alix–Cohen ed.). New York: Palgrave MacMillan, 2014, 224-44.

which Kant discusses more fully in *Religion I*. It afflicts even the best of us, for, as Kant already alluded to in the *Groundwork*, it expresses a “natural dialectic” in human practical reason –“a propensity to rationalize (*vernunfteln*) against those strict laws of duty and to cast doubt upon their validity, or at least upon their purity and strictness, and, where possible, to make them better suited to our wishes and inclinations” (G 4:405).⁷

The disagreement between Frierson and Grenberg regarding the role of phenomenology I believe hangs, in last instance, on the importance they attribute to moral self-examination in Kant’s practical philosophy. As Frierson sees it, the fact of reason in no significant way depends on “the ‘feel’ of it”: “that the Fact of moral obligation shows up in a particular way –via respect– is not essential to Kant’s argument” (p. 369). If we were more like Gods, and what we cognized as objectively necessary were also subjectively necessary (G: 412), “Kant’s argument from the fact of moral reasons to the possibility of acting in accordance with them would still work” (p. 369). Grenberg, on the other hand, insists that the path to “godliness (*Vergotterung*)” will never makes us Gods, “for no matter how far back we direct our attention to our moral state, we find that this state is no longer *res integra* and that we must rather start by dislodging from its possession the evil which has already taken up position there (as it could not have done, however, if it had not been incorporated by us into our maxims)” (R 6:58 n.). In her mind, Frierson’s conception of transcendental philosophy is built upon the questionable assumption that human practical reason could be made whole. Only constant counteraction and attentive self-examination could keep “the natural dialectic” in check: since human practical reason is not transparent to itself, the task of morality must begin by clearing the dust we have thrown on our own eyes (R 6:38). This is why Kant insists that we must “know [our own] 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*” (MS 6: 441).

If Grenberg’s reading is correct, transcendental philosophy has phenomenology as its condition for the possibility. If we side with Frierson, the Incentives chapter is an afterthought, an important though dispensable addendum for the transcendental project Kant undertakes in the second *Critique*. As I have tried to show here, the difference in their respective interpretations boils down to how radical they think is evil’s radicalism.

Works Cited

Carl, Wolfgang. *Der schweigende Kant: Die Entwürfe zu einer Deduktion der Kategorien von 1781*, Abhandlungen der Akademie der Wissenschaften zu Göttingen, Philologisch-Historische Klasse, Dritte Folge Series 182. Göttingen: Vandenhoeck and Ruprecht, 1989, 91-2.

Dyck, Corey. *Kant and Rational Psychology*. Oxford, UK: Oxford University Press, 2014.

⁷ See my *Kant’s Theory of Evil: An Essay on the Dangers of Self-Love and the Apriority of History*. Lathan, Maryland: Lexington Books, 2009, particularly chapter 2, pp. 43-84.

Frierson, Patrick. *Kant's Empirical Psychology*. Cambridge, UK: Cambridge University Press, 2014.

———. "Affective Normativity." *Kant on Emotion and Value*. Alix Cohen (ed.). UK: Palgrave MacMillan, 2014. 166-90.

Grenberg, Jeanine. *Kant's Defense of Common Moral Experience: A Phenomenological Account*. New York: Cambridge University Press, 2013.

Henrich, Dieter. "The Concept of Moral Insight and Kant's Doctrine of the Fact of Reason" *The Unity of Reason: Essays on Kant's Philosophy*, Richard L. Velkley (ed.). Cambridge, MA: Harvard UP, 1994. 55-88. Print.

Muchnik, Pablo. *Kant's Theory of Evil: An Essay on the Dangers of Self-love and the Apriority of History*. Lanham: Lexington, 2009. 43-84. Print.

———. "The Heart as Locus of Moral Struggle in *Religion*." *Kant on Emotion and Value*. Alix Cohen (ed.). UK: Palgrave MacMillan, 2014. 224-244.

Paton, H.J. *The Categorical Imperative: A Study in Kant's Moral Psychology*. London: Hutchinson and Co., 1947.

