

ARTICLE

# Kant's Radicalization of Cartesian Foundationalism: Thought Experiments, Transcendental Arguments, and Level Circularity in the Paralogisms

Murray Miles 

Department of Philosophy, Brock University, St. Catharines, ON, Canada, L2S 3A1  
Corresponding author. E-mail: [mmiles@brocku.ca](mailto:mmiles@brocku.ca)

## Abstract

Kant's critique of rational psychology is a thought experiment that targets no individual or school, but rather the natural tendency of human reason to "hypostatize" the highest intellectual condition of all cognition (the pure 'I think') as though it were itself a cognition of the 'I.' To do so is to violate the very anti-circularity stricture also at work in Kant's better-known transcendental critiques of Locke and Hume. Along with a new type of circularity (level circularity), this article proposes a conception of transcendental arguments different from that of most contemporary debates regarding empiricism, naturalism, and Cartesian foundationalism.

## Résumé

La critique kantienne de la psychologie rationnelle est une expérience de pensée visant ni un individu ni une école, mais une tendance de la raison humaine à « hypostasier » la condition intellectuelle suprême d'une connaissance quelconque (le « Je pense ») en connaissance du « moi ». Cette tendance implique une circularité qui est également la cible des critiques transcendantales bien plus familières qui visent Locke et Hume. De même qu'un nouveau type de cercle (dit « de niveau »), cet article propose une conception des arguments transcendantsaux différente de celle présupposée dans la plupart des débats contemporains sur l'empirisme, le naturalisme et le fondationnalisme cartésien.

**Keywords:** paralogisms; transcendental arguments; circularity; thought experiments; Cartesian foundationalism; naturalism;

*Friedrich-Wilhelm von Herrmann, in memoriam*

## 1. Introduction: Kantian Thought Experiments<sup>1</sup>

The Paralogisms chapter of Kant's first *Critique* is devoted chiefly to exposing the transcendental illusion at work in the fallacious reasoning behind three alleged

<sup>1</sup> Quotations from the *Critique of Pure Reason* are referenced in the customary manner, following the pagination of the original editions of 1781 and 1787, designated 'A' and 'B,' respectively. Translations

© The Author(s), 2022. Published by Cambridge University Press on behalf of the Canadian Philosophical Association/ Publié par Cambridge University Press au nom de l'Association canadienne de philosophie. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

synthetic *a priori* principles asserting the substantiality, simplicity, and numerical identity of the soul.<sup>2</sup> The principles form the conclusions of three syllogisms or dialectical inferences to the soul's absolute unity.<sup>3</sup> While Descartes is expressly identified as the target of a somewhat anomalous fourth paralogism, the target (or targets) of the first three is more controversial.<sup>4</sup> So is the exact nature — whether logical, non-logical, or both — of the fallacies involved, and the precise relations among the key terms 'transcendental illusion,' 'hypostatization,' and 'subreption.' What has gone largely unremarked is the radicalizing critique of Cartesian foundationalism that forms a secondary theme of the chapter.<sup>5</sup>

---

follow, with some emendations, *The Cambridge Edition of the Works of Immanuel Kant* edited by Guyer and Wood. Translations of other works follow the same edition, though references (e.g., 28:259) are to volume and page of the standard German edition, *Kants Gesammelte Schriften*, edited by the Königliche Preußische (later Deutsche) Akademie der Wissenschaften, Georg Reimer (later Walter de Gruyter), 1900–. Square parentheses within quotation marks indicate interpolations by the author, while 'e.a.' after a reference stands for 'emphasis added.'

<sup>2</sup> 'Chiefly' because, even ignoring the presence of a fourth paralogism, the "nature of our thinking being" (A345/B403) occupies only the first part of rational psychology. The second is devoted to the soul's "community" (cf. A384, A390, B427) with the living, organic body, the third to the soul's existence "outside this community (even after death)" (B427). There will be occasional references to these parts of the discipline in what follows.

<sup>3</sup> The categories predicated of the soul are those which "under each heading ... express absolute unity" (A401) or "ground the unity of the remaining ones" (A403): subsistence (as opposed to inherence, under the heading of relation), being simple (in contradistinction to being a real composite of parts, under that of quality), and numerical unity (as distinct from plurality, under the heading of quantity). Kant regularly adds the category of existence (A403, cf. also A402 and B409), although the categories under the fourth heading of his table of categories (modality) hardly mesh seamlessly with the others. Horstmann, following Kant, extends the parallelism to existence as "modaler Sinn der Einheit" (Horstmann, 1993, p. 417, cf. 414).

<sup>4</sup> The tendency among those, but not only those (cf. Düsing, 1987; Klemme, 2009) writing in English has long been to identify Descartes, or Descartes and Leibniz, as Kant's principal target(s). See Strawson (1966, p. 168) on the "*coup de grâce* to Cartesianism," and Bennett's (1974) strongly Descartes-focused interpretation. Kitcher (1982, p. 516, 1990, p. 181), following Bennett (1974, p. 83, note 4), cites Wilson (1974) in favour of including Leibniz, adding "and their followers" almost as an afterthought. Recently, attention has turned sharply to "their followers" (see, e.g., Longuenesse, 2008, p. 20, and the exhaustive, book-length study of the question by Dyck, 2014).

<sup>5</sup> Regarding the nature of the fallacy, Kannisto (2018, p. 194, note 1) provides exact references to the works of Bennett (1974), Ameriks (2000), Van Cleve (1999), Allison (2004), and Grier (2001), all of whom deem the inferences logically valid. Kitcher (1990), Longuenesse (2007), Proops (2010), and Dyck (2014) endorse what the latter calls Kant's "official diagnosis" (Dyck, 2014, p. 100): an (informal) fallacy of equivocation. Kannisto himself asserts the validity *and soundness* of the paralogisms by the canons of Kantian "general" logic, but their invalidity by those of transcendental logic. As for the other, terminological question, Allison, Grier, and Kannisto all follow Kant (cf. A341/B399) in distinguishing logical from transcendental paralogisms and the latter from transcendental illusion. Allison likens a transcendental paralogism to a "category mistake" (Allison, 2004, p. 337), which Kant (lacking the Rylean term) characterizes as that of "hypostatizing" (A395) the pure self by elevating it to the status of an intelligible object. This category (or, better, *level*) mistake lies *between* the inevitable "transcendental illusion" (from which Allison had not distinguished it sharply in the earlier edition of his book) and the perfectly avoidable "substantive metaphysical error" or "subreption" that constitutes the transcendental paralogism proper (Allison, 2004, p. 337). More recently, Dyck has shifted the focus of attention to the error with which Kant allegedly burdens the *Schulphilosophen* in particular: "hypostatizing" the 'I' of the pure 'I think' by raising it to the level of an *empirical* rather than intelligible object. This "original" or "*initial* failing" (Dyck, 2014, pp. 81 and 86, respectively) he distinguishes from the "subsequent error" or "foundational mistake" (Dyck, pp. 86 and 89, respectively) variously analyzed by Allison, Grier, and Kannisto. Both level mistakes (which have

In the sense intended here, the word 'radicalizing' refers to Kant's attempt to 'get beneath' that which Descartes and his followers considered the ultimate or fundamental sources ('roots') of epistemic justification. In this sense, Descartes himself had already radicalized Scholastic-Aristotelian foundationalism by 'getting beneath' its founding dimension (ordinary sense perception) and grounding all cognition of outer things upon inner cognition of the empirical *ego cogitans* (see Section 2). Kant's critique, however, is more radical. From his new, transcendental standpoint, *any* attempt to locate the ultimate ground of all object-level cognition in *a* (or *some*) empirical (or even pure) cognition *belonging to that very level* is circular. In positive terms: the foundational dimension of any *non-circular* justification of *all* first-order cognition of objects *as such* must be sought 'lower down' (or, varying the metaphor to fit Kant's project, 'higher up'), in meta-level reflection upon the knowing subject and the conditions of the possibility of *any* object-level cognition (inner as well as outer, either empirical or *a priori*) *whatsoever*. So put, the target of Kant's critique would appear to be identical to that of the fourth paralogism: the Descartes of the Second Meditation. But it applies equally to almost the entire early modern tradition on both sides of the channel, which had absorbed Descartes's foundationalism (with or without his rationalism) almost 'atmospherically,' so to speak, much as it would have been imbibed by the pre-critical Kant himself.

Breaking with Cartesian foundationalism, the mature Kant conceived and, in the Transcendental Analytic, executed a novel *transcendental* foundationalist project designed to 'get beneath' the empirical *cogito, ergo sum* that had been *the* first principle of Descartes and at least *a* first principle for all those who can justly be considered his heirs. In the Paralogisms chapter of the Transcendental Dialectic, moreover, he conducted a bold *thought experiment* designed to show that attempts to ground *all* first-order cognition in *some* first-order cognition commit a level mistake that renders them viciously circular. The thought experiment posits two *hypothetical* forms of foundationalism that neither existed in Kant's day nor had ever existed in the past. The first is presented in the guise of a *mixed* (both pure and empirical) rational psychology that circularly misconstrues the *Ich denke* of transcendental apperception, which is the highest intellectual condition of *all* cognition of objects as appearances, as though it were itself *a* cognition of an (inner) object or appearance — specifically, an inner *perception* of the *empirical* self. This violates the transcendental *anti-circularity stricture* formulated at A402 as: "I cannot cognize as itself an object that which I must presuppose in order to cognize an object at all."<sup>6</sup> Now, as Kant surely realized, neither the historical Descartes, nor anyone else, had had the least conception of an *a priori* transcendental condition of the possibility of all

---

not previously been designated as such in the literature) will be discussed in more detail in Section 4. Since the critique of Cartesian foundationalism concerns only the error identified by Dyck, no detailed consideration of either the fallacy or the terminological question is required here.

<sup>6</sup> On the expression 'anti-circularity stricture,' coined by Van Cleve, see Section 3 below. The important distinction between completely pure and mixed rational psychology was entirely overlooked until brought to light by Dyck, who, however, both exaggerates the role of the mixed variety of rational psychology and mistakenly identifies it with that the *Schulphilosophen* alone. See note 12 below.

empirical cognition *überhaupt* prior to his own *Copernican thought experiment*.<sup>7</sup> This ‘refutation’ would therefore be entirely without force if it targeted some particular historical figure or figures rather than the hypothetical discipline imagined in the first part of Kant’s thought experiment. Viewed as a thought experiment, however, the anti-circularity stricture turns out to be the operative principle in a *transcendental argument* that still represents a serious challenge to both empiricist (naturalist) and rationalist efforts to ground *all* first-order cognition of objects upon *some* (empirical or *a priori*) first-order cognitions of objects. It hardly matters whether the candidate explanatory or justificatory ultimates are sense-data, impressions, sensations, ideas (even *innate* ideas), or inner and outer perception; from the transcendental perspective, *any* such explanation or justification is circular. Whether the only way to ground first-order cognition *non-circularly* is through second-order reflection on the transcendental subject and its *a priori* intuitions and concepts is, of course, quite another matter. Here Kant’s procedure in the *Critique* betrays a Cartesian bias that held sway in modern philosophy at least until Husserl, and that was only seriously challenged by Heidegger (see note 38 below).<sup>8</sup>

The second and *main* form of foundationalism imagined by Kant in a further (or as part of the same, extended) thought experiment is explicitly said to be a merely *hypothetical*, or as Kant himself puts it, a “putative (*angebliche*)” (A342/B400) *completely pure* science of the soul based exclusively on the *a priori* “concept — or rather, if one prefers — the judgment **I think**” (A341/B399).<sup>9</sup> As Kant well realized,

<sup>7</sup> It is well known that Kant never used the phrase ‘Copernican revolution,’ but he does speak of trying (*versuchen*) “whether we get farther with the problems of metaphysics by assuming that the objects must conform to our cognition” (Bxvi), describing this *Versuch* (Bxxi and Bxviii), “which resembles the first thoughts of Copernicus” (Bxvi e.a.), as an *Experiment der reinen Vernunft* (Bxxi n.). In three notes (B xviii, B xxi, and B xxii), the last of which contains the only other mention of Copernicus in the Preface, Kant describes his experiment in terms of the transcendental distinction between appearances and things in themselves. There is thus ample textual warrant for speaking of an overarching Copernican thought experiment of which that of the Paralogisms chapter forms part. On the way in which the transcendental distinction is deployed in the Paralogisms to uncover three level mistakes and two forms of level circularity (corresponding to the two types of rational psychology distinguished by Dyck), see Section 4.

<sup>8</sup> By grounding *all* object-level cognition *as such* upon a meta-level *a priori* cognition of the *transcendental* self or subject, Kant transcendentalized the *ontological* principle that Leibniz formulated as: “The reasons for the world ... lie in something extramundane, different from the chain of states or series of things whose aggregate constitutes the world” (G VII 302). Along with the impossibility of a contingent (natural) *origin* of the universe, Leibniz had already insisted on the circularity (infinite regress) involved in naturalistic *explanations*. Against the naturalist “system” of Strato of Lampsacus and *les Stratoniciens* (cf. G VI 228), he maintained that the sufficient reason for the existence of the universe cannot be found in the series of contingent things ... [W]e will not make any progress in this way, however far back we go, for the same question always remains. Thus the sufficient reason, which needs no other reason, must be outside this series of contingent things, and must be found in a substance which is its cause, and which is a necessary being, carrying the reason of its existence with itself. Otherwise, we would not yet have a sufficient reason, *where one could end the series*. (G VI 602 e.a.)

Kant’s substitution of a *transcendental* for Leibniz’s *transcendent* onto-epistemological anti-circularity stricture is the subject of Section 3.

<sup>9</sup> Kant speaks of a “putative (*angebliche*) science” of the soul in two senses. First, it is a *supposed* science, that is, a *pseudo-science* that can only pretend (*angeben*) to be genuinely scientific or demonstrative; this “imagined (*eingebildete*)” (A395) or “would-be (*scheinbare*)” (A397) science rests on “four paralogisms ... which are falsely held to be a science of pure reason about the nature of our thinking being” (A346/B404).

no science of the soul that understood itself in quite this way existed or had ever existed.<sup>10</sup> The point of this further thought experiment is rather that universal human reason is inclined (though not necessitated) to develop such a completely pure doctrine of the soul, the ratiocinative faculty of the mind being naturally constrained (as formal logic shows) to proceed ever higher in search of the unconditioned to any given series of conditions (cf. A322–323/B379). Whereas the *Ich denke* is no more than the highest *intellectual condition* of all cognition of appearances, the human *ratiocinative* faculty, which “relates itself only to the use of the understanding,” aims at “a certain unity of which the understanding has no concept” (A326/B383). The main target of the Paralogisms is this natural propensity of human reason to misconstrue the *Ich* of the *Ich denke* as a purely intelligible or noumenal object that can be cognized in an absolutely unconditioned manner through mere categories.<sup>11</sup> For, as Kant points out, “[a]pperception is itself the ground of the possibility of the categories” (A401). Thus, the completely pure rational psychology of the second and main part of Kant’s thought experiment circularly misconstrues the pure *Ich denke*, “which grounds the categories” (A421), as though it were itself an *a priori* cognition of an intelligible object (the pure *Ich*), applying to it the very categories (‘substance,’ ‘simplicity,’ etc.) that *it* makes possible in the first place. This grounding in the absolutely unconditioned is not just a transcendental misuse of those categories whose sole legitimate use the Analytic had shown to be empirical; it constitutes a further violation of the above-cited anti-circularity stricture. This time, however, the alleged object is purely

---

Second, it is a *suppositious* or hypothetical science, and not an *actual* doctrine, or set of doctrines and arguments, of any historical figure or figures in the recent or remoter past, least of all Descartes (see next note).

<sup>10</sup> Kant states repeatedly that Descartes’s ‘I think’ is “an empirical proposition” (B422 n.). Cf. also B420, B421, B428, and the minute, clause-by-clause dissection of B422 n. by Kim (2019). Leibniz called his own reformulation of the Cartesian *cogito* as “*I think various thoughts*” a primary *truth of fact* (Leibniz, 1996, p. 367, cf. p. 411). Neither they (nor any of their followers) championed that *completely pure* science of the soul evoked in the second part of Kant’s thought experiment. True, *sentences* identical to the fallaciously derived synthetic *a priori* conclusions of the syllogisms examined by Kant can be found in Descartes (and many others). But even where identically worded, the *propositions* such sentences express are merely *analytic* for Kant (see notes 27 and 32 below) and do not belong to rational psychology proper. And as nobody had ever taken the *Ich denke* for a completely *a priori* cognition of the ‘I’ as an intelligible object, so no one had confused the *a priori* condition of any empirical cognition with an empirical cognition of the ‘I’ either. It is thus hardly surprising that Kitcher should suggest that Kant is issuing “*caveats* about his own doctrine of apperception” (Kitcher, 1990, p. 192, cf. p. 198) rather than “criticisms of his predecessors.”

<sup>11</sup> That the aim of the first three paralogisms is to hold in check a natural tendency of human reason rather than criticize any of his predecessors is apparent where Kant first introduces the term ‘paralogism.’ He will have to do, he states, with “sophistries not of human beings, but of pure reason itself” (A339/B397). “A fallacy of this kind will have its ground in the nature of human reason” (A341/B399). This fits well with the stated aim of the whole Transcendental Dialectic as “a true cathartic” (A486/B514) of *reason* conducted by *reason* itself, and with the remark, at the very beginning of the *Critique*, that the work is not “a critique of books and systems, but a critique of the faculty of reason itself” (A xii). Having discovered the *original* unity of apperception as the highest intellectual condition of empirical cognition of objects at an earlier stage in its journey toward self-knowledge (the Transcendental Deduction), human reason pursues its course in the Paralogisms by giving an exacting account of just how sterile this discovery is for the attainment of even a single synthetic *a priori* principle, and *a fortiori* for a purely theoretical cognition of those matters (see note 2 above on the third part of rational psychology) that are the ultimate ends of rational psychology. Rather than “a caution about his own doctrines” (Kitcher, 1990, p. 192), the Paralogisms present Kant as the spokesman of human-reason-reflecting-on-itself.

*intelligible*, and not that empirical object of inner *perception* of which mixed rational psychology purports to have both empirical and a *priori* cognition.

Only this second part of Kant's thought experiment (which occurs first and remains his dominant concern throughout the chapter) involves a dialectical inference to "the absolute (unconditioned) **unity** of the **thinking subject**" (A334/B392), since only such a completely pure rational psychology has its origin in pure reason's requirement that the cognition of a series of conditioned conditions be grounded in that of the unconditioned. By contrast, the error of that *mixed* rational psychology imagined in the first part of the thought experiment stems from mistaking the 'I' of the pure *Ich denke*, which is the supreme *condition* of all cognition, for a *cognition* of the self of *empirical* apperception. Since there is nothing dialectical about this, it is hardly surprising that Kant's radicalizing critique of Cartesian foundationalism has been eclipsed by the dialectical arguments of the four paralogisms (especially the fourth) in the scholarly literature on Kant's relationship to Descartes. It is nevertheless quite as important as the fourth paralogism both historically and in a systematic perspective.<sup>12</sup>

Its obvious historical significance aside, the first part of Kant's thought experiment epitomizes what he himself would most readily have recognized as a 'transcendental argument.' Today, that term is used almost exclusively to describe a type of argument whose prototype is the famous Refutation of Idealism, but which has little to do with the technical sense of 'transcendental' in Kant: "occupied not so much with objects but rather with our mode of cognition of objects insofar as this is to be possible *a priori*" (B25, cf. A11). In Section 6, it will be argued that the radicalizing critique of Cartesian foundationalism is a transcendental argument as Kant defines the term and that this general type of argument has an important bearing on contemporary philosophical debates regarding empiricism (or naturalism), rationalism, and foundationalism — debates in which transcendental arguments modelled on the Refutation have long occupied a pre-eminent place.

---

<sup>12</sup> Dyck clouds the issue of Kant's historical target(s) by identifying "pure" with "the narrowly rationalistic psychology of, for instance, Descartes and Leibniz" (Dyck, 2014, pp. 226–227). From the "pure" he distinguishes the "mixed" variety of the Wolffians, who assigned much greater "importance ... to experience in the rational investigation of the soul" (Dyck, 2014, p. 9). But, on Dyck's own showing, the Cartesian *cogito* and *sum* are empirical propositions; of this, not only Descartes himself, but Leibniz (Dyck, 2014, p. 181; also p. 234), Wolff (Dyck, 2014, p. 178), Baumgarten (Dyck, 2014, p. 179), and Kant (Dyck, 2014, p. 183) were all quite as well aware as Dyck shows himself to be. His extended argument that the mixed rational psychology of the Wolffians sheds valuable light on the specific detail of the individual paralogisms can stand on its own considerable merits; but it is mistaken to suggest that (i) the Wolffians (and *not* Descartes and Leibniz) and that (ii) mixed (and *not* pure) rational psychology are the best answers to the (i) Whom? and (ii) What? questions. It is not just the historical fact that *all* the above figures are proponents of a *mixed* rational psychology that Dyck overlooks; it is also the fact that the *main* focus of the Paralogisms is upon a (likewise non-existent) completely pure science of the soul. On this latter point, see Kannisto, who endorses Dyck's distinction between two kinds of rational psychology (without comment on his historical attributions), but rightly insists "that pure rational psychology is Kant's main target in the Paralogisms" (Kannisto, 2018, p. 224). The correct answers to the (i) Whom? and (ii) What? questions are thus: (i) *no* historical figure or figures (so too Kitcher, but also Horstmann, 1993, p. 416) and (ii) *no existing* type of rational psychology, but rather a merely hypothetical discipline that mistakes (a) the highest intellectual *condition* of all empirical cognition *überhaupt* for (b) a cognition of an intelligible object. The critique of that mixed rational psychology (or Cartesian foundationalism) that mistakes (a) for a cognition of an *empirical* object is only a secondary theme of the Paralogisms.

The procedure will be as follows. After an initial survey of the ramified *family* of critiques of Cartesian foundationalism, including those nowadays called 'transcendental arguments' (Section 2), the *epistemological* anti-circularity stricture of the Paralogisms will be contrasted with its better-known *ontological* counterpart in rational theology (Section 3). Following consideration of three different *level mistakes*, two of which amount to *level circularity* (Section 4), the textual basis of the re-interpretation of the Paralogisms as a two-part thought experiment will be presented (Section 5). A rapid glance at Kant's use of similar arguments against empiricist accounts of space and causality will confirm that Kant's epistemological anti-circularity stricture, while lacking in strictly refutative force outside his own two-part thought experiment, still has a certain weight when it comes to the systematic question of whether to naturalize or, alternatively, transcendentalize, foundationalist epistemologies of a Cartesian stamp (Section 6). A brief conclusion follows (Section 7).

## 2. Foundationalist and Other Critiques of Cartesian Foundationalism

Criticism of a foundationalist epistemology like Descartes's generally takes one of two forms: either it advances (1) an alternative, more radical theory of the same general kind — an *alternative foundationalism* — or it proposes (2) an anti-foundationalist *alternative to foundationalism*.<sup>13</sup> Coherentist theories of justification directly or indirectly inspired by 'naturalized' epistemology (cf. Quine, 1953, 1969) are anti-foundationalist critiques of the latter sort; so are those transcendental arguments modelled on the Refutation of Idealism and/or the fourth paralogism that today pass for transcendental arguments *tout court* (cf. Stroud, 1968, 1979). The Paralogisms, by contrast, form part of an alternative foundationalist strategy aimed at radicalizing traditional Cartesian foundationalism by exposing its circularity.<sup>14</sup>

The single most striking example of (1) in the history of philosophy is Descartes's own reversal of what had long been the established order of knowing. With his *mens notior corpore* doctrine, succinctly formulated in the *Principles* as "our knowledge of our thought is prior to, and more certain than, our knowledge of any corporeal thing" (AT VIII-1 7: CSM I 195), Descartes gave *self-knowledge*, as expressed in empirical,

<sup>13</sup> To adapt a distinction drawn by Allison, who argues that Kant's transcendental idealism should not be regarded as an "an alternative ontology," but as "an alternative to ontology" (cf. Allison, 2006, p. 123, 2004, p. 98). However that particular question may be decided, transcendental idealism is clearly an alternative foundationalism and an extreme form of anti-naturalism that seeks to ground all empirical in pure, and all pure cognition of objects in pure transcendental cognition of the subject.

<sup>14</sup> While (i) "so-called 'transcendental arguments'" and (ii) Quinean naturalized epistemology remain the chief *alternatives to* "a straightforward Cartesian or 'foundationalist' theory of knowledge" (Stroud, 1979, p. 277), efforts to supplant Cartesian with an *alternative foundationalism* have received less attention. Note 17 below distinguishes a number of different foundationalisms. To Stroud's simple binary categorization, Strawson adds (iii) arguments to "the best available explanation of the phenomena of experience" (Strawson, 1985, p. 20) as a further possible response to Cartesian scepticism. See Stroud (1968) for the two main sub-varieties of transcendental arguments (nicely summarized by Strawson, 1985, p. 9), and Bennett (1974, pp. 66–69) for a spate of others. By far the most exhaustive classification of types of transcendental arguments and of the types of scepticism against which they are and are not effective is that of Stern (2000). On the varieties of naturalism, see Strawson (1985, *passim*).

contingent propositions (the various values of the variable *cogito*, 'I am perceiving,' 'I am imagining,' etc.), an absolute priority over cognition of the external world present to the five bodily senses. Such cognition of material things as can be reliably obtained must be derived by inference from non-inferential cognition of the self, using logical and/or metaphysical "axioms," "common notions," "eternal truths," or "principles of the natural light."<sup>15</sup> In the Aristotelian tradition, by contrast, knowledge of the self had always been regarded as requiring a special act of consciously directing one's attention *away* from sensible things (the first and most reliably known objects), a *bending back* (*reflectere*) of one's cognitive gaze upon one's actually occurring cognitive acts. As thus dependent upon cognition of ordinary sensible things, empirical cognition of the self is 'later' in the order of knowing. Yet though 'earlier,' cognition via the outer senses had never been considered a source of *principles* having the same degree of certainty as the universal, necessary, and hence absolutely first *non-existent* truths of logic and/or metaphysics. Accordingly, Descartes's (and later Leibniz's) elevation of the contingent existential propositions *cogito* and *sum* to the status of absolutely certain first *principles* was revolutionary: in redefining 'absolute certainty' *operationally* as 'the ability to withstand the illusion, dreaming, and deceiving God arguments of the First Meditation' (rather than in terms of universality or necessity), Descartes re-drew the extensional boundaries of 'principle' so as to make room for a pair of contingent propositions alongside the necessary principles of the natural light.<sup>16</sup>

The Cartesian paradigm confirms that the shared trait of all radicalizing critiques of pre-existing foundationalisms is the effort to 'go back behind' those alleged first truths whose greater reliability serves to render all other cognitions more certain. It is true that at least one 20<sup>th</sup> century form of foundationalism took the basic truths to be empirical propositions *about bodies* (the so-called *Protokoll-Sätze* of 20<sup>th</sup> century physicalists); but the influence of Descartes has been such that philosophers have generally looked to empirical first truths about the *mental*, including the sense-data of the phenomenologists, for foundational or first truths endowed with existential import. Seen in this wider context, the most novel thing about Kant's foundationalism is his insistence that the supreme or ultimate principles *must* be meta-level cognitions of the conditions of the possibility of both empirical and pure object-level cognitions as such, and not themselves again cognitions of any either empirically (phenomenally) or absolutely (noumenally) *real* object, not even of the transcendental subject *as an object*. Behind this lies the conviction that *any* attempt to found *all* first-order cognition *as such* upon that which is itself just *an item* (or kind) of first-order cognition is viciously circular. The aim of Kant's *transcendental* foundationalism is thus not just to reach 'rock-bottom'; if circularity is to be avoided, *first or founding* principles must be *second-order* cognitions about the transcendental subject

<sup>15</sup> On the equivalence of these technical locutions, see my lexicon entry 'common notions' in Nolan, 2016, pp. 138–140).

<sup>16</sup> The point made by Curley (1978, pp. 172–175), among others, that Descartes's *mens notior corpore* doctrine (like the *cogito* itself) had important precedents in Augustine and Montaigne, hardly diminishes his stature as an innovator who effected a radical reversal of the traditional order of knowing. On the Aristotelian order of knowing, see EN 1170a 29–b 1, *Met.*, XII, 1074b 35–36, *Parva Naturalia*, 448a25f. and *De An.* 425b12f.

and the possibility of all first-order *a priori* cognition of objects (whether metaphysical, mathematical, or natural scientific) *überhaupt*. Thus, Kant not only locates the founding dimension of human knowledge 'further back' (or 'higher up'); he does so in view of a kind of circularity that no one before him had ever even glimpsed. This emerges most clearly in the Paralogisms.<sup>17</sup>

### 3. Ontological and Epistemological Anti-Circularity Strictures

Kant's neglected transcendental critique of Cartesian foundationalism has nothing to do with the familiar Cartesian circle (or circles) to be described presently; it is also a different *type* of transcendental argument from that exemplified by the Refutation of Idealism. Both points can be clarified by contrasting Kant's transcendental anti-circularity stricture with its better-known ontological counterpart (see note 8).

In a qualified defence of the point of view of "cosmological arguers," Van Cleve formulates what he calls Kant's "counterprinciple" (Van Cleve, 1999, p. 206) to a principle defended by Hume in the *Dialogues Concerning Natural Religion*.<sup>18</sup> Hume's principle states (in Van Cleve's terms): "the existence of a totality is always adequately explained when the existence of each member is explained" (Van Cleve, 1999, p. 205). The juxtaposition of Kant's counter-principle to a *corollary* of this yields Kant's ontological "*anticircularity stricture*" (Van Cleve, 1999, p. 206). The corollary affirms that "if there is an infinite totality of objects or events each of which is explained by the causal efficacy of some other member(s) of the totality, then the existence of the whole totality is thereby adequately explained — no recourse to anything outside the totality is needed" (Van Cleve, 1999, p. 205). The operative terms here are obviously 'infinite' and 'outside,' for what Hume rightly affirms of finite classes — that there is no need to go outside them in order to furnish a complete explanation of everything contained in them — no longer holds when the *explanans* is the existence of the infinite class of all things of a certain kind (say, all contingent things in

<sup>17</sup>To the varieties of anti-foundationalism (see note 14) compare some of the forms an alternative foundationalism may take. Not all are strictly *radicalizing* attempts to 'get beneath' those first truths in which an earlier foundationalism saw the 'roots' or 'seeds' of all cognition; of those mentioned already, physicalism is not, while phenomenism is, since (in its sense-data elaboration, at least) it locates the foundational propositions at an even lower level than outer and inner perception. Developing Descartes's famous metaphor of a tree whose roots are metaphysics, whose trunk is physics, and whose branches are all the special sciences, Heidegger asserts the need to go back into the very ground from which the roots receive their nourishment (Heidegger, 1998, p. 277). For the later Heidegger (the above-referenced essay dates from 1949), it was no longer a question, as it had been for his teacher, of providing something like an "absolute grounding of science" (Husserl, 1960, p. 7 e.a.). Insisting on the scientific inspiration behind his attempt to radicalize the Cartesian *ego cogito* through a phenomenological return ("reduction") to the primordial sphere of "the transcendental Ego" (Husserl, 1960, p. 25), Husserl tacitly acknowledged his indebtedness to Kant.

<sup>18</sup>Van Cleve quotes from Section IX of the *Dialogues*, where Hume attacks the cosmological argument. As he summarizes the argument:

The proponent of the cosmological argument allows that each member of the set of contingent beings is caused by some member of the set that existed earlier. In that case, claims Hume, it is unreasonable to seek any further explanation for the series as a whole. (Van Cleve, 1999, p. 205)

general — or *überhaupt*, to use Kant's term). Van Cleve formulates Kant's anti-circularity stricture this way: "The existence of Fs in general (or the fact that Fness is instantiated) can be explained only by appeal to the existence of something that is *not* an F" (Van Cleve, 1999, p. 206). Elsewhere he substitutes the expression "*at all*" (e.g., Why are there any Fs at all?) for Kant's *überhaupt*. The reason that such explanations are circular is simply that the "explanans invokes the existence of Fs, but Fs are the very beings for whose existence an explanation is sought" (Van Cleve, 1999, p. 205). In the simplest terms, the *explanans* itself forms part of the *explanandum*.<sup>19</sup>

Van Cleve's anti-circularity stricture is straightforwardly *ontological*; it concerns the possible *existence* of an infinite class of all contingent entities without something that exists necessarily. What motivated Kant's transition from (his own pre-critical) Cartesian to a new, transcendental foundationalism was an anti-circularity stricture that ruled out attempts to justify — or even just to explain — the *cognition* of all objects of a certain kind (all *x*'s *überhaupt*) by recourse to a first or absolutely certain cognition of an object of that very kind (*an x*). Van Cleve's anti-circularity stricture can be generalized to include cognition in the following way: if *y* is to serve as the justificatory or explanatory *ground* of some class of *x*'s,  $\{x_1, x_2, x_3 \dots x_n\}$  (where *n* is some finite number), then, while *y* cannot (without circularity) be a member of *that very* class, it may still be some other *x* *outside* the class, for instance, a 'first' *x* that exists prior to all others, or an *x* that is more certain than the rest, depending on whether what is in question is the efficient cause of the existence of things or the explanatory and/or justificatory ground of cognitions. Thus, whether Descartes's manner of executing the foundationalist project of the *Meditations* is *premise*-circular depends on whether he uses his truth rule ('whatever I perceive clearly and distinctly is true') as a premise in the proof of God's existence and veracity, while using God's existence and veracity to establish the certainty of his truth rule; and whether his procedure is *rule*-circular depends on whether he first *deduces* that God exists and then uses that conclusion to guarantee the reliability of deduction.<sup>20</sup> Finally, whether it is *faculty*-circular depends on whether he first doubts his reason and then attempts to reason his way out of doubt. However, there is nothing either *premise*- or *rule*- or *faculty*-circular about Descartes's *project* of placing the cognitions of God and the world on the secure foundation of the cognition of the self. Only where the *justificandum* is the *infinite* set of all cognitions of any object

<sup>19</sup> The context being rational theology, Van Cleve concludes that "the existence of contingent beings in general can be explained, if at all, only be reference to a *noncontingent* being," which, he adds, "may or may not be God." It follows that "cosmological arguers have always been right to maintain that we can explain everything that needs explaining only if there is a necessary being" (Van Cleve, 1999, p. 206); some, however, have been wrong in concluding (without further argument) that there must be some *one* being that is necessary, namely God. This is precisely the quantifier shift fallacy that Leibniz presses against Locke's cosmological argument in the *New Essays* (Leibniz, 1996, p. 436). Kant argues that the real work of the cosmological argument for God's existence *a contingentia mundi* is done by the ontological argument; but he takes Leibniz's (and Van Cleve's) point: the arguments, he writes, lead us "only to the concept of a necessary being, but not so as to establish this concept in any determinate thing" (A607/B635). The proof that the non-contingent being is God derives from a surreptitious use of the ontological argument.

<sup>20</sup> The distinction between *premise*- and *rule*-circularity is usually attributed to Braithwaite (cf. 1955, pp. 274–278).

whatsoever (*überhaupt*) does it become viciously circular for the *justificans* to be itself a cognition (or set of cognitions) of an object, since it then forms part of the *justificandum*. This type of *level circularity* has to do with the two level mistakes touched on earlier (see note 5) and to be considered further in the next section.

#### 4. Transcendental Distinction, Level Mistakes, and Level Circularity

The level mistakes noted earlier correspond to two ways of “hypostatizing” (A395) the ‘I’ of transcendental apperception. The (a) *Ich* of the *Ich denke* may be hypostatized either as (b) an appearance (an object given in empirical intuition), as in the first part of Kant’s thought experiment; or it may be hypostatized as (c) a thing in itself (an intelligible object), as in the second part. Both level mistakes violate the anti-circularity stricture of A402 by taking what is an *a priori* condition of *all* cognition *überhaupt* for a cognition of an object; and both differ from the much more prominent level mistake relentlessly exposed throughout the Transcendental Aesthetic and Analytic, namely hypostatizing (b) appearances, which are at bottom “representations” for Kant, as though they were (c) things in themselves. While Kant understands his “transcendental distinction” (A45/B62) as a simple binary distinction between (b) and (c), certain of the *uses* to which he puts it (notably in the Paralogisms) require a distinction among three elements: hypostatizing (b) appearances by raising them to the level of (c) things in themselves, which is non-circular; hypostatizing (a) the *Ich* of the *Ich denke* by elevating it to the level of (b) an appearance (the empirical *ego cogito* of Descartes), which is circular; and hypostatizing (a) the *Ich* by raising it to the level of (c) a thing in itself, which is again circular. The first level mistake is that which Kant regarded as historically actual (indeed pervasive in the transcendental realist tradition), while the other two, on whose circularity he focuses in the Paralogisms, are merely hypothetical and belong to the two parts of his thought experiment.<sup>21</sup>

Within the Paralogisms, the level mistake on which Kant insists throughout the Transcendental Aesthetic and Analytic figures most prominently in the “Observation” (A381–396) devoted to the “three dialectical questions” (A384) regarding the mind-body relationship that form the subject-matter of the second and third parts of rational psychology (see note 2 above). Given its subject-matter, the “Observation” provides Kant with an occasion to reiterate the teaching of Analytic regarding *bodies*, even though bodies lie outside the scope of the first part of rational psychology (cf. note 2 again), to which the bulk of the Paralogisms are devoted. According to the Analytic, bodies are only outer *appearances* and hence mere representations; yet in the debate regarding the mind-body relation, Kant notes, “one

<sup>21</sup> Hypostatizing is, in general, “making thoughts into things” (A395), that is, into “real object[s] outside the thinking subject” (A384, cf. also A386 and A695/B723 n.). Depending upon the thoughts (whether *a priori* conditions or empirical cognitions) and on the things in question (appearances or things in themselves), three forms of hypostatizing or level mistake have been distinguished. Note that in the present section, the second part of the Kant’s thought experiment, dealing with the hypostatization committed in a hypothetical completely pure rational psychology, will be discussed first. The next section will revert to the earlier order, interpreting the text passage devoted to a (likewise hypothetical) mixed rational psychology first, even though, as already noted, it appears only toward the end of the Paralogisms.

hypostatizes what exists merely in thoughts, and — assuming it to be a real object outside the thinking subject — takes the same quality, namely extension, which is nothing but appearance, for a property of external things subsisting even apart from sensibility” (A384). Later, he returns to this tendency to “hypostatize outer appearances, no longer relating them to our thinking subject as representations, but rather relating them to it **in the same quality as they are in us as things external to us and subsisting by themselves**” (A386). The words “relating them to it **in the same quality**” pick out what, for Kant, is the *root* error of all “transcendental realism” (A369): hypostatizing things as appearances that are “**outside us**” in the *empirical* sense (cf. A373 on the empirical and transcendental senses of ‘outside’) as though they possessed *the very same spatio-temporal and categorial features or qualities* that they have as appearances (Kant mentions extension and motion) even when considered as things existing outside “our thinking subject” (A386) in the *transcendental* sense of ‘outside.’ According to the “Observation,” this error is the source “of all the difficulties that one believes one finds in these questions” (A384) regarding the mind-body relation during and after life; but while it may bedevil all three main theories (interactionism, occasionalism, and pre-established harmony), the transcendental realist’s “hypostatizing” of (b) outer appearances as though they were (c) things in themselves does not involve circularity.

Circularity arises from hypostatizing (a) rather than (b) in either of the two ways considered in Kant’s thought experiment. The error that is Kant’s main focus in the Paralogisms is that of *circularly* elevating (a) the supreme intellectual *condition* of all cognition through the categories, the ‘I’ of the ‘I think,’ to the level of (c) an *intelligible* object of purely intellectual cognitions *in which those very categories* occupy the predicate position. This, in Kant’s words, is to mistake the (a) “merely subjective condition [‘on which every experience depends and which precedes it’]” (A354) for (c) “a **concept** of a thinking being in general” (A354) or “a thinking being *in itself*” (A360 e.a.).<sup>22</sup> It is this unexampled circle that occupies Kant in the second part of his thought experiment. Where he speaks of “making thoughts into things and hypostatizing them” for the third and last time in the Paralogisms, his focus has shifted from that of the “Observation” (regarding the mind-body relation) to this other level mistake whose blatant circularity vitiates the whole first part (concerned with the *nature of a thinking being*) of that “imagined science” (A395) of the soul which is completely pure. Thus, “everyone either presumes to know something about [purely intellectual, noumenal] objects about which no human being has any concept” (A395), as when outer (b) appearances are hypostatized as (c) things in themselves in those parts of this science dealing with the mind-body relation; “or else makes his own representations into [intellectual] objects” (A395), as when (a) the pure *Ich denke* is hypostatized as (c) a thing in itself in the part of rational psychology devoted to the nature of the soul. Either way, Kant notes, one only “goes round and round in an eternal circle of ambiguities and contradiction” (A395). But while, strictly speaking, no circle (but only a level mistake) is involved in baselessly

<sup>22</sup> Speaking of (a), Kant also uses the expressions “transcendental subject of thoughts = *x*” (A346/B404), “transcendental subject” (A350), “Something in general (a transcendental subject)” (A355), a “mere Something” (A355), all of which refer to (a) “the transcendental object of inner sense” (A361) which is *hypostatized* as (c) a thing in itself, noumenon, or intelligible object.

presuming to know the human *body* as it is in itself, Kant's anti-circularity stricture is violated when (a) the 'I' of the *Ich denke* is elevated to the level of (c) an object of a purely intellectual cognition through the very categories that it makes possible.<sup>23</sup>

Less noticed than either the first or second is the different level mistake ascribed to the mixed rational psychology of Descartes and his followers in the first part of Kant's thought experiment. This third level mistake involves "hypostatizing" that which is only (a) the subjective and necessary intellectual condition of all cognition of appearances whatever (the *Ich denke*) as though it were *itself* (b) a cognition of the 'I' as an appearance. Like the second, then, the third level mistake concerns (a) the *Ich denke* as the "sheer spontaneity of combining the manifold of a merely possible intuition" (B428) which is the highest intellectual condition of all appearances whatsoever; this time, however, the 'I' of the 'I think' is hypostatized as though it were (b) "the subject of consciousness *as an appearance*" (B428 e.a.), that is, an *empirical* object that is not "outside us" at all, not even in the empirical sense of A373, but 'in us' as the object of inner perception or the *empirical* self. For that which is "obviously not an experience, but rather the form of apperception on which every experience depends" (A354), namely (a) "the formal proposition of apperception, **I think**," (A354), cannot (without circularity) be said to be (b) an experience of an object (the Cartesian *ego cogitans*), that is, *an instance of the very experience that it makes possible*.

Now according to the doctrine of transcendental idealism, to represent myself via the 'I think' of pure apperception is to "represent myself to myself *neither* as I am [in myself] *nor* as I appear to myself" (B429 e.a.). Despite Kant's frequent insistence that the Paralogisms are wholly devoted to a completely pure rational psychology, these words make it clear that he is *also* targeting that level mistake which he *supposes* to be the original sin of a mixed rational psychology of the type inaugurated by Cartesian foundationalism. But since neither the historical Descartes nor his rationalist and empiricist successors can be said to have hypostatized, as though it were an inner appearance, a pure or transcendental 'I think' of which they had not the slightest conception, Kant's critique only serves to highlight the fact that all historical forms of Cartesian foundationalism (all mixed and *a fortiori* all merely empirical psychologies) operate *only* with an empirical concept of the 'I' as an object of inner experience, attempting (non-circularly) to ground all *other* cognition on that which is itself *an* ostensibly more secure and reliable cognition — with the aid, in the case of mixed rational psychology, of certain *a priori* axioms of the natural light. The conclusion of this section is therefore that, while Kant's level-circularity argument is strictly refutative when directed against the mixed varieties of rational psychology *envisaged in his thought experiment*, the Paralogisms do not appear to lay the groundwork for any telling criticism of the *actual* Cartesian project of founding all *other* first-order cognition on a first-order cognition. That the circularity stricture has nevertheless a certain force against both historical and contemporary forms of empiricism and naturalism will be argued in the Section 6. First, however, its actual use in the text of the Paralogisms must be examined.

<sup>23</sup> The talk of an "eternal circle" here would appear to be completely non-technical; Kant might as well have written 'an endless string' or 'interminable series' rather than "an eternal circle" of ambiguities, etc. Accordingly, this passage is not included among the applications of the anti-circularity stricture to be considered in the next section.

## 5. The Anti-Circularity Stricture in the Thought Experiments of the Paralogisms

Kant gives his anti-circularity stricture its most general — and eye-catching — formulation at the end of the “Observation” (appended to the four paralogisms of the first edition). This will be referred to as the A passage (A402). It is there immediately applied to a *mixed* doctrine of the soul of the *sort* first espoused by Descartes. A similar application of the anti-circularity stricture is found in (what will be called) the B passage (B421–422) appended to the condensed restatement of the four paralogisms in the second edition. A further, but different, use of the same stricture is found much earlier, at the very outset of the Paralogisms, in a passage that was taken over unaltered in the second edition (A346/B404). In the A/B passage (as it will be referred to), the target is a hypothetical completely pure science of the soul that hypostatizes the ‘I’ of the ‘I think’ as an *intelligible* object.<sup>24</sup>

The above-mentioned general formulation appears as the very first sentence of the A passage. In the following sentence, however, Kant narrows the focus to the last of the three level mistakes (or the second type of level circularity) considered in the previous section: hypostatizing (a) the pure *Ich denke* as (b) an empirical cognition of an object. The passage reads in full (numbering the sentences for ease of reference):

Now it is indeed very evident (*einleuchtend*) that [1] I cannot cognize as itself an object that which I must presuppose in order to cognize an object at all (*überhaupt*); and that [2] the determining Self (thinking) is different from the determinable Self (the thinking subject) as cognition is different from its object. Nevertheless, [3] nothing is more natural and seductive than the illusion of taking the unity in the synthesis of thoughts for a *perceived* [e.a.] unity in the subject of these thoughts. [4] One could call it the subreption of hypostatized consciousness (*apperceptionis substantiatae*) (A402).<sup>25</sup>

In (2), the perfectly general anti-circularity stricture of (1) is applied to an imagined mixed rational psychology based on the cognition of the empirical *ego cogitans* or “thinking subject”; equally as “evident” as (1), Kant maintains, is the fact that (2) “the determining Self,” which he here refers to as “thinking (*das Denken*)” instead of ‘the pure *Ich denke*,’ “is different from the determinable Self,” or ‘the thinking subject’ *qua* empirical object of inner perception, “*as cognition is different from its object*” (A402 e.a.). Now the latter *would* indeed be evident if it meant — as it seems it must, coming hard on the heels of the general circularity stricture — that the determining

<sup>24</sup> Despite having been described above as “eye-catching,” the first sentence of the A passage has been largely overlooked. Dyck does not cite it anywhere in his monograph; nor does Kitcher in hers (1990). The exception is Ameriks (2000, p. 52, cf. also p. 54), who paraphrases it with specific reference to both the A and B passages as “the ‘I’ itself cannot be known objectively because it is a condition of all knowing” (A402, B422). He does not, however, mention circularity or differentiate (in this context) between cognition of an empirical and that of an intelligible object.

<sup>25</sup> Altering the Guyer/Wood translation of *einleuchtend* as ‘illuminating’ (*einleuchtend* is not *erhellend*) and changing the word order from “as an object itself” to “as itself an object” in order to highlight the circularity of the move Kant describes here (without so labelling it). Guyer/Wood’s “in order to cognize an object at all” renders Kant’s *um überhaupt ein Objekt zu erkennen*, which differs hardly at all from *um ein Objekt überhaupt zu erkennen*. *Das Denken* has been rendered as ‘thinking’ rather than ‘the thinking.’

Self or 'I' of *pure* apperception differs from the determinable Self of *empirical* self-consciousness as (a) the highest *intellectual* condition of the cognition of all appearances *überhaupt* differs from (b) the inner cognition of the self *as an appearance or object of empirical introspection*. To hypostatize the former as though it were the latter is precisely the third level mistake considered in the previous section. But to say, as Kant does here, that they differ "as cognition ... from its object" is, frankly, not evident at all. Instead of "cognition" and "its object," one would have expected a contrast between a *condition* of all empirical cognition *überhaupt* and an *item* of empirical cognition. After all, Kant draws just such a contrast, using the same terminology, in the B edition, when he writes:

All *modi* of self-consciousness in thinking as such (*im Denken an sich*) ... provide thought with no object at all, and hence do not present my Self as an object to be cognized. It is not the consciousness of the **determining** Self or *Ich denke*, but only that of the **determinable** Self, i.e., *of my inner intuition* ... that is the object." (B407, e.a. and 'self' capitalized)

In light of its reoccurrence in this passage, the substitution of *das Denken* for *das Ich denke* at A402 is less troubling; and the words "the unity of consciousness in which *alles Denken* consists" just a few lines earlier (B406) makes it even less so. By contrast, the ill-chosen opposition of "cognition" and "its object" is made *more* troubling by the fact that, in sentence (3), Kant immediately goes on to juxtapose that which is only a *condition* of all cognition with the empirical *cognition* of an object as inner appearance: "Nevertheless nothing is more natural and seductive than the illusion of taking the unity in the synthesis of thought," that is, the merely intellectual unity represented in the *determining* Self or 'I' of *pure* apperception, "for a *perceived* unity in the subject of these thoughts" (A402 e.a.). What is in question in both (2) and (3), then, is precisely the third of the level mistakes distinguished earlier, *circularly* hypostatizing (a) the highest intellectual condition of all cognition of *appearances* in general as (b) a cognition of an *empirical* object or appearance. Taking (a) the 'I' of pure apperception for (c) the cognition of an *intelligible* object existing *in itself* is, of course, just as gross a level mistake (the second in the earlier enumeration), and just as evident a violation of the general anti-circularity stricture stated in (1). But *that* violation is not Kant's focus here, even if, in (4), he turns the reader's attention to a "subreption of hypostatized consciousness (*apperceptionis substantiatae*)" (A402) that presumably covers *both* kinds of level mistake or level circularity.

To turn now from the A to the B passage, the first thing to note is that, although its opening sentence refers only to a "misunderstanding," Kant's attention turns to circularity already in the second sentence, coming to rest on that fallacy in the penultimate and final sentences, even though the word 'circle' is absent. The final sentence also provides an example of a different error analogous to that with which most of the passage is concerned: construing (d) that which is only a pure *sensible* condition of all empirical intuition of objects *qua* appearances existing in determinate time relations as though it were (e) *an* object empirically intuited as existing in determinate time relations. While (e) is just a specific instance of (b) or an

appearance, (d) is only the *sensible* analogue of (a) “the unity of consciousness” or highest *intellectual* condition of the categories. The passage reads (without the final sentence):

[1] From all this one sees that [mixed] rational psychology has its origin in a mere misunderstanding. [2] The unity of consciousness, which grounds the categories, is here taken for *an intuition* of the subject *as an object*, and the category of substance is applied to it. [3] But this unity [of transcendental apperception] is only the unity of **thinking**, *through which no object is given*; and thus the category of substance, which [i.e., whose *legitimate* application] always presupposes a given **intuition**, cannot be applied to it, and hence this [transcendental] subject *cannot be cognized* at all. [4] Thus the [transcendental] subject of the categories cannot, by thinking them, obtain a concept of itself *as an object* of the categories; for [5] in order to think them, it must take its pure self-consciousness, which is just what is to be explained, as its ground. (B421–422 e.a.)

The final sentence gives what Kant presumably considers an intuitively more obvious example of the same *sort* of circle (again, without using the word):

[6] In just such a manner (*Ebenso*), the subject, in which the representation of time originally has its ground, cannot thereby determine its own existence in time, and if the latter cannot be, then the former, as the determination of its self (as a thinking being in general) through categories, can also not take place. (B422)

That *mixed* rational psychology is intended in (1) is indicated by the words “an intuition of the subject as an object” in (2) and by “given **intuition**” in (3); it would be very surprising if an intellectual rather than an empirical intuition were meant here. Were it not for the easily overlooked allusion in (2) to the circularity of applying the category of substance to the purely intellectual “unity of consciousness, *which grounds the categories*” (B421 e.a.) themselves and all cognition by their means, it might seem that the *only* point being made here were the familiar one highlighted in (3) and (4): since no such object as the pure ‘I’ of the ‘I think’ is actually given in empirical intuition (but only hypostatized as so given), the category of substance cannot be employed to obtain the sort of synthetic *a priori* cognition of the empirical self that has always been supposed to be attainable in the pure part of mixed rational psychology. After all, the words “which always presupposes a given **intuition**” clearly evoke the restriction placed on the use of the categories in the Transcendental Analytic. But, in fact, *both* the circularity alluded to in (2) *and* the limitation of the categories to *given* objects of intuition in (3) are intended to support the conclusion drawn in sentence (4): the (a) pure ‘I’ cannot itself be the object of a synthetic *a priori* cognition by means of the categories. For so to understand it would be a misemployment of the categories, to be sure, but it would also be an example of the “subreption of hypostatized consciousness (*apperceptionis substantiatae*)” mentioned at the end of the A passage. It is the clause numbered (5) that shifts attention from

the limitation imposed by the Transcendental Analytic to the *circle* inherent in the procedure of the rational psychologist who applies the categories to the 'I' of the 'I think' of transcendental apperception: "for [5] in order to think them [the categories], it must take its pure self-consciousness, which is just what is to be explained, as its ground."

Using "just what is to be explained" (the *explanandum*) as (part of) the explanation (the *explanans*) is roughly what is understood by a logically circular explanation. As Kant describes the imagined rational psychologist's procedure here, a category ('substance') is predicated of the unity of consciousness (the 'I' of the pure 'I think') that is itself the "ground" of all cognition of objects through categories. This is not *just* the *mis*application of the categories proscribed by the Transcendental Analytic; according to (5), it is also circular. The tenor of the whole passage, with its frequent references to intuition, suggests that what Kant has in mind here is the hypostatization of (a) the pure 'I' as though it were (b) an *empirical* object given in inner intuition or perception. One might therefore speak of a 'recuperation' of the essential content of the deleted A passage here in the B edition, though with the unhelpful talk of a "misunderstanding" now doing duty for the precise formulation of the anti-circularity stricture in the A edition.

That the anti-circularity stricture of A402 is really at the heart of this passage is made even clearer by sentence (6), which draws a parallel with the circle that occurs when (d) the transcendental subject whose pure intuition of time is the "ground" or condition of all temporal determination of objects *überhaupt* is said to be itself (e) a temporally determined object or inner appearance. Kant apparently considers this an even more evident instance of the sort of level circularity with which he is concerned in the A and B passages; otherwise, he would hardly have resorted to it to reinforce the point that (a) the "determining Self" of pure apperception cannot be said to be *determined* as an appearance with respect to the category of substance-accident. As all efforts to determine (a) "the nature of our thinking being" (A345/B403) through the categories are circular, so too, according to (6), is the attempt to determine (d) the transcendental subject whose pure intuition is the source of all temporal determination of objects *überhaupt* as itself (e) a temporal object or inner appearance.<sup>26</sup>

<sup>26</sup> The point only roughly sketched in (6) is more fully elaborated in a *Reflection* (R 5661) that Kant entitles "Answer to the question, is it an experience that we think?" His reply is unequivocal: "The consciousness of having an experience (*eine Erfahrung anzustellen*) or even just of thinking (*oder auch überhaupt zu denken*) is a **transcendental consciousness, not experience**" (18:319). The argument first lays it down as a postulate that being "an experience" entails being empirically determined with respect to time: "experience, without a temporal determination annexed to it, is impossible" (18:319). From this it follows that if the consciousness "that we think" were itself an experience (and the thinking self therefore an object that is empirically determined with respect to time), "there would be a[nother] time in which and with which, at the same time, a given time flowed by, which is absurd (*ungereimt*)" (18:319). The absurdity might be thought to reside in the violation of the (for Kant) self-evident time-axiom that there can only be one time (of which all particular time intervals are just the successive phases, much as all particular spaces are just simultaneously existing segments within the one all-encompassing space). But *this* absurdity — a plurality or, if an infinite regress starts, an infinity, of times within times — is just the consequence of violating the anti-circularity stricture of A402 by taking that which is the *a priori* subjective condition of *all* being-in-time of objects (*transcendental* subjectivity) to be *itself* in time. Kant could have developed the same line of reasoning in the B passage. Instead, he left matters at a hastily drawn parallel.

So much for the B passage. The third, much earlier A/B passage does not refer to a “misunderstanding” but explicitly to “a constant circle.”<sup>27</sup> It is not, of course, that this circular reasoning is the *only* thing that renders otiose the natural tendency of reason to develop a completely pure *scientific* rational psychology based on the *Ich denke* alone. This earliest of the three passages is only the opening gambit in the complex etiology of error that occupies the entire chapter. Still, the circle that Kant highlights in this introductory passage is enough to render null and void all the reasonings of this “imagined science” (A395) based on “nothing but the simple and in content for itself wholly empty representation I.” Since a completely pure science is in question, the A/B passage contains no such references to being given *in intuition* or *perception* as are found in both the A and B passages.

The third passage begins by juxtaposing (a) the “wholly empty representation I,” that is, the “mere consciousness ... that accompanies every concept” and (c) “a concept” through which the ‘I’ is *cognized* (understood: as a purely intellectual object). The former is “wholly empty” because the ‘I’ of pure apperception is no more than the merely *logical* form of all concepts of objects *überhaupt*; it can only be described at all in terms of the various forms of a merely *possible* synthetic unity of representations, that is, those unifying logical *functions* of the understanding which characterize discursive thinking in general.<sup>28</sup> These subjective *forms* of thinking underlie the categories, which *are* (or can become) concepts or cognitions of objects, provided the latter are not just thought, but also *given* in intuition. By themselves, however, the “*modi* of self-consciousness in thinking are ... mere functions, which provide thought with *no object at all*, and hence do not present my Self as an object” (B407 e.a.).

Nevertheless, the immediately following sentence in the German text of the A/B passage misleadingly states that the ‘I’ *is*, after all, *erkennt*, albeit only indirectly, “through the thoughts that are its predicates, ... in abstraction [from which thoughts] we can never have even the least concept” of it. There are two difficulties here. For one thing, if Kant is not to be said to violate his own anti-circularity stricture, *erkennt* must be understood in its other, perfectly ordinary, German meaning of ‘recognized’ (so in the translation of Guyer/Wood), not as ‘cognized’ (Dyck 2014, p. 78) or ‘known’ (Kemp Smith). The sense of ‘recognized’ here can be glimpsed through a

---

<sup>27</sup> There are five passages in which Kant *either* mentions *or* alludes to circularity in the Paralogisms. In addition to the A, B, and A/B passages (at A402, B422, and A346/B404, respectively), only the last of which *mentions* “a constant circle,” there is a passage at A366 and another at A395 in which the word ‘circle’ occurs. It was suggested in note 23 that the phrase “an eternal circle,” etc. at A395 is a *façon de parler* rather than a deliberate allusion to a *petitio principii*. The words “revolves in a circle around itself” at A366 cannot be so lightly dismissed, since the passage ends in a manner clearly indicative of a *petitio* (“merely presuppose what one demanded to know”). Nevertheless, Kant’s point seems to be that the conclusion of the third paralogism (like those of the first two) is merely an *analytic* consequence of the pure concept ‘I think.’ He speaks accordingly of “tautological answers” (A366), and it seems likely that the words “revolves in a circle” at A366 echo those many passages (see note 32 below for examples of this Pickwickian sense of ‘circle’) in which he declares some conclusion of an individual paralogism an “indisputably correct” (A400), but “completely empty” (A400) or analytic proposition.

<sup>28</sup> Elsewhere Kant calls the ‘I’ of pure apperception “the poorest representation of all” (B408) and “in every way indeterminate” (B426), which is equivalent to its description here as “wholly empty” and as “the transcendental subject of thoughts= x.”

consideration of the second difficulty. Despite the distinct echo of a famous passage in Descartes's *Principles* (I, 52), the second sentence cannot be a replica of Descartes's observation that "we cannot initially become aware of a substance merely through its being an existing thing, since this alone does not of itself have any effect on us," although we can "easily come to know a substance by one of its attributes," since the latter *do* affect us (AT IX–A 25: CSM I 210). Since Descartes immediately goes on to gloss 'attribute' with "properties or qualities" (AT IX–A 25: CSM I 210), his point, as regards *thinking* substance in particular, is that the *ego cogitans* can only be cognized indirectly via the *modi cogitandi* through which it affects itself, not directly, through its mere existence. Now since the *ego* of which Kant is speaking is plainly not this empirical, but rather the transcendental 'I' of *pure* self-consciousness, the "thoughts" or "predicates" *via* which it can be "recognized" (not "cognized") can only be those "*modi* of self-consciousness in thinking" (B407), which are "mere functions" or "forms" of all thinking in general. If this is correct, then the *Ich* of the pure *Ich denke* can only be "recognized ... through the thoughts that are its predicates" in the second-order, transcendental reflection of the *Critique*. Like formal logic, the *Critique* is not a science "properly and *objectively* so called" (B ix e.a.), that is, not a science containing first-order cognition *of objects*, but second-order cognition of the subject itself.<sup>29</sup> And here it is the logical table of judgements that provides second-order reflection with the requisite "clue (*Leitfaden*)" (A76/B102) to the pure forms of synthetic unity that underlie the tables of judgements and categories alike. However, the 'I' that is "recognized" through these forms cannot (on pain of circularity) be said to be *cognized* as an object of first-order cognition, not even cognized *indirectly* through these "predicates." For, as Kant notes elsewhere in the Paralogisms, "this I is no more an intuition than it is a concept of any object; rather, it is the mere form of consciousness which accompanies both sorts of representations" (A382). While it "can elevate *them* to cognitions ... insofar as something is given in intuition, which provides the material for the representation of an object" (A382 e.a.), the pure 'I' or Self can never be "elevated" to a concept or first-order cognition of an object. Still, just such a circular "hypostatizing" of "the transcendental subject of thought = *x*" takes place in the *objective* (see note 29) science of pure rational psychology toward which pure reason, in its pursuit of an unconditioned cognition to complete the series of conditioned cognitions, is naturally driven in its search for a stopping-point. In this way, a *condition* of all intuitive or discursive cognition of objects *whatsoever* is mistaken for "a representation distinguishing a particular object," the latter being, as Kant notes in the final sentence of the passage, what is meant when a representation is called "a cognition" of an object.

With the two main difficulties addressed in advance, it remains only to discuss the "constant circle" of the last three sentences of the A/B passage, which reads in its entirety:

[1] At the ground of this doctrine [a hypothetical, *completely pure* rational psychology] we can place nothing but the simple and in content for itself wholly

<sup>29</sup> On critique as a *subjective* science that resembles "logic ... in which reason has to do with itself alone" (B x), see A xiv and B23.

empty representation I, of which one cannot even say that it is a concept, but a mere consciousness that accompanies every concept. [2] Through this I, or He, or It (the thing) which thinks nothing further is represented than a transcendental subject of thoughts =  $x$ , which is recognized (*erkannt*) only through the thoughts that are its predicates, and of which, in abstraction, we can never have even the least concept; [3] around which (*um welches*) we therefore turn in a constant circle, since we must always already avail ourselves of the representation of it at all times in order to judge anything about it; [4] we cannot separate it ["this I, or He, or It"] from this inconvenience (*eine Unbequemlichkeit, die davon nicht zu trennen ist*) because consciousness in itself is not even a representation distinguishing a particular object, but rather the form of representation in general, insofar as it [the representation] is to be called a cognition; for [5] of it [namely, of cognition] alone can I say that through it I think anything [i.e., an object]. (A346/B404)<sup>30</sup>

The circularity problem here is obviously akin to that addressed in both the A and the B passage, though it is now posed with a completely pure rational psychology in view: we must inevitably "turn in a constant circle" should we "judge anything" about the 'I' of the *pure* 'I think' as an intelligible object. This is because "this I, or He, or It" is that which makes *every* concept and every judgement *as such* possible. While this is perfectly general ("judge *anything*"), it concerns chiefly those bogus synthetic *a priori* judgements that are the conclusions of the first three paralogisms and whose predicates are the categories of relation, quality, and quantity, respectively.<sup>31</sup> These pure concepts are themselves made possible by those "wholly empty" forms of discursive thinking through which the pure 'I' can be "recognized" in second-order, transcendental reflection on the table of judgements. The pure 'I think' cannot therefore *itself* be an intelligible object of first-order synthetic *a priori* cognitions through the very categories that it makes possible. For, in the words of the general formulation of the transcendental anti-circularity stricture of the A passage: "I cannot cognize as itself an object that which I must presuppose in order to cognize an object at all" (A402).

So much for the three passages. It remains to consider in more detail why the argument of the first two lacks all force against historical forms of Cartesian foundationalism, whether empiricist or rationalist, and why this type of transcendental argument may nonetheless have greater weight in the context of contemporary discussions of empiricism (naturalism) and rationalism than the type to which that title is nowadays commonly assigned.<sup>32</sup>

<sup>30</sup> Substituting "of which" for "about which" (*wovon*) in sentence (2), and "around which" for "because of which" (*um welches*) in sentence (3) of the Guyer/Wood translation. The use of *trennen* in (4) is transitive, not reflexive; "separate ourselves from" has accordingly been replaced with "separate it" (namely, the 'I' of the 'I think') from the "inconvenience" (i.e., circularity).

<sup>31</sup> While there is no mention of the categories in the A/B passage, the references to having "a concept" of the 'I' and to cognizing the 'I' "as a particular object" *allude* to the categories and their *misuse* in synthetic *a priori* pseudo-cognitions of the soul "though mere concepts" (A361, *passim*).

<sup>32</sup> Historically, there never was (as Kant realized) an *actual* instance of the application of the categories to the very "determining Self" that makes all synthetic cognition through the categories possible, since no one

## 6. Kantian Transcendental Arguments

Contemporary transcendental arguments modelled on Kant's Refutation of Idealism seek to demonstrate the absurdity of the bodiless, worldless, solipsistic *ego* of Cartesian scepticism by showing that the truth of those beliefs about one's body, the external world, and other minds that Descartes declared doubtful is entailed by those beliefs about the self that he deemed certain beyond the rational possibility of a doubt, outer being a necessary condition of inner cognition in general.<sup>33</sup> That the Refutation should be considered a *transcendental* argument is at least somewhat odd, since outer is not a *transcendental* condition of the possibility of inner cognition; indeed, it is so far from being an *a priori* and merely formal subjective *condition* of inner cognition *überhaupt* that it is rather itself a *cognition*, and an empirical one at that. In this section, a more recognizably Kantian candidate for the honourific title 'transcendental argument' will be examined, first in an historical, and then in a systematic perspective.<sup>34</sup>

It is a characteristically Kantian retort to the broadly foundationalist (or in Hume's case, sceptical) epistemologies of his empiricist and rationalist predecessors to insist that  $x$  cannot itself *be* a  $y$  (or be justified or explained by any number of  $y$ 's) if  $x$  is that which makes all  $y$ 's *überhaupt* possible. Sometimes he couches his arguments in fittingly modal language, for instance, in this remark from the Transcendental Aesthetic

---

before Kant had ever so much as glimpsed that transcendental Self. Yet three *sentences* with the same predicates ('substance,' 'simple,' etc.) and the same subject ('the soul') as the first three paralogisms are recognized as "identical," "analytic," or tautologically true propositions (and hence as circular in that other, Pickwickian sense mentioned in note 27 above). That is the likely source of the mistaken view that Kant has an historical target in mind. In the B edition, Kant presses this trivial form of 'circularity' charge (analyticity) against a "putative" completely pure science almost to the point of obscuring the circularity charge that really counts. But while indeed cognitions *of a sort*, the analytic truths generated by predicating the logical forms of the categories *of the pure I* (in the representation of which they are indeed all analytically contained) are not science "properly and objectively so called" (B ix; cf. note 29 above) and are therefore nothing to the purpose of the hypothetical pseudo-science under consideration. On their analyticity, see A400 and especially B407–409. Also 23:39: "lauter identische Sätze." Similarly, Düsing: "diese Prädikate sagen nach Kants Kritik nichts über das Dasein des Ich aus, sondern bilden nur analytische Bestimmungen seines reinen Denkens" (Düsing, 1987, p. 101).

<sup>33</sup> See Kitcher:

Transcendental arguments are intended to refute skepticism .... Their special characteristic is that they ensnare skeptics in self-contradiction. Once a skeptic grants that he has experience, an analysis of the concept of possible experience shows that it is a part of the meaning of 'experience' that the proposition the skeptic doubts must be true. (Kitcher, 1990, pp. 27–28)

Kitcher's description *straddles* the two forms of transcendental argument distinguished by Strawson:

a philosopher who advocates such an argument may [1] begin with a premise that the skeptic does not challenge ... and then proceed to argue that a necessary condition of the possibility of such experience is, say, knowledge of the existence of external objects ...; or he may [2] argue that the skeptic could not even raise his doubt unless he knew it to be unfounded. (Strawson, 1985, pp. 8–9)

<sup>34</sup> 'Argument' suggests a formal structure with discrete premises and conclusion. As so-called transcendental arguments have been (or can be) given such a structure (see previous note), so the anti-circularity stricture of A402 can serve as major premise in an argument whose conclusion states that a certain type of justification is level-circular. A paradigm would be the first metaphysical space argument, in which the major is only implicit.

(directed against the Lockean and Leibnizian views that ‘space’ is an empirical idea acquired through sensation): “That within which sensations can alone be ordered and placed in a certain form cannot itself be in turn sensation” (A20/B34). This version of the rule of A402 figures as a suppressed premise in the first space argument of the *Metaphysical Exposition*, where the *a priori* character of space is demonstrated.<sup>35</sup>

Hume made certain that Locke’s foundationalism would founder on the shoals of scepticism by substituting impressions for ideas as the sole explanatory and/or justificatory ultimates. A transcendental argument similar to that of the *Metaphysical Exposition* is at work in Kant’s emblematic reply to Hume in the *Prolegomena*: “[T]heir possibility,” he writes of the categories in general and cause-and-effect in particular, “is founded solely in the relation of the understanding to experience; not, however, in such a way that they are derived from experience, but that experience is derived from them, a completely reversed type of connection that never occurred to Hume” (4:313, cf. A301/B357). Thus, Hume too commits the error of “hypostatizing” what is for Kant (a) an *a priori condition* of all experience of outer events *überhaupt*, circularly elevating it to the level of a merely subjective impression or feeling of necessity produced in the mind by (b) a regular sequence of outer *perceptions* of A-like events followed by B-like events, when in fact (a) is that which makes (b) — the perception of *events* as distinct from enduring states of affairs — *überhaupt* possible.

There is no overlooking the parallel between these better-known transcendental arguments and the neglected thought experiment by which the Kant of the *Paralogisms* sought to radicalize Cartesian foundationalism. As he there introduces his own conceptions of space and causality in place of those of his opponents, so the Kant of the *Paralogisms* blithely substitutes his own ‘I think’ and radicalizing foundational project for those of Descartes, as though *the former* were at issue in the mixed rational psychology of his predecessors or contemporaries and not just in his own thought experiment. The key sentence reads: “[T]his proposition [*Ich denke*] is indeed certainly not an experience, but rather the form of apperception on which every experience depends and which precedes it” (A354). The unstated reason that the apperception in question *cannot* be an inner experience or empirical cognition of the self is that it is instead a *transcendental* condition of the possibility of *all* experience of objects (whether inner or outer) *as such*. Yet far from rigidly designating the universal and necessary form of thinking that “precedes” (A354) all particular experiences and on which experience *überhaupt* (as a form-matter composite) “depends” (A354), the self-consciousness expressed in Descartes’s *cogito* is a variable whose concrete values are all those particular inner experiences of the self that he considered the first and most reliable cognitions of an empirical object. In substituting his own *Ich denke* for the Cartesian *cogito* and introducing the novel idea of a necessary condition of all empirical cognition or experience *überhaupt*, Kant, one

<sup>35</sup> Allison adverts to the inevitable circularity of any attempt to account for our conceptions of space and time in a purely empirical or psychological manner” when he writes: “Reduced to its simplest terms, Kant’s main point is that we cannot hold with the empiricists that these conceptions [of space and time] are derived from our prior experience of externality and succession, because they are necessarily presupposed in any such experience. (Allison, 1973, p. 50)

is sorely tempted to say, is not so much laying bare a fallacy of the mixed rational psychology of Descartes and the tradition as simply changing the subject.<sup>36</sup>

Given that his predecessors not only lacked, but probably could have made no sense of, an *Ich denke* that was *not* a reliable cognition of the self on which to ground the cognition of other things, it would obviously be question-begging on Kant's part to tax any existing doctrine of the soul with "hypostatizing" his own *Ich denke* as though it were a first-order cognition of a sensible or intelligible object. However, if Kant is engaged in an imaginative thought experiment involving two merely hypothetical sciences of the soul that he *conceived* in terms of the very level mistakes he ascribed to them, then, however question-begging the critique of Descartes and his followers in an historical perspective, the Paralogisms still retain a certain interest as a rational suasion in favour of a transcendental radicalization of Cartesian foundationalism. For like the ontological, Kant's epistemological anti-circularity stricture has a positive sense that can be put this way: if there is to be stopping-point in the process of justification, if all object-level cognition *überhaupt* is to be non-circularly grounded in something that is not itself part of the *explanandum*, then what is required is a higher-order cognition of the *a priori* conditions of the possibility of any object-level cognition as such. To the extent that Kant himself suggests that he is engaged in a critique of an existing doctrine or doctrines in the Paralogisms, he has both given historical scholarship a false scent and done his own thought experiment a disservice. For his reflections on level-circularity serve above all to reinforce the thesis, developed throughout the *Critique*, that just as (1) empirical requires grounding in pure cognition if it is to be 'science' in the sense that Kant and his rationalist contemporaries still gave that word (roughly: a discipline that is either completely *a priori* or at least rests on *a priori* first principles), so (2) *all* first-order cognition must be grounded in something "higher" (B131) if science itself is to have a stopping-point that renders the justificatory process non-circular and (so Kant believed) absolutely secure against the relentless attacks of the sceptics.<sup>37</sup>

Since the failure to ascend to meta-level cognition of the subjective conditions of the possibility of object-level cognition is only *logically* circular where the *justificandum* or *explanandum* is the infinite totality of first-order cognition *überhaupt*, it is

<sup>36</sup> It is not merely that no one before Kant had any conception of a transcendental condition; for Descartes and Leibniz, as for both their eighteenth-century German followers and their British critics, *the very idea* of an explanatory or justificatory *condition* of all first-order cognition of objects that is *not* itself an item of first-order cognition of an (inner or outer) object would simply have made no sense. And it apparently still makes none to many today. When Klemme asks: "Wie kann das Ich Grund aller Erkenntnis sein, wenn wir es seinerseits nicht erkennen?" (Klemme, 2009, p. 180), he is thinking as Descartes (and everyone else) thought about foundations prior to Kant — indeed, as not only foundationalists, but even coherentists still think about the "structure of empirical knowledge" today (cf. Ayer, 1958; BonJour, 1985). And he takes himself to be expressing Kant's point of view as well. In a Kantian, transcendental perspective, however, the question is just the reverse: How could anything be the ultimate ground of all (object-level) cognition as such, if it *were* itself just an item of (object- rather than meta-level) cognition?

<sup>37</sup> The first *Critique* has fostered some notable attempts to show either (1) that some synthetic *a priori* cognition is necessary in order to guarantee the objective validity of empirical cognition and/or (2) that all cognition requires ultimate grounding in meta-level, synthetic *a priori* cognition of the subject. For a modern-day defence of (1) by an analytic philosopher (whose chief argument is precisely that the only alternative is an unacceptable form of scepticism), see BonJour (1998) on *a priori* justification. The best-known application of the lesson of (2) is the foundationalist project of Husserlian transcendental phenomenology, but examples in Fichte and Schelling abound.

easy enough to evade the force of Kant's argument by denying that such ultimate grounding of cognition *überhaupt* is even possible, let alone necessary. Still, as a *normative injunction* to push epistemological inquiry from the object- to the meta-level, Kant's epistemological anti-circularity stricture deserves the same respect that its ontological counterpart has always received in rational cosmology and natural theology. It is not without reason that Van Cleve endorses the demurrer of "cosmological arguers" who balk at efforts to explain the whole natural and contingent order of things *without* recourse to an ultimate existent of a *different* and higher (transcendent) order. The force of the epistemological anti-circularity stricture has been similarly felt (and continues to be felt) by what may be called 'transcendental arguers,' beginning with Kant. At a minimum, Stroud's picture of naturalism, on the one hand, and transcendental arguments (in *his* sense), on the other, as the only promising lines of attack upon Cartesian foundationalism requires supplementation by an alternative foundationalism based on the more characteristically Kantian sort of transcendental argument examined here.<sup>38</sup>

## 7. Conclusion

The conclusion of all this is that anticipated at the end of Section 4. Since Kant's level-circularity arguments are strictly refutative only with respect to the peculiar forms of rational psychology envisaged in his thought experiment, the Paralogisms fall far short of a decisive critique of the *historical* Cartesian project of founding all *other* first-order cognition on *a* first-order cognition. Nevertheless, the circularity stricture formulated there is highly relevant to the question of whether it is not perhaps more desirable to transcendentalize than to naturalize Cartesian foundationalism.

In this regard, it is worth recalling that Gadamer once claimed (personal recollection) to have learnt from Heidegger that what matters in philosophy is not going beyond, but rather going back behind, one's predecessors. Heidegger himself will have learnt this lesson from Husserl, and Husserl, in turn, from Descartes and Kant. While Kant no doubt believed himself to have gone beyond his predecessors, and even to have shown all previous foundationalisms to be viciously circular, refutations that do not beg the very question at issue are rare in philosophy. Kant did, however, go back behind all his predecessors, opening up a deeper-lying, transcendental dimension to philosophical research, and therein consist both the historical importance and the systematic relevance of his Paralogisms today.

## References

- Allison, H. (1973). Kant's critique of Berkeley. *Journal of the History of Philosophy*, 11(1), 43–63.  
 Allison, H. (2004). *Kant's transcendental idealism: An interpretation and defense*. Yale University Press.

<sup>38</sup> The point is that the epistemological anti-circularity stricture *itself* deserves such consideration, whatever the extent to which the same is true of Kant's employment of it. To say that (1) the ultimate foundation of *all* natural (empirical) and non-natural (*a priori* and synthetic) cognition (in mathematics, pure natural science, and metaphysics) *überhaupt* cannot itself be some particular item or items of first-order synthetic *a priori* cognition of an object or objects is one thing; that it must be second-order synthetic *a priori* or transcendental cognition of the *knowing subject* and the *subjective* conditions of all first-order cognition of objects whatsoever is quite another matter. The latter is presumably what Guyer (1989, p. 59) means by Kant's "Cartesian presumption" when he comments, justly, that it "is a big one indeed."

- Allison, H. (2006). Kant's transcendental idealism. In G. Bird (Ed.), *A companion to Kant* (pp. 111–124). Oxford University Press.
- Ameriks, K. (2000). *Kant's theory of the mind: An analysis of the paralogisms of pure reason*. Oxford University Press.
- Ayer, A. J. (1958). *The foundations of empirical knowledge*. Macmillan & Co.
- Bennett, J. (1974). *Kant's dialectic*. Cambridge University Press.
- BonJour, L. (1985). *The structure of empirical knowledge*. Harvard University Press.
- BonJour, L. (1998). *A defense of pure reason*. Cambridge University Press.
- Braithwaite, R. B. (1955). *Scientific explanation*. Cambridge University Press.
- Curley, E. (1978). *Descartes against the sceptics*. Harvard University Press.
- Descartes, R. (1984–1991). *The philosophical writings of Descartes* (J. Cottingham, R. Stoothoff, & D. Murdoch, Trans. and Eds., 3 volumes). Cambridge University Press. (References to Descartes's works are to volume and page of the definitive Franco-Latin edition of C. Adam & P. Tannery [*Oeuvres de Descartes*, 11 volumes. Paris, 1897–1913, nouvelle présentation par P. Costabel et B. Rochot, Vrin-CNRS, 1964–1976], followed by volume and page of the now standard English translation by Cottingham et al., e.g., AT VII 121: CSM II 86.)
- Düsing, K. (1987). Cogito, ergo sum? Untersuchungen zu Descartes und Kant. *Wiener Jahrbuch für Philosophie*, 19, 95–106.
- Dyck, C. W. (2014). *Kant and rational psychology*. Oxford University Press.
- Grier, M. (2001). *Kant's doctrine of transcendental illusion*. Cambridge University Press.
- Guyer, P. (1958). Psychology and the transcendental deduction. In E. Förster (Ed.), *Kant's transcendental deductions: The three "Critiques" and the "Opus posthumum"* (pp. 47–68). Stanford University Press.
- Guyer, P., & Wood, A. (1998). *The Cambridge edition of the works of Immanuel Kant*. Cambridge University Press.
- Heidegger, M. (1998). *Pathmarks* (W. McNeill, Ed.). Cambridge University Press.
- Horstmann, R. P. (1993). Kants paralogismen. *Kant-Studien*, 84(4), 408–425.
- Husserl, E. (1960). *Cartesian meditations* (D. Cairns, Trans.). Martinus Nijhoff.
- Kannisto, T. (2018). Transcendental paralogisms as formal fallacies — Kant's refutation of pure rational psychology. *Kant-Studien*, 109(2), 195–227.
- Kim, H. (2019). Was heißt "Ich denke is ein empirischer Satz"? *Kant-Studien*, 110(1), 136–159.
- Kitcher, P. (1982). Kant's paralogisms. *The Philosophical Review*, 91(4), 515–547.
- Kitcher, P. (1990). *Kant's transcendental psychology*. Oxford University Press.
- Klemme, H. (2009). 'Das Ich denke ist ... ein empirischer Satz': Kants Auseinandersetzung mit Descartes. In J. Ferrari, P. Guenancia, M. Ruffing, R. Theis, & M. Vollet (Eds.), *Descartes and Deutschland — Descartes et l'Allemagne* (pp. 177–186). Europea Memoria, Series 1, Volume 71. Georg Olms Verlag.
- Leibniz, G. W. (1875–1890). *Die Philosophischen Schriften von Gottfried Wilhelm Leibniz* (C. I. Gerhardt, Ed.). Reprint, Georg Olms. (Abbreviated as G with volume and page number.)
- Leibniz, G. W. (1996). *New essays on human understanding* (P. Remnant & J. Bennett, Eds.). Cambridge University Press.
- Longuenesse, B. (2007). Kant on the identity of persons. *Proceedings of the Aristotelian Society*, 107, 149–167.
- Longuenesse, B. (2008). Kant's "I think" versus Descartes' "I am a thing that thinks." In D. Garber & B. Longuenesse (Eds.), *Kant and the early moderns* (pp. 9–31). Princeton University Press.
- Nolan, L. (Ed.) (2016). *The Cambridge Descartes lexicon*. Cambridge University Press.
- Proops, I. (2010). Kant's first paralogism. *Philosophical Review*, 119(4), 449–495.
- Quine, W. V. O. (1953). Two dogmas of empiricism. In *From a logical point of view* (pp. 20–46). Harvard University Press.
- Quine, W. V. O. (1969). Epistemology naturalized. In *Ontological relativity and other essays* (pp. 69–90). Columbia University Press.
- Stern, R. (2000). *Transcendental arguments and scepticism: Answering the question of justification*. Oxford University Press.
- Strawson, P. F. (1966). *The bounds of sense: An essay on Kant's Critique of pure reason*. Methuen & Co.
- Strawson, P. F. (1985). *Skepticism and naturalism: Some varieties*. Columbia University Press.
- Stroud, B. (1968). Transcendental arguments. *Journal of Philosophy*, 65(9), 241–256. Reprinted in T. Penelhum & J. J. MacIntosh (Eds.), *The first critique: Reflections on Kant's Critique of pure reason* (pp. 54–69). Wadsworth Publishing.

- Stroud, B. (1979). The significance of scepticism. In P. Bieri, R.-P. Horstmann, & L. Krüger (Eds.), *Transcendental arguments and science: Essays in epistemology* (pp. 277–297). D. Reidel Publishing Company.
- Van Cleve, J. (1999). *Problems from Kant*. Oxford University Press.
- Wilson, M. (1974). Leibniz and materialism. *Canadian Journal of Philosophy*, 3(4), 495–513.

---

**Cite this article:** Miles, M. (2022). Kant's Radicalization of Cartesian Foundationalism: Thought Experiments, Transcendental Arguments, and Level Circularity in the Paralogisms. *Dialogue* 61(3), 493–518. <https://doi.org/10.1017/S0012217322000348>