

## *Reason's Self-Knowledge and Kant's Critical Methodology*

[The fashion of professing indifference to metaphysics] is a demand for reason to take on anew that most difficult of all her tasks, namely that of self-knowledge, and to institute a tribunal that will secure her in all her rightful claims, while being able to dispatch all her groundless pretensions, not through despotic decrees but through her eternal and unchangeable laws; and this [tribunal] is none other than the **critique of pure reason** itself.

Axi–xii, original emphasis

Kant's *Critique of Pure Reason* is fundamentally an exercise in self-knowledge. Yet, the self-knowledge it seeks is of a peculiar sort. Kant ascribes this self-knowledge to reason itself, as a cognitive faculty. He figures our capacity for reason as an epistemic agent in its own right. It is far from clear how this sort of facultative self-knowledge relates to the more familiar kind that an individual person might have about her own thoughts, intentions, personal history, and so on. What does it mean for a cognitive faculty to enjoy self-knowledge? What is it that reason knows in knowing itself? And how does it come by such knowledge?

The aim of this chapter is to unpack Kant's idea of critique as "reason's self-knowledge". I will argue that one of the distinguishing features of such knowledge is that it demands a special philosophical methodology. The status of critique as the self-knowledge of reason turns not so much on *what* is known but on *how* it is known. What makes critique *facultative* self-knowledge is not merely that reason is the subject matter of the inquiry as well as the central means by which it is pursued. It is because the inquiry relies on a special sort of cognitive ground: namely, apperceptive insight into reason's constitutive norms. That is what makes critique the sort of self-knowledge that it makes sense to ascribe to a cognitive *faculty* rather than an individual person. For it exploits the essentially self-conscious character of reason in order to justify claims about the constitutive essence (i.e. the form) of that very faculty.

In this respect, Kant's conception of critique is closely analogous to his conception of pure general logic (PGL). Kant claims that both sciences embody reason's self-knowledge of its own form. Yet, while PGL concerns the form of reason in *all* its manifestations – practical as well as theoretical, empirical as well as pure – the *Critique of Pure Reason* is specifically concerned with the form of reason as it manifests itself in a priori theoretical cognition, just as the *Critique of Practical Reason* is concerned with the form of reason in its practical use, i.e. insofar as it is capable of determining the will. Critique is thus a formal science of the material use of reason – its use in cognizing objects a priori, either practically or theoretically. PGL, by contrast, is a formal science of reason *überhaupt* – reason “in general” or merely “as such” – without taking its various material uses into consideration.

A central challenge for this reading is to explain how an inquiry into the *material* use of reason can nevertheless count as *formal*. The form of reason, I contend, is that through which an activity is constituted as rational: The form of reason is its *essence*. In order for knowledge of this essence to count as *self-knowledge* in the relevant sense, the form of reason must not only be its topic but also its source. Critique is thus a rational investigation of the constitutive principles that make an investigation rational. This is why Kant associates the *formality* of such knowledge – in PGL and in critique – with the idea that the principles of the relevant science can be known exhaustively and with certainty: namely, because the inquiry is guided by the very principles it seeks to articulate, so that the task is just to make this implicit guidance explicit. Because critique and PGL manifest the very form they seek to characterize, rational self-reflection alone provides a sufficient cognitive ground for each science.

Making sense of the formality of critique is a significant step toward understanding it as the sort of self-knowledge that belongs to a faculty and not to individuals, as such. For knowledge of these principles is grounded in the nature of reason, as a self-reflective capacity, and not in any special endowment or experience. Of course, only individual persons are cognitive agents. So all knowledge, including knowledge of logic and of critical philosophy, is possessed by individual persons alone. But it is not *as* individual persons that we know such things. Rather, it is in virtue of being rational that we can enjoy reason's self-knowledge.

The chapters that follow will explore how specific doctrines of Kant's critical philosophy – in particular, his characterizations of human intuition – fit into this self-knowledge of reason. The peculiarly reflective, self-conscious methodology that Kant identifies as central to PGL and critique is not restricted to uncovering features of our “higher” cognitive faculties.

I will argue that Kant employs the very same methodology in articulating his critical theory of sensible intuition. Paradoxical as it may seem, part of what reason knows, in knowing *itself* (its own form), is the general character of a capacity *distinct* from it – a capacity for receptive intuition, on which reason depends in order to attain its constitutive cognitive ends.

### 1.1 Unfolding A Priori the Mere Concept of a Faculty of Knowledge

When the *Critique* first appeared, it encountered more incomprehension than opposition. This, at least, was Kant's view. In the appendix to the *Prolegomena*, Kant complains about one anonymous reviewer who "seems not at all to see what was really at stake in the investigation with which I (felicitously or infelicitously) occupied myself" (4:373.7–8; cf. 376.19–21, 261.9–11). The reviewer seizes on Kant's idealism but does not appreciate that this "so-called (actually critical) idealism is of a quite peculiar sort" (375.15–16). In particular, the reviewer fails to grasp the overarching problem for which Kant's critical idealism aims to provide a solution: namely, how we can enjoy synthetic knowledge a priori (377.19–31). Instead, the reviewer takes Kant to be engaged in just the sort of traditional metaphysics that the *Critique* calls into question. So what the reviewer is missing is the very idea of a critique of pure reason. His local misunderstandings stem from a general blindness to the special character of the critical enterprise, as reason's self-interrogation of its own capacity for a priori knowledge: "The reviewer thus understood nothing of my text and perhaps also nothing of the spirit and essence of metaphysics itself" (377.31–33).

Distressed by Kant's public castigation of his review, Christian Garve wrote to Kant, revealing himself as its author and conceding Kant's main objection: "I [Garve] believe that I rightly grasped the sense of most passages considered singly; I am not so sure that I had a proper overview of the whole."<sup>1</sup> Kant was mollified. His conciliatory reply to Garve is especially valuable, since it attempts to enlighten Garve about the central point the review missed, concerning the distinctive character of the critical philosophy and its method:

Please be so good as to cast another quick glance upon the whole and note that what I develop in the Critique is by no means metaphysics but an

<sup>1</sup> Garve to Kant, 13 July 1783, 10:330.2–4. Garve shunts responsibility for the review onto the editors of the *Göttingische gelehrte Anzeigen*, claiming (inaccurately) that they "mutilated" his original text (332.31; cf. 330.29). On the Garve controversy, see Kuehn (2001, 250–252, 267–268).

entirely new and hitherto unattempted science, namely, the critique of a reason that judges a priori. Others have admittedly touched on this faculty, such as Locke and also Leibniz, but always muddled together with other cognitive powers[.] [Y]et no one has even entertained the thought that this [faculty] may be an object of a formal and necessary and, indeed, quite extensive science, which (without departing from this restriction merely to assess the sole faculty of pure knowledge [*des alleinigen reinen Erkenntnisvermögens*]) demands such a multiplicity of subdivisions and simultaneously (which is marvelous) can derive from its [the cognitive faculty's] nature all objects to which it extends and can enumerate them [and] prove their completeness through their interconnection in a whole faculty of cognition. This absolutely no other science is capable of doing, namely, unfolding a priori out of the mere concept of a faculty of knowledge (if it is determined precisely) all objects and everything that one can know of them [..]. Logic, which would most closely resemble this science, is in this respect infinitely beneath it. For it [logic] admittedly pertains to every use of the understanding whatsoever, but cannot at all indicate to which objects and how far intellectual cognition [*Verstandeserkenntnis*] will extend[.] (Kant to Garve, 7 August 1783; 10:340.2–25)

There is a lot to unpack here. Kant claims (1) that critique concerns “a [faculty of] reason that judges a priori”; (2) that it is “a formal and necessary science” of this faculty; (3) that this science analyzes (“unfold[s] a priori”) “the mere concept of a faculty of knowledge (if it is determined precisely)”; and (4) that this analysis reveals “all objects and everything that one can know of them”. Kant aims to elucidate these claims by comparing critique to logic. On the one hand, “[l]ogic [..] would most closely resemble this science” – presumably because both are “formal and necessary”. On the other hand, what is “marvelous” about critique is that it “can derive from the nature [of the cognitive faculty] all objects to which it extends and can enumerate [and] prove their completeness”; whereas logic “cannot at all indicate to which objects and how far intellectual cognition will extend”, since logic “pertains to every use of the understanding whatsoever”.

I think this discussion of critique – and especially the comparison with logic – is a helpful guide to Kant’s difficult but crucial conception of reason’s self-knowledge. I will first outline the sense in which logic is “formal and necessary” (Section 1.2). With this in place, we can then unpack Kant’s comparison of critique to logic (Section 1.3). This will clarify in what sense logic and critique embody reason’s self-knowledge. It will also position us to appreciate what is “marvelous” about critique (Section 1.4): namely, its ability to “[unfold] a priori from the mere

concept of a cognitive faculty (if it is determined precisely) all objects and everything that one can know [*wissen*] of them”.

## 1.2 Pure General Logic as a *Selbsterkenntnis* of Reason

Kant's discussion of “logic” in his letter to Garve clearly refers to what he elsewhere calls “pure general logic” (PGL). For he tells Garve that logic “pertains to every use of the understanding whatsoever” (10:340.23–24). This recalls his description of general logic in the *Critique* and elsewhere as concerned with “the absolutely necessary rules of thinking without which no use of the understanding takes place at all”.<sup>2</sup> By contrast, a “special” or “particular” (“*besondere*”) logic concerns only “the rules for thinking correctly about a certain kind of object” (A52/B76.1–2). So the logic Kant describes to Garve must be general rather than special: It concerns all thinking, whatever its object may be.

In addition to being general, the logic Kant mentions to Garve must be pure. “General logic”, the *Critique* tells us, “is either pure or applied”:

A **general** but **pure logic** has to do with pure principles [*lauter Prinzipien*] a priori and is a **canon of the understanding** and of reason, but only with respect to what is formal in their use[.] [. . .] A **general** logic is called **applied** when it is directed at the rules for the use of the understanding and the subjective empirical conditions that psychology teaches. It thus has empirical principles, even though it is general insofar as it pertains to the use of the understanding without distinguishing between its objects. (A53/B77.25–03)

An applied logic rests on empirical grounds and so is not a necessary science. Pure general logic, by contrast, promulgates its principles a priori and thus articulates necessary truths about “what is formal” in our use of the understanding and reason. Kant surely has such a pure logic in mind when he tells Garve that “logic [. . .] most closely resembles” critique (10:340.21–22), which he has just called a “formal and necessary [. . .] science” (340.9–10). So it is to PGL that Kant finds it instructive to compare the critical philosophy in order to bring out the distinctive status that Garve's review failed to register.

### 1.2.1 PGL as *Selbsterkenntnis* with Respect to Subject Matter

The generality of PGL consists in the fact that it “abstracts from all content of intellectual cognition [*Verstandeserkenntnis*] and the differentness

<sup>2</sup> A52/B76.33–35; cf. A131/B170.12–21; R1620 (1780s) 16:40.17–27; L-Jäsche 9:12.19–24, 13.27–14.1; L-Pö 24:503.15–22; L-Wien 24:790.22–28; L-DW 24:694.14–16.

[*Verschiedenheit*] of its objects, and has to do with nothing but the mere form of thinking”.<sup>3</sup> But this does not mean that PGL is devoid of content or that it lacks a proper subject matter.<sup>4</sup> Pure general logic is not vacuous; it is formal. Its object is the form of thinking *überhaupt*. Pure general logic discloses “the absolutely [*schlechthin*] necessary rules of thinking, without which utterly no use of the understanding takes place” (A52/B76.33–35). These rules are formal in the sense that they concern the *essence* of thinking, the characteristics that constitute mental activity *as thinking* (rather than *as feeling pain*, say).<sup>5</sup> Such characteristics pertain to thought merely as an act of intellect,<sup>6</sup> regardless of its content, the sorts of objects it concerns, or how it relates to those objects. Because PGL investigates the form, the essence, of acts of thinking as such, it can be understood as an inquiry into the faculty of thinking itself:

[PGL] is a rational science not merely [1.] with respect to form but [2.] with respect to matter, since [1.] its rules are not taken from experience and since [2.] it also has reason for its object. Logic is hence a self-cognition [*Selbsterkenntniß*] of the understanding and of reason, but not with respect to their capacities in regard to objects, but merely with respect to form.<sup>7</sup>

All correct exercises of the intellect, by definition, obey “the absolutely necessary rules of thinking”: rationality is their *form*.<sup>8</sup> But not all correct

<sup>3</sup> A54/B78; cf. Bvii.9–ix.21, A299/B355.25–28; *Groundwork* 4:387.8–12.

<sup>4</sup> On the proper subject matter of PGL as a science, see Lu-Adler (2018a, 154–161). I am convinced by Tolley (2012) that transcendental logic (TL) is not a special logic, i.e. that it is not distinguished from PGL by having a narrower object domain. Both PGL and TL share the same object domain; they differ in that TL takes into account the relation of thinking to the objects in that domain, specifying that this relation must be possible a priori, whereas PGL abstracts from all relation of thought to its object, whether this relation can be established a priori or only on the basis of experience (cf. Lu-Adler 2018a, 157–160). Thus, PGL is more *abstract* than TL, but need not be more *general*, since the determinations from which it abstracts do not narrow the object domain. I will formulate my account in accordance with this view, though it is also possible to do so *mutatis mutandis* on the assumption that TL is a special logic concerned with thinking about the sort of objects that can be given in pure sensible intuition (cf. MacFarlane 2002, 42n.35). For instance, Merritt treats TL as a special logic (2018, 22–28) but develops an account of reflection that my interpretation echoes in numerous points (cf. Merritt 2018, chs. 1–3).

<sup>5</sup> For this use of “formal” as meaning *essential*, see *Tone* 8:404.12–21; M-Mr 29:826.2–7, 847.9; *Postumum* (1800) 22:11.20–22. For discussion, see Graubner (1972, 37–45); Pollok (2017, ch. 4); Boyle (forthcoming-a).

<sup>6</sup> I use “intellect” to refer generically to the spontaneous cognitive powers, understanding, and reason, as I take Kant to employ the term “reason” in the title “Critique of Pure Reason”. See Willaschek (2018, 21–23).

<sup>7</sup> L-Jäsche, 9:14.22–27; cf. R3939 (1769) 17:356; R1612 (mid 1770s) 16:36.5–9; L-Ph (1772) 24:315.5–14, 316.7–10, 316.21–317.2; L-Blom (early 1770s) 24:24.25–27, 24:38–39; L-Wien 24:791.30–32, 792.1–2, 792.6–14; L-DW 24:695.24–29.

<sup>8</sup> I will not address the special sort of rational form that elevates merely correct cognition into science, i.e. into a systematic hierarchy of explanatory principles organized under a single idea. For an

exercises are *about* the formal laws of thought: Rationality is not their matter, is not the object or topic about which they entertain claims. Pure general logic, however, has rationality as both its form and its matter. Pure general logic is a “self-cognition of the understanding and of reason” in the sense that it brings to explicit consciousness – it “cognizes” as its “object” – the rules and principles that (consciously or unconsciously) inform every correct use of the intellect, merely as such.<sup>9</sup>

### 1.2.2 PGL as *Selbsterkenntnis* with Respect to Cognitive Ground

Pure general logic counts as a *Selbsterkenntnis* in a more specific sense as well – not merely with respect to the object it cognizes (namely, the form of thinking) but with respect to the *cognitive grounds* it relies on in cognizing this object.<sup>10</sup> Yet, apart from his repeated insistence that PGL is an a priori science and does not rely on experience, Kant provides no positive characterization of the cognitive grounds, the justificatory basis, we rely on in bringing to explicit consciousness the constitutive laws that in-form all thinking.<sup>11</sup> Indeed, he writes as though no special cognitive ground were necessary to move from an unconscious employment of the laws of thought in concreto to an explicit cognition of them in abstracto:

For there can be utterly no doubt that we cannot think or use our understanding otherwise than according to certain rules. Now these rules we can, in turn, think in their own right [*für sich selbst*], i.e. we can think them without their application or *in abstracto*. (L-Jäsche 9:12.3–7, my underlining)

This suggests that our explicit consciousness of the laws of thought in abstracto relies on nothing – no further evidence, faculty, or other

illuminating discussion of insight (*Einsehen*) and comprehension (*Begreifen*) as central to Kantian scientific cognition, see Schafer (2022).

<sup>9</sup> I use “self-cognition” and “self-knowledge” interchangeably, though I mostly leave “*Selbsterkenntnis*” untranslated. I agree with Schafer (2020a, 19–22) that the cognitive achievements embodied in Kant’s critical epistemology are importantly different from those embodied in his critical metaphysics, but I do not think this requires us to contrast *Erkenntnis* with *Wissen* in the present case (cf. Watkins and Willaschek 2017, 2020 [2017]; Schafer 2022, 2023).

<sup>10</sup> By “cognitive ground” I mean what Kant calls the “source” of a cognition (*Prolegomena* 4:266), i.e. what one would appeal to in justifying a particular judgment as knowledgeable: the evidentiary basis on which one stakes one’s claim to know. I use “cognitive ground”, “justificatory basis”, and “epistemic warrant” interchangeably to get at this idea.

<sup>11</sup> Maimon challenges Kant on this point and calls for a critique of logic to “determine those forms and make them complete by reflecting on the faculty of cognition” (Maimon to Kant, 2 December 1793, 11:471, my underlining; discussed in Lu-Adler 2018a, 161–169). Kant apparently felt it unnecessary to spell out the cognitive grounds of PGL, for reasons I will try to explain.

cognitive ground – beyond what we already rely on in “us[ing] our understanding” in the first place. This is true only of faculties that are essentially self-conscious. For it means there is a constitutive link between the faculty’s first-order acts “in concreto” and its representation “in abstracto” of the principles governing those acts. Only an intellect that is intrinsically capable of both (i) representing its acts, in the first person, as its own and (ii) representing those acts as satisfying certain normative principles can form a representation in abstracto of the rules governing its acts solely on the justificatory, cognitive basis of the first-order acts that are so governed. Granted, one need not *actually* represent one’s acts of thinking or the rules governing them in order to think at all. The cognitive acts in concreto merely make their *in-abstracto* counterparts possible. But it is essential to the first-order acts to do this. A mental episode cannot count as an act of thinking unless it provides a sufficient cognitive ground or epistemic warrant for an act of thinking (actual or only potential) about the constitutive formal principles governing that first-order act.<sup>12</sup> That’s why an abstract cognition of the laws of thought, *if* it occurs, relies on no cognitive grounds beyond those contained in the first-order act, whose form it characterizes.

Pure general logic is thus a *Selbsterkenntnis* not merely in the sense that it takes the intellect’s form as its object, but in the sense that it relies on the essentially self-conscious character of the intellect: its capacity for pure apperception. As Kant puts it in the *Anthropology*, “in logic we investigate according to what the intellectual consciousness offers up [*an die Hand giebt*]”, clarifying that by “the intellectual consciousness” he means “the I as subject of thinking (in logic), which signifies pure apperception (the merely reflecting I)” (7:134.21–26; cf. 141.1–4).

This account of PGL as cognitively grounded in the essentially self-conscious character of concrete acts of thinking provides a richer understanding of the priority of PGL. For such *Selbsterkenntnis* is a priori in the

<sup>12</sup> An acute reviewer asks, “What is Kant’s justification for claims such as this one?” The only answer I think Kant can consistently give is this: Such claims are known and justified *through apperception*. It is perhaps frustrating that the snake eats its tail here. But accepting such claims is, I take it, part of what it means to treat the intellect as essentially self-conscious: namely, to treat its acts as licensing a representation of their own constitutive essence. This is a defining commitment of Kant’s innovative, apperceptive method. Though there is no “further” justification to be had here, we can nevertheless elucidate such claims by elaborating Kant’s theory of apperception. I can only gesture at such an elaboration in the present work (see Section 1.3.3). For more fulsome discussions to which I’m indebted, see Merritt (2009, 2011); Kitcher (2011, ch.9); Engstrom (2013, 2016); Dyck (2017); and Land (2021 [2018], section 2.2).



“cognition from grounds” sense of the term.<sup>13</sup> Although self-conscious reflection on concrete acts of thinking obviously depends on such concrete acts as an enabling condition, what we reflect on, in articulating the formal laws of thought in PGL, are not the particular acts of thinking themselves but the lawful cognitive disposition that expresses itself in them. And we reflect on this cognitive disposition not only as something general, unlike the concrete acts themselves, but as something that determines and, in that sense, grounds the particular acts on which we are reflecting *as* the specific type of acts they are – namely, as *thoughts*, as acts of intellect. In bringing the form of thinking to reflective consciousness, then, we cognize the formal ground of our acts of thinking. So the self-conscious reflection we are considering is cognition from grounds. In such reflection, we know the concrete acts through the form of the capacity that gives rise to them; we know the effect through the cause, not the reverse. As Kant puts it in one of his logical Reflections: “Logic thus does not predate the use [of the understanding], but its rules, once they are cognized [*erkant*] at all, are indeed clear through themselves [*aus sich selbst*], because they contain the ground of all judgments, namely, their form” (R1602, mid-1770s, 16:32.2–4).

This sort of reflective self-consciousness is also a priori in the negative sense that it does not appeal to experience as source of evidence or justification.<sup>14</sup> Pure general logic not only cognizes effects (acts of thinking) through their ground (their form) but also cognizes this ground without appealing to experience to justify its claims, relying instead on the intellect's ability to make its own activity and form into an object of thought. This capacity for apperception, according to Kant, cannot be derived from experience, for it is a precondition of all experience. So PGL is a priori not merely in the negative sense that its principles are not empirical generalizations, but in the positive senses that (i) they cognize the formal ground of all thinking, as such, and that (ii) they do so by exercising a capacity for self-consciousness that is essential to the intellect and, thus, not derived from experience.

<sup>13</sup> See especially M-Mr 29:747.34–748.24 as well as *Critique* A758/B786.30–4; *Metaphysical Foundations* 4:470.18–19. For discussion, see Smit (2009) and Melamedoff-Vosters (2023, sections 1–2). All empirical cognition is a posteriori in this sense because experience is itself a *consequence* (not a ground) of the thought-independent realities we seek to cognize. So cognition based on experience is not cognition from the grounds of what is cognized but from its consequences (M-Mr 29:748.17–19).

<sup>14</sup> See A2.20–23 and B2f.15–20. As Pippin helpfully puts it, “a priori does not mean ‘not derived from experience’ but ‘known without appeal to experience’” (1982, 102).

This account of PGL's cognitive grounds as a pure *Selbsterkenntnis* of reason can initially seem rather obscure and mysterious. So it is important to register that this self-cognition of the form of thought in abstracto is not attained through some unmediated *Wesensschau* or a rarefied form of gnostic interiority, but through painstaking reflection upon, and reflective engagement in, concrete acts of thinking.<sup>15</sup> Kant emphasizes that although “the necessary and universal rules of thinking [...] can and must be cognized a priori independently of the natural use of the understanding and of reason in concreto”, it is nonetheless true that “they can only first be found [*gefunden*] through observation of that natural use”.<sup>16</sup> Reason's cognition of its own form is mediated through, though not justified by, “observation” of concrete exercises that manifest (because they are informed by) the laws in question. Clearly, such “observation” cannot consist in an empirical, inductive survey of de facto features we encounter in our thinking. Pure general logic is, as Kant likes to put it, “**abstracted** from the **empirical** use of the understanding, but not **derived** from it”.<sup>17</sup>

The fact that the formal rules of thinking *can* “be found” – and, indeed, can *only* “be found” – by “observ[ing]” the natural use of reason implies that acts of thinking, as such, involve an implicit conception of the formal rules that constitute thinking as thinking. If a conception of the laws of thought were not contained in these acts, then it would be impossible to recover those laws merely by reflecting on those acts. Indeed, when we say that acts of thinking, as such, involve an “implicit” awareness of the formal, constitutive laws of thought, all we (can) mean is that an explicit (clear) consciousness of those laws can be attained and *justified* merely through self-conscious reflection on that act – through mere “observation” and “abstraction” – without the introduction of further cognitive grounds or “data” beyond what is already at hand in the relevant act of thinking. This, as we saw, is what distinguishes PGL as a kind of self-knowledge with respect to its cognitive grounds and not merely with respect to its subject matter.

The laws treated of in PGL are *internal* to the acts of thinking that they govern in the sense that they are *represented in* those acts, though typically

<sup>15</sup> See Section 4.4 for an elaboration of what I take this to involve.

<sup>16</sup> L-Jäsche 9:17.27–31; cf. L-Ph (1772), 24:316.21–317.2; L-Wien 24:791.30–32, 792.1–2; L-DW 24:697.10–19.

<sup>17</sup> R1612 (mid-1770s) 16:36.9, original emphasis. This is one of a series of mid-1770s Reflections expressing the same idea: R1602, 16:31.22–32.4; R1603, 33.10; R1607, 34.11. See also R1620 (1780s) 41.14–16; R1627 (1790s) 43.9–10; L-Wien 24:792.1–2; *Anthropology* 7:133.31–33.

unconsciously.<sup>18</sup> That is to say, every act of thinking, merely as such, must involve an appreciation – conscious or unconscious, explicit or implicit – of the formal laws that constitute it as an act of thinking, even if it only imperfectly satisfies the laws it represents (perhaps unconsciously) as constitutively normative for it. By “appreciation”, I mean not only that these laws (or the cognitive disposition they characterize) exercise a *de facto* influence on our thinking, nor merely that our thoughts involve a representation of such laws, albeit unconscious. I mean that these laws exercise an influence on our thinking, though not an insuperable one, and that they exercise this influence *in virtue of* our representing them, albeit unconsciously, as binding on our thought.<sup>19</sup> Acts of thought are, in this sense, guided by an internal representation of the constitutive, formal principles of thinking. Such guidance manifests itself not only in a subject's disposition to assent to such principles when explicitly presented with them but also in her disposition to recognize the relevance to her thinking (or others') of certain kinds of questions or challenges – about, say, contradictions or logical entailments – and in her readiness to revise her thinking in response to such challenges (or to urge such revisions on others).<sup>20</sup>

This conception of internal guidance is compatible with the fact that particular acts of thinking may not perfectly satisfy the laws they internally, if only unconsciously, represent as unconditionally binding on themselves.<sup>21</sup> Similarly, acts of moral turpitude do not undermine Kant's claim that a representation of the moral law belongs to the form of practical reason, as such. For what it means to represent a law as unconditionally binding on one's act is to accept its authority as a standard of correctness against which one's act is to be assessed. This does not entail that every act meets the standard it sets for itself. Even when one adds that the logical laws are, like the moral law, formal and thus constitutive with respect to the acts they govern, logical blunders remain possible and explicable.

<sup>18</sup> On unconscious representations, see L-Jäsche 9:33–34; *Anthropology* 7:135–137; R176–177 (prior to 1770?) 15:64–66; A-Mrongovius 25:1221.16–21. For discussion, see La Rocca (2008a, 2008b) as well as the essays by Crone, Heidemann, Kitcher, Rockmore, and Schulting collected in Giordanetti et al. (2012).

<sup>19</sup> Compare Kant's remark that “[e]verything in nature operates [*wirkt*] in accordance with laws. Only a rational being has the capacity to act [*handeln*] in accordance with the representation of laws, i.e. in accordance with principles” (*Groundwork* 4:412). Though Kant makes this point specifically about the faculty of practical reason (i.e. the will), I take an analogous claim to hold for all rational capacities, as such. Here I follow Schafer (2020a).

<sup>20</sup> For elaboration of this idea, see Land (2021 [2018], esp. section 2.2).

<sup>21</sup> This is well-trodden ground (MacFarlane 2002, 37; Smit 2009, 214; Nunez 2018; Land 2021 [2018]; Boyle forthcoming-a).

Logical laws characterize the proper functioning of the *faculty* of thought.<sup>22</sup> They thereby set a standard of correctness for particular acts of thinking, as such. A mental act counts as an act of thinking just insofar as it is subject to this standard of correctness. But an act can be subject to a standard without perfectly satisfying it.<sup>23</sup> And in the case of essentially self-conscious capacities, it suffices that the act *represent itself* as subject to the relevant standard in order for it to *be* subject to that standard. Thus, the special character of reason's *Selbsterkenntnis* in PGL reflects not only the essentially apperceptive character of the intellect but also its essential *autonomy*, in that its acts are subject only to laws that they internally represent as binding on themselves.<sup>24</sup> Indeed, its acts constitute themselves as the types of acts they are – as acts of thinking, or as exercises of practical reason, say – precisely in promulgating, albeit unconsciously, the laws by which their own correctness is to be assessed. Even acts that fail to fully satisfy the formal principles that constitute them as the types of acts they are nevertheless represent those principles, albeit unconsciously, as unconditionally binding on themselves. Such acts, which are all too possible, are flawed by their own lights. That is why Kant characterizes such cases not as violating some independent principle or as conflicting with reality but as a failure of reason to agree *with itself*.<sup>25</sup>

### 1.2.3 The Internality of Selbsterkenntnis

We can get a more concrete sense of the internality of logical laws to all acts of thinking by considering a particular logical principle: the law of non-contradiction. It is partly constitutive of what it is to think that *p* that one's thought is, consciously or unconsciously, guided by the principle

<sup>22</sup> For a trenchant exploration of this point, see Boyle (forthcoming-a). Cf. also Smit (2009, 214, 222n.13); Land (2021 [2018], 7).

<sup>23</sup> Kant does seem to hold that some degree of actual compliance with the relevant standard is required for an act to count as the type of act that is constitutively governed by that standard. My point is just that perfect conformity is not required. Indeed, perfect compliance may be impossible or unverifiable for finite minds. See L-Blom (early 1770s) 24:93.28–94.20; L-Wien 24:825.33–826.5; L-DW 24:721.5–6, 722.6–8; L-Jäsche 9:54.22–28; as well as note 26.

<sup>24</sup> Apperception, Kant writes, is "*autonomia rationis purae*" (Postumum, 1800–1803, 21:81.30). The autonomy of reason is a consequence of Kant's view that reason is essentially self-conscious, because this implies that its acts are self-constituting. Kant claims credit for being among the first to recognize this consequence and grasp its significance, but if he is right, it is something all rational creatures – including Kant's predecessors – are implicitly (no doubt unconsciously) committed to. Compare Kant's clarification that his categorical imperative is not a new moral principle but a new *formulation* of a principle everyone already (and necessarily) accepts (*Practical* 5:8n.).

<sup>25</sup> B115.3–4; R1628 (1780s) 16:46.4–5; R1629 (1780s) 16:47.17–19. See Engstrom (2009, 131–134).

that one ought not conjointly think *not-p* (nor anything that would entail *not-p*).<sup>26</sup> That is to say, it would not so much as count as an act of thinking if the subject thinking *p* did not, in that very act, represent the principle of non-contradiction as unconditionally binding on her thought that *p* and suitably related contents. Part of what it *is* to think that *p* is to appreciate that one's thought is subject to the schema  $\langle \textit{not} (p \textit{ and } \textit{not-p}) \rangle$ .

To claim that the law of non-contradiction must be internally represented in all acts of thinking is tantamount to denying that it is a distinct thought in its own right. According to Kant, the laws of logic cannot be conceived as a set of separate, if privileged, thoughts that stand alongside all our other thoughts, as Quine's "web of belief" would have it.<sup>27</sup> Now a Quinean account might preserve Kant's idea that a standing belief in certain logical principles is a normative, even constitutive, condition on one's status as a thinker. Its core idea is rather that these "logical beliefs" are distinct from a subject's day-to-day beliefs, such as her belief that meerkats bark.

What Quine's account fails to appreciate, from a Kantian standpoint, is the complete generality of logical laws as formal principles. The principle of non-contradiction governs the relations in which *all* thoughts may permissibly stand – including purportedly distinct thoughts about the laws of logic. If one treats logical commitments as distinct from other determinate acts of thinking, then a subject who simultaneously holds the belief that *p* and the belief that *not-p* is merely committed to an inconsistent triad, made up of this pair of contradictory beliefs along with her belief in the law of non-contradiction.<sup>28</sup> This makes it seem as though she could restore consistency to her web of belief by rejecting the law of non-contradiction, while retaining the beliefs that *p* and that *not-p*. What is absurd about this proposal is not that such a subject would, by rejecting the principle of non-contradiction, undermine her status as a thinker and, thus, cease to *think* the contents in question. There may be something tragic in a subject's self-exile from the space of reasons, but it is not incoherent. From a Kantian perspective, what is absurd in the Quinean scenario is that the subject's *reason* for rejecting the principle of

<sup>26</sup> In proscribing *anything* that would entail *not-p*, the principle of non-contradiction implicates an indefinite set of cognitions. It thereby imposes a sort of systematicity, a total classification, on one's thoughts, both actual and possible. This is analogous, I take it, to the universalizing import of the moral law, both in its "natural law" formulation and insofar as a finite agent is morally compelled to represent her maxims within a "kingdom of ends".

<sup>27</sup> Quine (1980 [1951]).

<sup>28</sup> Quine speaks of a web of beliefs, rather than of mere thoughts, but the point is the same.

non-contradiction could only be that *it contradicts* other beliefs she holds. So her rejection of the principle would, absurdly, manifest her allegiance to it. This enduring commitment to the principle of non-contradiction, even in the act of rejecting it, demonstrates the “internality” of logical laws to all her acts of thought – including the thought that the principle is false.

The subject’s adherence to logical laws cannot be sequestered in a special set of beliefs, separate from the other thoughts she entertains. An implicit appreciation of the principle of non-contradiction will always insinuate itself into (or “in-form”) her other acts of thinking. For if a subject is capable of appreciating that there is a contradiction between her belief in the principle of non-contradiction and her conjoint beliefs that  $p$  and that  $\text{not-}p$ , then she must have the more basic capacity to directly recognize the contradiction between her belief that  $p$  and her belief that  $\text{not-}p$ . She must have this latter capacity because the sentence letters we are using are purely schematic, so we can just rewrite the former scenario (the inconsistent triad) as the latter (the dyadic contradiction) by interpreting  $\langle p \rangle$  to be her contradictory belief pair and  $\langle \text{not-}p \rangle$  to be the principle of non-contradiction itself. Thus, even though it is possible, as Kant himself claims, for a subject to represent the principle of non-contradiction on its own and in abstracto, the ability to do so presupposes a more fundamental recognition of the principle that is internal to her capacity for thinking anything at all. That is to say, only a subject whose logical knowledge is internal to – that is, implicit in, but explicatable through mere self-conscious reflection on – her acts of thinking is in a position to frame and apply logical principles in abstracto as a distinct and privileged set of beliefs. The Quinean account of logical knowledge is otiose. For it presupposes just the sort of internal, apperceptive, autonomous grounding of the laws of logic that it proposed to supplant.

#### 1.2.4 *The Resulting Picture of PGL*

What picture of PGL does this leave us with, so that we may return to Kant’s comparison of logic to critique? Pure general logic has a proper subject matter, namely the formal rules governing all thinking, regardless of its object, its (empirical or a priori) relation to that object, or its content. This is what makes it a general logic. And it relies on a special sort of cognitive ground in cognizing this subject matter: namely, the essentially self-conscious character of the intellect. This is what makes it a pure logic, an a priori science. It is a priori both (i) in the sense that a self-conscious cognition of the form of thought is a cognition of what grounds all

particular thoughts and (ii) in the sense that the capacity for such self-consciousness is not derived from experience. But this independence from experience merely concerns the justification or epistemic warrant of such self-conscious cognition. It does not mean that such self-consciousness is possible in the absence of experience. Quite the contrary. It means that the possibility of such a self-conscious representation of the form of thinking is inscribed into all acts of intellection, as such. The knowledge PGL presents in abstracto is distilled through painstaking reflection on particular acts of thinking in concreto – reflection that aims to bring to explicit consciousness the appreciation of the formal rules of thought that necessarily, if unconsciously, informs and guides those acts, thereby constituting them *as* acts of thinking. The underlying assumption of this reflective method is that all acts of thinking contain a conception of the essence of thought *überhaupt*. It may take great intellectual labor to isolate and unpack this conception. But such labor is fundamentally analytical in that its sufficient cognitive ground is contained in the acts whose form it seeks to characterize.

### 1.3 Critique as a *Selbsterkenntnis* of Reason

In what respects does critique resemble PGL and how, precisely, do they differ? We are helped toward an answer not only by Kant's letter to Garve but also by his various discussions of PGL, which present the same picture. Notably, the passages from L-Jäsche we have just been considering, where Kant characterizes PGL as reason's *Selbsterkenntnis* of its own form, draw the same comparison between logic and critique:

[PGL] is a science of reason not merely with respect to form but with respect to matter, since its rules are not taken from experience and since it also has reason for its object. Logic is hence a self-cognition of the understanding and of reason, but not with respect to their capacities in regard to objects, but merely with respect to form. In logic I do not ask what the understanding knows [*erkennt*] and how much it can know or how far its knowledge extends. For that would be self-cognition with regard to its material use and thus belongs in metaphysics. In logic, there is only the question: how will the understanding cognize itself? (L-Jäsche 9:14)<sup>29</sup>

The “material use” of the intellect is its use in cognizing objects that do not depend, for their existence or inner constitution, on the acts through

<sup>29</sup> Cf. L-Wien 24:857.1–10; M-Mr 29:755.35–756.4; R5644 (1783–1784) 18:286.15–16.

which they are thought or cognized.<sup>30</sup> This material use comprehends the practical use of reason, in cognizing what is good (pursuit-worthy, what ought to be), as well as the theoretical use of reason, in cognizing what is true (belief-worthy, what is).<sup>31</sup> Kant mentions several questions one might raise about the material use of the intellect: namely, “what the understanding knows”, “how much it can know”, and “how far its knowledge extends”. These are precisely the questions that define the critical project.<sup>32</sup> They correspond to Kant’s claims to Garve that critique “derives, from the nature of [sc. ‘the sole pure cognitive faculty’ (10:340.11–12), i.e. ‘a reason that judges a priori’ (340.6)], all objects to which it extends” (340.14–15) and that critique can do this by “unfolding a priori out of the mere concept of a faculty of knowledge (if it is precisely determined) all objects and everything that one can know of them” (340.17–21).<sup>33</sup> So it is *critique* that L-Jäsche characterizes as “self-cognition [of the understanding] with regard to its material use”.

L-Jäsche thus presents us with the same elucidatory comparison of PGL and critique as Kant’s letter to Garve, but helpfully expands on the idea that these sciences are “formal and necessary” (10:340.9–10), adding that each is “a self-cognition of the understanding and reason” (9:14.25).<sup>34</sup> The difference is that PGL treats of reason in all its uses, whereas critique considers reason specifically in its pure but material use, i.e. in its a priori cognition of objects that exist independently of our acts of cognizing them. Critique is the *Selbsterkenntnis* of reason as a faculty of a priori yet synthetic cognition.<sup>35</sup>

<sup>30</sup> See B23.30–31, quoted subsequently. This is what I will mean by “thought independence”. Of course, thought-independent objects may be mind-dependent in various other respects. In particular, they may depend for their *intelligible form* (i.e. for their cognizability), on the *capacities* of sensibility and understanding (though perhaps not on particular acts of those capacities), so that any knowable object must exhibit the forms of space and time as well as the forms of unity prescribed by the categories. I will not, however, address Kant’s idealism in any detail.

<sup>31</sup> Bix-x.31–02; *Groundwork* 4:387.8–16; M-Mr 29:753.24–754.3.

<sup>32</sup> See, for instance, A3.18–8/B7.30–8. Kant admittedly says the answer to these questions “belongs in metaphysics” (9:14.30). But I take this to be a capacious use of the term, on which critique, as prolegomenon and canon, is both the first part and the complete idea of a scientific metaphysics. See especially A841/B869.20–29; cf. Bxxiii.11–xxiv.24; *Prolegomena* 4:365.25–28; *Metaphysical Foundations* 4:469.30–470.1; R4284 (1770–1775) 17:495.9–16.

<sup>33</sup> See also *Prolegomena* 4:365.17–25; *Metaphysical Foundations* 4:475n.37–42.

<sup>34</sup> Further characterizations of critique as the *Selbsterkenntnis* of reason include A735/B763.13–15, A763/B791.6–8, A849/B877.14–18; *Prolegomena* 4:328.20–25; *Tone* 8:390.30; R4892 (1776–1778) 18:21.2–5; R2667 (after 1790) 16:459.20–22; notes for *Metaphysics of Morals* 23:402.11–14; M-Mr 29:756.10, 783.21–22.

<sup>35</sup> Cf. A10.16–21, B19.28–02; *Prolegomena* 4:274.6–22; *Judgment* 5:289.3–5; R5133 (late 1770s or early 1780s) 18:101.



Now, the previous section argued that PGL is *Selbsterkenntnis* with respect to (i) its object (its subject matter) as well as (ii) its cognitive ground (its justificatory basis or epistemic warrant). Pure general logic (i) takes the form of the intellect as its object of inquiry and (ii) it bases its claims about this object on self-conscious reflection upon, and reflective engagement in, particular acts of intellection, merely as such. Pure general logic is *Selbsterkenntnis* in the special sense that has its sufficient cognitive ground in the essentially apperceptive character of the intellect, as realized in particular acts of thinking. If critique is likewise *Selbsterkenntnis* in this twofold sense, then it should be possible to adduce the principles of the critical philosophy through painstaking but a priori reflection on particular acts of material cognition. The principles of the critical philosophy ought to implicitly in-form all object-directed cognition, as such. Accordingly, if critique is *Selbsterkenntnis* in this twofold sense, the principles of the critical philosophy will be constitutive of what is to count as discursive knowledge of objects in the same way that the principles of logic are constitutive of what is to count as thought *überhaupt*. Here, too, it may turn out that particular acts of cognition fail to perfectly satisfy the principles that constitute them as material cognitions. But such cases will be flawed not with respect to some external standard but insofar as the intellect fails to agree with itself – insofar as its acts flout principles that they internally represent as binding on themselves.<sup>36</sup> This, I will argue, is precisely Kant's view.

### 1.3.1 Critique as *Selbsterkenntnis* with Respect to Subject Matter

Now, one might think this account of critique is a non-starter. For Kant is clear that PGL is a purely formal science, whereas he specifically associates critique with the material use of the intellect. However, critique can be *concerned with* the material use of the intellect without itself *constituting* a material use of the intellect. Kant is quite explicit about this:

This science [“the critique of reason” (B23.21)] also cannot be terribly wide-ranging, because it does not have to do with objects of reason, whose variety is infinite, but merely with [reason] herself, with tasks that spring wholly from her own womb and are set for her not by the nature of things

<sup>36</sup> Axii.16–18, 21–22, A484/B512.24–25, A293f./B350.20–25; L-Ph (1772) 24:315.23–33. Such failures of self-agreement range from mild forms of heteronomy, in which sensibility exercises an undue influence on one's act of intellection (A294/B350f.4–9), to blatant antinomies, in which the intellect's own internal standards come into conflict through “subreptive” misapplication (A643/B671).

that are distinct from her but through her own [nature]. For when she has already [zu]vor] come to know completely her own capacity in regard to the objects that may crop up in experience, it must be easy to determine completely and with certainty the scope and the boundaries [Grenzen] of her attempted use beyond all bounds of experience. (B23)<sup>37</sup>

Critique is not concerned with “objects of reason” as its subject matter but “merely with [reason] herself”. Now the faculty of reason that critique undertakes to study is, for its part, concerned to cognize “objects that are distinct from it” and to cognize them “beyond all bounds of experience”. So critique is *indirectly* concerned with objects of reason. For it investigates the faculty of reason in its material use. But this does not make critique itself a material use of reason. It rather means that critique is not concerned with reason *überhaupt* but specifically with “a reason that judges a priori”, as Kant writes to Garve (10:340.6). Such an investigation is still *formal* in the sense that it concerns reason’s essence, its fundamental character. Critique addresses “tasks that spring wholly from [reason’s] own womb”, tasks that are set for it “through her own nature” (my underlining), rather than issues that may arise from “things that are distinct from her”. Critique is reason’s attempt to “come to know completely her own capacity in regard to the objects that may crop up in experience” (my underlining). And it is this concern with the fundamental nature of reason – its form – that enables critique “to determine completely and with certainty the scope and the boundaries of [reason’s] attempted use beyond all bounds of experience”.

Kant’s initial attempt to distinguish his critical enterprise from traditional metaphysics draws a contrast between their respective subject matters.<sup>38</sup> The subject matter of metaphysics consists in thought-independent objects, which reason purports to cognize a priori. The subject matter of critique, by contrast, is the *faculty of reason* that aspires to such a priori cognition of thought-independent objects. Critique consists of what Kant calls “transcendental cognition”, which “essentially concerns [*überhaupt beschäftigt*] itself not so much with objects [*Gegenständen*] but rather with our manner of cognizing [*Erkenntnisart*] objects, insofar as this [sc. cognition] is supposed to be possible a priori” (B25.1–5).<sup>39</sup> This distinction in subject matter applies to “critical” as well

<sup>37</sup> See also Aiv.20–28, Bxxiii.11–01, A12.27–A13.3/B26.28–4; *Prolegomena* 4:274.27–35.

<sup>38</sup> See also *Progress* 20:343.29–31; M-Mr 29:750.18–751.10, 752.11–18, 756.1–17, 779.24–37, 780.27–31; M-Vi/K<sub>3</sub> (1794–1795) 29:949.14–24.

<sup>39</sup> Cf. A11–12.1–4, A56/B80.11–19; A751/B779.10–16; *Judgment* 5:176.19–23. Kant’s hedge “not so much with objects” suggests that this distinction may not be exclusive. Indeed, it turns out that

as “dogmatic” metaphysics.<sup>40</sup> Critical metaphysics consists of those synthetic a priori judgments that are genuinely knowable, in light of the faculty analysis prosecuted by critique. But metaphysics and critique proper (which I will call “critical epistemology”) nevertheless have different topics, and Kant accordingly treats them in different sections of his critical texts.

The core of Kant's critical epistemology is established in the Transcendental Aesthetic and the Analytic of Concepts. It is there that claims directly addressing the nature of our cognitive faculties are concentrated: for example, that sensibility and understanding are the two stems of human knowledge (A15/B29.7–12), that intuition involves an object's being given to the mind (A19/B33.5–10), that objects are given to us only insofar as they affect us (A19/B33.9–12), that the understanding is spontaneous (A51/B75.1–6), that it can be regarded as a capacity to judge (A69/B94.19–26), that the “I think” must be able to accompany all my representations (B131.4–132.8; A107.25–29, A118n.18–32), and so on. This critical epistemology vindicates the possibility of a genuinely knowledgeable metaphysics of nature, which is subsequently outlined in the Analytic of Principles. Here we encounter claims that are directly concerned with thought-independent objects, such as the principle that all appearances are extensive magnitudes (B203.18–19), that real properties are continuously gradable (B211.6–11), that substance endures throughout all change (A182.23–25, B222.4–6), and so on. The Dialectic, of course, is likewise preoccupied with claims about thought-independent objects, though Kant's aim there is to expose them as illusory, insofar as they are incompatible with the critical epistemology, and to diagnose our temptation to believe them. Here we encounter traditional metaphysical dogmas, e.g. that the soul is a substance (A344/B402), that the world has a beginning in time (A426–8/B454–6), and so on.

With respect to subject matter, then, critical epistemology is no less a *Selbsterkenntnis* of reason than is PGL. Critique does not exhibit the thoroughgoing formality of PGL, since it is concerned with the material

an investigation of our *faculty* for a priori cognition yields an a priori account of the *objects* such a faculty can cognize. But this move requires Kant's signature idealism: “a priori cognition can attribute nothing to objects except what the thinking subject takes from within herself” (Bxxiii). Only within such an idealistic framework can transcendental philosophy, as a science of the faculty of speculative reason, amount to an ontology, as a science of being (A247/B303.17–23; cf. A845/B873.29–02).

<sup>40</sup> Kant's use of the term “metaphysics” varies somewhat, but its core meaning is the material use of reason a priori, i.e. the pure rational cognition of thought-independent objects. This comprehends the constitutive principles of practical reason, the subreptive speculations of traditional metaphysics, and the immanent principles of transcendental ontology. See *Groundwork* 4:388.4–8.

use of the intellect, but neither is it a sort of material cognition. In both sciences, reason investigates its own form. Their difference lies in the degree of abstraction with which they consider their common subject matter, namely reason's form.<sup>41</sup> Reason has just one form, of course, but it can be characterized more or less abstractly. Pure general logic characterizes this form in abstraction from all material use of the intellect, whereas critique takes account of reason's use in cognizing objects a priori and thereby offers a more concrete, but still formal, account of what reason most essentially is.<sup>42</sup> Now there are, Kant says, just two ways that material cognition can relate to its object: namely, (i) theoretically, by "merely **determining** it" through concepts, and (ii) practically, by "also **making** it **actual**" (Bix-x, Kant's emphases). So there will be a critique of theoretical reason as well as a critique of practical reason. And these critical inquiries into the form of reason in its theoretical and its practical applications are specifically concerned with reason's capacity to cognize its object a priori – that is, with reason's character as a "pure faculty of knowledge [*reinen Erkenntnisvermögen*]" (Kant to Garve, 10:340.12).<sup>43</sup> So while both PGL and critique are a priori inquiries into the form of the intellect, PGL abstracts from both the *objects* of thought and the *manner* in which thought relates to its objects – empirically or a priori, theoretically or practically. Critique abstracts from the former but not the latter. Pure general logic is reason's *Selbsterkenntnis* of its form as a faculty of *thinking*; critique is reason's *Selbsterkenntnis* of its form as a faculty of *a priori cognition*, either practical or theoretical.

### 1.3.2 Critique as *Selbsterkenntnis* with Respect to Cognitive Ground

Thus far, we have fleshed out the idea that critique is a *Selbsterkenntnis* of reason with respect to its subject matter: Critique is reason's investigation

<sup>41</sup> This reading is controversial (see note 4). Others hold that critique – and, in particular, Transcendental Logic – has a different subject matter from PGL. The approach I am sketching can accommodate this. My main point is not that PGL and critique have the very same faculty as their object, but that, whatever faculty each has as its object, each embodies a *Selbsterkenntnis* of that faculty with respect to subject matter and cognitive ground. The core of my interpretation is that, as a *Selbsterkenntnis* with respect to cognitive ground, the claims of Kant's critical epistemology are justified through apperceptive reflection on (self-conscious engagement in) material cognition of objects.

<sup>42</sup> Kant calls critique a "formal science" in *Prolegomena* 4:262.5; cf. *Tone* 8:404.3–21.

<sup>43</sup> For reasons that do not affect the present argument, Kant prefers to speak simply of a critique of practical reason, not of *pure* practical reason. Kant argues that, if reason can determine the will at all, it must do so a priori. So the critical question comes down to whether reason can be practical *at all*, i.e. whether our will is free (*Practical* 5:15.16–16.2).

of its own essential principles, as a faculty of a priori material cognition. Whether critique is also a *Selbsterkenntnis* with respect to its cognitive ground or justificatory basis is a more difficult question. For, as in the case of PGL, Kant provides almost no positive characterization of the epistemic warrant enjoyed by critique, beyond repeatedly insisting that it is an a priori science. But some of his remarks are quite suggestive.

For instance, Kant claims that we can reasonably “hope to complete such a system” of critical epistemology because:

the object here is not the nature of things, which is inexhaustible, but rather the understanding, which makes judgments about the nature of things – and the understanding [is the object] only with respect to its cognition a priori – the advantage of which is that, since we clearly need not seek for it externally [*ihm doch nicht auswärtig suchen dürfen*], [the understanding] cannot remain hidden from us[.] (A13/B25)<sup>44</sup>

Kant's first point is familiar to us: Critique has a special “object” (i.e. subject matter) – namely, “the understanding [. . .] with respect to its cognition a priori”. But he now adds that the nature of this subject matter is such that “it cannot remain hidden from us” because the understanding is not something that we must seek out “auswärtig” – “externally”, or “elsewhere”. This recalls Augustine's dictum: “Let not the mind, therefore, seek itself as though it were absent, but let it take care to discern itself as present.”<sup>45</sup> Kant contrasts “seek[ing] elsewhere” with the epistemic procedure of critique, which, by implication, need not seek, since its object is already manifest: its object is not “elsewhere” but *here* in what already lies before it. And what already lies before us, in critique, are precisely the understanding's “judgments about the nature of things”. That is why the

<sup>44</sup> See also Kant's remarks in the A-edition Preface: “Nothing can escape us here, because what reason produces entirely of herself cannot conceal itself but will be brought to light by reason herself as soon as one has only discovered the common principle of it [sc. what reason produces of itself].” (Axx.14–18) Again: “[In the Critique] I have solely to do with reason herself and her pure thinking, for extensive acquaintance [Kenntnis] with which I need not search far beyond myself, because I encounter her in me myself, and common logic already provides me an example of how to completely and systematically enumerate all her [sc. reason's] simple acts” (Axiv.18–25).

<sup>45</sup> *De Trinitate* 10.9.12, cited in Boyle (2011). Augustine continues:

Let it not know itself as though it did not know itself, but [let it know] how to distinguish itself from that which it knows to be another thing. When it hears the command “Know thyself”, how will it be able to carry it out if it does not know what “know” means, and what “thyself” means? If, however, it knows what both mean, then it also knows itself.

It is customary to align Augustine with Descartes (e.g. Menn 1998; Matthews 2003); yet Kant's conception of critical epistemology also bears suggestive similarities to Augustine's account of intellectual self-inquiry. A direct influence of Augustine or Augustinianism is doubtful, however.

understanding “cannot remain hidden from us”: because it is already manifest in our judgments about objects.

This is not to say that we might not overlook important features of our understanding or that we are immune to error in our characterizations of it. As Kant stresses early in the Introduction, it is only “once long practice has made us attentive to it and adept at separating it out” that we come to appreciate “what our own faculty of cognition (merely occasioned by sensible impressions) provides [*hergibt*] of itself” (B1f.21–26). The sense in which the understanding “cannot remain hidden from us” is rather that everything we require to give a correct and complete account of its form is already open to view in our intellectual judgments (*Verstandesurteile*) about objects. If the nature of the understanding is hidden, it is hiding in plain sight.

Critique, then, is a *Selbsterkenntnis* of the understanding with respect to its cognitive ground. Its claims, like those of PGL, are based on the essentially apperceptive character of the understanding, for what they bring to light is already contained in the exercises of the faculty whose form they characterize. The claims of critique are established through mere reflection upon, and reflective engagement in, concrete acts of judgment.<sup>46</sup>

Kant makes a similar point in the B-edition Preface, which likewise indicates that critique is distinguished not merely by its subject matter but by a special epistemic relation to that subject matter:

[The critique of pure speculative reason] is a treatise on method, not a system of the science [sc. of pure speculative reason] itself; but it nevertheless outlines the entire structure of that science, with respect to both its boundaries and its internal articulation. For pure speculative reason has this particularity: that she can measure out [*ausmessen*] her own capacity, according to the different manners in which she selects objects for thought [. . .]; because, with respect to the first point [sc. that reason can measure out her own capacity], nothing can be attributed [*beigelegt*] to the object in a priori cognition except what the thinking subject takes from within herself [. . .] Accordingly, metaphysics has the rare fortune, which redounds to no other rational science that has to do with objects (for **logic** busies itself only with the form of thinking in general), that, once critique has set it upon the sure path of a science, it [sc. metaphysics] can [. . .] complete its work [. . .] for it has to do merely with the principles and limitations of her [sc. pure speculative reason’s] use, which are determined through [reason] herself. (Bxxii–xxiv, Kant’s emphasis; cf. A467f./B504f.; *Prolegomena* 4:366.12–22)

<sup>46</sup> For discussion of how such reflection proceeds, see Section 4.4.

Kant is again claiming a principled completeness for critique – and, by extension, for the reformed metaphysics that critique exhaustively “outlines”. Kant links this completeness to his claim that a priori cognition of objects can only concern “what the thinking subject takes from within herself” (my underlining). Thus, the content and extent of metaphysics, as a priori cognition of objects, will be determined by the nature of the thinking subject, viz. by the essential character of reason. The task of critique, then, is to explicate the essence of speculative reason, the constitutive principles that characterize its form. This much is familiar. What is new is the suggestion that we are in a position to explicate these principles – to prosecute a critique of pure speculative reason – precisely because they “are determined through [reason] herself”.<sup>47</sup> This is an expression of reason's autonomy – of the fact that reason's exercises are governed by principles that they internally represent as binding on themselves.

We have already seen that the purely formal laws of PGL are self-legislated in this sense (Section 1.2.2). Kant is now claiming that the same autonomy is involved in *material* uses of reason. Reason “determines” for herself “the principles and limitations” of her own use “according to the different manners in which she selects objects for thought”. This is her autonomy. Accordingly, the epistemic basis of critique, which delivers a cognition of the form of reason in its material use, will consist in self-conscious reflection on, and reflective engagement in, concrete acts of material cognition – the acts through which its constitutive principles are self-legislated. Its aim is to bring to explicit consciousness the principles that such acts internally, if unconsciously, represent as binding on themselves, insofar as they stem from an autonomous rational faculty – a faculty that “determines” its own principles of operation.

### 1.3.3 *The Immanence of Critique as Selbsterkenntnis*

These discussions in the introductory sections of the *Critique* are merely suggestive, but they are reinforced by the accounts of apperception, synthesis, and judgment that Kant subsequently develops in the

<sup>47</sup> Cf. *Prolegomena* 4:349.31–38; *Critique* A751/B779.20–21. See also Kant's remark in the *Anthropology*:

If by the word “understanding” is meant the faculty of cognition by means of rules (and thus through concepts) in general, so that it comprehends the entire higher faculty of cognition, then these rules are not to be understood as those through which nature guides the human being in his conduct [...] but only those that he [the human being] himself makes (7:197, my underlining).

Analytic.<sup>48</sup> Consider Kant's well-known claim that "[t]he **I think** must **be able** to accompany all my representations" (B131, Kant's emphases). This says that I can, in principle, become explicitly conscious of any of my representations and can do so without relying on empirical observations. The power to do this is called pure apperception, which Kant considers an essential characteristic of rational faculties, as such.<sup>49</sup> As Kant emphasizes, however, apperceptive consciousness need not actually accompany all my representations.<sup>50</sup> Some of my representations may remain unconscious – indeed, Kant thinks the vast majority do.<sup>51</sup> What is essential to them, as representations belonging to a rational faculty, is that it is *possible* for me to become explicitly conscious of them.<sup>52</sup> Now, on the interpretive approach I've been sketching, the principles of critical epistemology are supposed to emerge through apperceptive reflection on material uses of the intellect. So the question we now face is whether Kant takes acts of material cognition to essentially involve a *representation*, albeit unconscious, of the constitutive principles governing material cognition, as such. If he does, then critique can explicate those principles through pure apperception.

Kant's discussions of synthesis and judgment in the B-edition Transcendental Deduction suggest that he does indeed take acts of material cognition to essentially involve a representation of the general principles that constitute them as material cognitions – just as acts of thinking are, as such, guided by an appreciation of the laws of logic that constitute them as thoughts. Material cognition, as opposed to mere thought, is based on synthesis: i.e. the combination (*Verbindung*) of a manifold that is given independently of the act of thinking and combining it (B130.19–24).<sup>53</sup>

<sup>48</sup> I can offer nothing approaching an adequate treatment of these matters here. My aim is not to defend this reading against all alternatives but to outline an account of Kant's critical methodology that is intelligible and plausible. The more important test of my interpretation is the light it sheds on Kant's actual prosecution of the critique, not the degree of credence it can compel in advance and in abstracto. What I say here is indebted to and supported by the more detailed discussions of Engstrom (2009, ch. IV; 2013; 2016); Merritt (2009; 2011; 2018, ch. 3); Land (2021 [2018]); and Schafer (2019; 2020a; 2021 [2018]; 2022).

<sup>49</sup> See B134n.35–38, A114.32–02, A117n.30–32; *Anthropology* 7:127.4–14.

<sup>50</sup> Cf. A117n.27–32; Kant's letter to Beck, 4 December 1792, 11:365.6–8. <sup>51</sup> See note 18.

<sup>52</sup> Precisely *how* we are to become conscious of them (i.e. what is involved in actually exercising our power of pure apperception) is a question we will touch on in Sections 2.5 and 4.4. A full account would take us too far afield.

<sup>53</sup> I do not mean to suggest that thinking does *not* involve combination, but merely that PGL takes no account of this. Because PGL abstracts from the origin of our representations, it ignores whether the concepts combined in a judgment are "given" independently of that act of combination or whether they are "made" through that act. Accordingly, the laws of PGL hold for every intellect whatsoever, whether discursive (i.e. dependent on receptivity) or intuitive (i.e. spontaneously productive).



Material cognition, for Kant, is not a passive absorbing of information but an active sense-making and stance-taking that involves the forging of rule-governed connections. As Kant says later in the *Dialectic*, “the senses cannot err [...] because they do not judge. Thus truth as well as error [...] are only to be met with in judgment” (A293/B350). One of Kant's pivotal claims early in the *Deduction* is that combination is an act of spontaneity, which he takes to entail that “we can represent nothing as combined in the object [*Objekt*] without having previously combined it ourselves” (B130). But the act of synthesizing or combining a manifold, Kant claims, always involves a representation of the unity that is to result from such synthesis:

Combination is the representation of the **synthetic** unity of the manifold. The representation of this unity cannot therefore arise out of the combination, rather it makes the concept of the combination possible in the first place by being added to the representation of the manifold. (B131, Kant's emphasis)

What is important here for our purposes is that combination involves a representation of unity – some concept of oneness that guides the synthesis of the manifold.<sup>54</sup> Since I can, through pure apperception, become explicitly conscious of any of my representings, it follows that I can, in principle, become explicitly conscious of the representation of unity that guides my synthesis in a given act of material cognition.

Now, most such representations of unity are not candidates for constitutive, formal principles of material cognition, because they are not common to all acts of material cognition. The concept <rattlesnake>, for instance, is a representation of unity that may guide the synthesis of an intuited manifold in a particular act of material cognition (cf. A141/B180.25–30). But most material cognitions (thankfully) do not involve representing a rattlesnake. There is, however, one representation of unity that Kant argues is essential to all synthesis and, thus, to all material cognition. This is the synthetic unity of apperception, which expresses itself in the pure concepts of the understanding – in “the category” (singular), as Kant often puts it.<sup>55</sup> This suggests that a theory of “the category” can be elaborated on the basis of pure apperception, since “the category” is a

By contrast, Kant's doctrine of synthesis and the associated account of material cognition pertain only to a discursive intellect (B135.20–27, B138f.3–14, B145.27–15, B153.7–18, B159.16–20).

<sup>54</sup> A few pages later, Kant offers an example: The concept of a line gives unity to the act of combining a particular spatial manifold into a determinate one-dimensional segment (B137f.16–22).

<sup>55</sup> B131.22, B144.6–7, B145.21, B146.24, A139/B179.15–16, A181/B224.15–17, A253.1–2, 5, A326/B383.20–23, B423n.17–18, A409/B436.5–6, A770/B798.23–25.

representation of unity that is essential to all material cognition. Such a theory would set out constitutive principles governing all material cognition, as such. These principles will be internally represented in any act of material cognition, since they characterize the unity that is represented as the aim of all synthesis. This is why an apperceptive consciousness of those principles would be “pure”: because it would rely only on what is contained in any act of material cognition, as such. And insofar as the synthesis it investigates is an “act of the understanding” (B130.23, B130.17–19), this apperceptively grounded theory of the category will be “essentially concerned not so much with objects as with our manner of cognizing objects” (B25.2–4, cf. A11f.11–13). So what we have described is a transcendental account of the constitutive principles of material cognition – a critique of speculative reason – based on pure apperception, in the sense of analytical reflection upon, and reflective engagement in, concrete acts of judgment about thought-independent objects.

Much the same picture emerges from Kant’s discussion of judgment in the B-Deduction.<sup>56</sup> Judgment is the root of all material cognition, for Kant, because it is only in judging that one claims *objective validity* for one’s thought:

**Judgment** [is] a relation [of representations] that is **objectively valid** and that sufficiently distinguishes itself from the relation of the same representations in which there would be merely subjective validity, e.g. in accordance with laws of association. For in accordance with the latter I can only say, “when I carry a body, I feel a pressure of heaviness”, but not, “it, the body, **is** heavy”, which amounts to saying that these two representations [sc. <body> and <heavy>] are combined in the object [*Objekt*], i.e. regardless of differences in the state of the subject, and not merely compresent [*beisammen*] in perception (however often it [sc. the perception of them together] may be repeated). (B142, cf. B168.33–15)

Truth, Kant says, consists in “the agreement of cognition with its object [Gegenstande]” (A58/B82). Objective validity, by contrast, consists in a *claim* to truth. A judgment not only represents an object, with which it may or may not agree; it also represents *itself* as agreeing with its object. Judgment, for Kant, involves a kind of “taking condition”.<sup>57</sup> Even an association of representations that does in fact agree with the object – as some subjective associations no doubt will – falls short of judgment and,

<sup>56</sup> The picture I sketch here agrees with Aquinas’s view of rational judgment, according to Coope (2013, esp. 8–10); cited by Land (2021 [2018]).

<sup>57</sup> The term “taking condition” is due to Boghossian (2014), but the idea is widespread and can be fleshed out in a variety of ways; see Valaris (2020).

thus, of material cognition. To cognize an object, one must also *take* one's combination of representations to indicate the state of the object. That is, one must represent one's own act of synthesizing the representations as grounded in the object to which they refer. Otherwise one is not really making a *claim* about the object, even if one's mental state contains information about the object and this information is accurate. This means that a judgment, as a vehicle of material cognition, essentially involves a representation of the standard of correctness against which that very act of judgment is to be assessed: namely, the standard of agreeing with its object. For a judgment intrinsically represents itself as *satisfying* this standard. Indeed, representing itself as subject to this standard of correctness is, for Kant, precisely what constitutes a combination of representations as a judgment in the first place: "That is the aim of the connective term [*Verhältniswörtchen*] 'is' in [judgments]: to distinguish the objective unity of given representations from the subjective"<sup>58</sup> (B141F).

Kant takes judgment to essentially involve a representation of the standard of correctness that governs all material cognition. Therefore, since (i) the "I think" must be able to accompany all my representations and (ii) it is constitutive of all material judgment to represent the standard of correctness that characterizes the form of all material cognition, pure apperception must suffice to bring this standard to explicit consciousness. Critique consists in performing just this apperceptive task. In thus spelling out the standard of correctness that is internally represented in every act of potentially knowledgeable judgment, critique elaborates the form (the essence) of material cognition.

Though they are couched in different terms, this account coincides with the previous discussion of combinatory synthesis. Recall, first, that any act of combination presupposes some representation of unity that guides it and, second, that Kant takes "the category" to be a representation of unity that is common to all acts of synthesis concerned with thought-independent objects. This led us to describe critique as a theory of "the category" that can be elaborated on the basis of pure apperception. We then turned to judgment and sketched a view of critique as an apperceptive explication of the constitutive standards of correctness governing all material cognition, based on Kant's view that it is essential

<sup>58</sup> This is why the senses do not judge (A293/B350.18–20). Though acts of sensing are, of course, subject to certain standards of correctness – namely, regarding the senses' proper (healthy) operation – an act of sensing does not, merely as such, involve representing itself as satisfying any such standard. Sensations convey information only about the current state of the sensing subject (cf. A320/B376.26–28).

to all potentially knowledgeable judgment to represent itself as agreeing with its object.

These accounts coincide because Kant conceives of the categories as “concepts of objects in general [*überhaupt*]” (A93/B126.2–5, cf. B128.8–11). Kant puts the same point the other way around when he says that an “**object** is that in the concept of which the manifold of intuition is **united**” (B137.22–02, Kant’s emphasis). The categories are those concepts of the object with which all judgment, as a vehicle of material cognition, essentially takes itself to agree. That is why Kant can say that “the principles of the objective determination of all representations” – that is, the standards of correctness for all material cognition – “are all derived [*abgeleitet*] from the principle of the transcendental unity of apperception” – that is, the unity that finds expression in the categories (B142.21–24). And it is why a theory of the category amounts to an account of the constitutive standards of correctness governing all material cognition – and why such an account can be epistemically based on pure apperception alone.

#### 1.4 Prosecuting a Critique via Pure Apperception

This account of critique raises several questions, two of which are especially pressing for my project in this book. First, the idea that the constitutive principles of material cognition can be known via pure apperception can seem mysterious, in much the way that calling PGL a *Selbsterkenntnis* of reason may seem to generate more heat than light. This worry is not allayed by reassurances about what apperceptive reflection is *not* – for example, that it is not an unmediated *Wesensschau*, not a mystical form of introspection, not an infallible or exhaustive form of self-omniscience. What one wants, in order to assess the textual and philosophical plausibility of the proposal, is a positive account of what it *is* to prosecute a critique of speculative reason via apperception, accompanied by concrete illustrations of such a procedure, with running commentary on how and why it unfolds as it does.

A second pressing question for my project in this book is how pure apperception could yield insight into the character of *sensibility*, even if one grants that it can somehow explicate the form of *the intellect* in its pure material use. After all, the accounts of synthesis and judgment sketched in the previous section suggest that pure apperception only enables a kind of “maker’s knowledge” of my spontaneous intellectual activity. I can apperceive only what I enact. Or, more precisely, I can apperceive only those

representations that guide me in my spontaneous synthetic activity, i.e. the representations that manifest the self-activity of my intellect, the autonomy and self-determination of reason in its material use. So apperception can perhaps yield knowledge of the principles essential to all combination – that is, a theory of the category (or categories). But Kant is explicit that “**combination** [...] can never come into us through the senses and thus also cannot be contained in the pure form of sensible intuition” (B129.14–15). Indeed, Kant appears to hold that sensible representations do not require combination:

The manifold of representations can be given in an intuition that is merely sensible, i.e. nothing but receptivity, and the form of this intuition can lie a priori in our faculty of representation, without being anything other than the manner [Art] in which the subject is affected. (B129.9–13, cf. B132.8–9)

This seems to place sensibility, as a capacity for representation, outside the scope of what we can cognize through pure apperception. And this is reflected in Kant's division of the Doctrine of Elements into (i) a Transcendental Aesthetic, which treats of sensibility insofar as it helps make synthetic a priori knowledge possible, and (ii) a Transcendental Logic (TL), which treats of the intellect insofar as it purports to cognize thought-independent objects a priori. It is therefore tempting to think that what is properly analogous to PGL, as a *Selbsterkenntnis* of reason, is not critique per se but TL to the *exclusion* of Transcendental Aesthetic and the theory of sensible intuition it elaborates.

The chapters that follow attempt to answer these worries. Chapter 2 reconstructs a series of arguments from pure apperception that support conclusions about the nature of receptive intuition. These arguments are located in the Introduction to the *Critique* and establish Kant's *Vorerinnerung* about the two “stems” of human cognition: (i) a receptive capacity for intuition (termed “sensibility”) through which objects are given and (ii) a spontaneous faculty of thought (termed “understanding”) through which they are cognized. Chapter 3 turns to the opening section of the Transcendental Aesthetic (A19–A22/B33–B36) and identifies apperception-based arguments for key elements of Kant's theory of sensible intuition: in particular, his claims that (iii) intuitions, as object-giving representations, relate immediately to their objects and that (iv) any *sensible* intuition must involve not just particular sensible affections but also an affection-independent “form” that enables these to be ordered in determinate relations. These claims are based on an apperceptive grasp on

the nature of our own cognitive powers combined with very general conceptual considerations. It is only with the claim that human intuition is based on *sensible affection* (B33.10–12) that Kant must appeal to a different sort of evidence, such as phenomenological reflection or empirical observation.

These discussions address the first worry by illustrating in detail how critique can be prosecuted via pure apperception. But they do less to allay the second worry, since they only establish *that* we possess a non-intellectual capacity for intuition and that any sort of receptive intuition must have a *pure form*, without delving into the determinate character of such representations or how these might, in virtue of specifiable features, contribute to synthetic, a priori cognition of thought-independent objects. I take up that challenge in Chapters 4 and 5, which present an interpretation of the Metaphysical Expositions at the heart of the Aesthetic. I argue that Metaphysical Exposition is a special sort of conceptual analysis, in which apperceptive reflection plays a central role. The Metaphysical Expositions aim to show that our representations of space and time originate in pure intuition. Kant does this by making claims (i) about the marks of our concepts, <space> and <time>, and (ii) about the essential characteristics of pure intuitions. Apperception plays a leading role in establishing both kinds of claims. The first arise from conceptual analysis as described in Chapter 4; the second from the reflections on our cognitive capacities outlined in Chapters 2 and 3.

This addresses the second worry, by illustrating how specific features (givenness, immediacy, infinity, singularity) can be established as essential to human intuition on the basis of apperceptive reflection. But it is only to be expected that, in the Transcendental Aesthetic and elsewhere, Kant blends together considerations that I want to ascribe to pure apperception with considerations of a phenomenological, metaphysical, mathematical, or even broadly empirical character. The argument from apperception that I aim to trace is not the only sort of argument Kant offers. But it is one that rewards our scrutiny.