Julian Wuerth - Kant's Immediatism, Pre-Critique

Julian Wuerth

Journal of the History of Philosophy, Volume 44, Number 4, October 2006, pp. 489-532 (Article)

Published by Johns Hopkins University Press
DOI: 10.1353/hph.2006.0074

Kant's Immediatism, Pre-*Critique*

JULIAN WUERTH*

AS THE AUTHOR OF A COPERNICAN REVOLUTION IN PHILOSOPHY,[1] Kant argues that philosophy begins with the study of the self.[2] To grasp the scope and nature of

[1] The analogy to Copernicus was, of course, first offered by Kant in the preface to the second edition of his *Critique of Pure Reason* (*CPR*, B xvi).

[2] Citations to Kant's works are inserted parenthetically in the text and refer to the title of the work with an abbreviation listed here and include a volume and page number from *Kant's gesammelte Schriften, herausgegeben von der Berlin-Brandenburgischen Akademie der Wissenschaften* [formerly the *Deutschen Akademie der Wissenschaften zu Berlin*, and before that the *Königlichen Preussischen Akademie der Wissenschaften*], 29 volumes (Berlin: Walter de Gruyter [and predecessors], 1900–). Translations from the German are my own, unless a translation of the source is listed here, in which case I often follow the translation. All translations by David Walford are found in *The Cambridge Edition of the Works of Immanuel Kant, Theoretical Philosophy, 1755–1770* (Cambridge: Cambridge University Press, 1992). All translations by Karl Ameriks and Steve Naragon are found in *The Cambridge Edition of the Works of Immanuel Kant, Lectures on Metaphysics* (Cambridge: Cambridge University Press, 1997). Citations to the first *Critique* are to the pagination of the first edition, A, and/or to the second edition, B, as translated in *The Cambridge Edition of the Works of Immanuel Kant, Critique of Pure Reason* [*CPR*], ed. and trans. Paul Guyer and Allen Wood (Cambridge: Cambridge University Press, 1997).

A	*Anthropologie in pragmatischer Hinsicht* (7) [1796–97]. *Anthropology from a Pragmatic Point of View*, trans. Victor Dowdell (Carbondale: Southern Illinois University Press, 1978).
AC	*Anthropologie-Collins* (25) [1772/73].
AF	*Anthropologie-Friedländer* (25) [1775/76].
AM	*Anthropologie-Mrongovius* (25) [1784/85].
APa	*Anthropologie-Parow* (25) [1772/73].
APh	*Anthropologie-Philippi* (25) [1772/73].
APi	*Anthropologie-Pillau* (25) [1777/78].
C	*Kants Briefwechsel* (10–13). *Philosophical Correspondence, 1755–99*, trans. Arnulf Zweig, of selections (Chicago: University of Chicago Press, 1970).
CJ	*Kritik der Urtheilskraft* (5) [1790]. *Critique of the Power of Judgment*, trans. Paul Guyer and Eric Matthews (Cambridge: Cambridge University Press, 2000).
CPrR	*Kritik der praktischen Vernunft* (5) [1788]. *Critique of Practical Reason*, trans. Lewis White Beck (New York: Macmillan Publishing Company, 1956).
DSS	*Träume eines Geistersehers, erläutert durch Träume der Metaphysik* (2) [1766]. *Dreams of a Spirit-Seer elucidated by Dreams of Metaphysics*, trans. Walford.
FS	*Die Falsche Spitzfindigkeit der vier syllogistischen Figuren* (2) [1762]. *The False Subtlety of the Four Syllogistic Figures*, trans. Walford.
IC (2)	*Untersuchung über die Deutlichkeit der Grundsätze der natürlichen Theologie und der Moral* [1764]. *Inquiry concerning the Distinctness of the Principles of Natural Theology and Morals*, trans. Walford.

* **Julian Wuerth** is Assistant Professor of Philosophy at the University of Cincinnati.

Journal of the History of Philosophy, vol. 44, no. 4 (2006) 489–532

knowledge in natural science, ethics, and aesthetics, we must first understand the self and its faculties of representation, pleasure, and desire, respectively.[3] While the literature acknowledges the foundational role of Kant's account of the self in his system of philosophy, it typically focuses on Kant's isolated conclusions about *a priori* concepts, or on the ontology of the self solely as Kant presents it in his rejection of rational psychology. The result is a gap in the commentary. The place of these conclusions in a single, broader account of the self, and the positive ontological status of this self, remains largely a mystery. These shortcomings are related: Kant's account of the ontology of the self provides the background for understanding the unity and interrelations of the many different powers of the self and their accidents, including *a priori* concepts.

ID	*De mundi sensibilis atque intelligibilis forma et principiis* (2) [1770]. *Concerning the Form and Principles of the Sensible and Intelligible World* (*The Inaugural Dissertation*), trans. Walford.
Me	*Menschenkunde* (25) [1781/82].
MD	*Metaphysik Dohna* (28) [1792–93], trans. Ameriks and Naragon, of 28:656–90.
MH	*Metaphysik Herder* (28, 29) [1762–64], trans. Ameriks and Naragon, of 28:39–53.
MK2	*Metaphysik K2* (28) [early 1790s], trans. Ameriks and Naragon, of 28:753–75.
ML1	*Metaphysik L1* (28) [1777–80], trans. Ameriks and Naragon, of 28:195–301.
ML2	*Metaphysik L2* (28) [1790–91], trans. Ameriks and Naragon, of 28:531–94.
MFNS	*Metaphysische Anfangsgründe der Naturwissenschaft* (4) [1786]. *Metaphysical Foundations of Natural Science*, trans. James Ellington (Indianapolis and New York: Hackett Publishing Company, 1970).
MMr	*Metaphysik Mrongovius* (29) [1782–83], trans. Ameriks and Naragon, complete.
MVi	*Metaphysik Vigilantius (K3)* (29) [1794/95], trans. Ameriks and Naragon, complete.
MVo	*Metaphysik Volckmann* (28) [1784-5], trans. Ameriks and Naragon, of 28:440–50.
MvS	*Metaphysik v. Schön* (28) [1780s].
MoMr	*Moral Mrongovius* (27) [1782].
NE	*Principiorum primorum cognitionis metaphysicae nova delucidatio* (2) [1755]. *New Elucidation of the First Principles of Metaphysical Cognition*, trans. Walford.
NM	*Versuch den Begriff der negativan Größen in die Weltweisheit einzuführen* (2) [1763]. *Attempt to Introduce the Concept of Negative Magnitudes into Philosophy,* trans. Walford.
PE	*Philosophische Enzyklopädie* (29) [1775].
Pr	*M. Immanuel Kants Nachricht von der Einrichtung seiner Vorlesungen in dem Winterhalbenjahre von 1765–1766* (2) [1765]. *M. Immanuel Kant's Announcement of the Programme of his Lectures for the Winter Semester 1765–6*, trans. Walford.
Pro	*Prolegomena zu einer jeden künftigen Metaphysik* (4) [1783]. *Prolegomena to Any Future Metaphysics*, trans. Ellington (Indianapolis: Hackett Publishing Company, 1977).
R	*Reflexionen* (15, 17, 18, 19, 23).
RP	*Welches sind die wirklichen Fortschritte, die die Metaphysik seit Leibnitzens und Wolf's Zeiten in Deutschland gemacht hat?* (20) [1795]. *What Real Progress Has Metaphysics Made in Germany since the Time of Leibniz and Wolff?*, trans. Ted Humphrey (New York: Abaris Books, 1983).

[3] That Kant takes this investigation of natural science and metaphysics, ethics, and aesthetics and teleology, to concern *a priori* concepts as they relate to representations, pleasures and displeasures, and desire, in particular, owes to Kant's tripartite division of the soul's faculties into these three, as consistently maintained by Kant throughout his career (e.g., *AC*, 25:30; *Religionslehre Pölitz*, 28:1059 [1783–84]*; Danziger Rationaltheologie*, 28:1275 [1784]; *C*, 10:513–15 [1787]; *CPrR*, 5:12,15; *CJ*, 20:245–46; *MVi*, 28:836), in opposition to the rationalists, who Kant argues erred in collapsing these three into the faculty of representation (*MH*, 28:145; *ML1*, 28:261–62; *MK2*, 28:815). This tripartite division of faculties is mirrored in Kant's division of labor between the three *Critiques*, the first emphasizing representations, the second desire, and the third pleasures and displeasures.

A recent spate of literature on Kant's account of the self addresses important components of this account but perpetuates some traditional methodological weaknesses. The often original commentaries by Patricia Kitcher, Thomas Powell, and Wayne Waxman stick to frontline issues such as Kant's Paralogisms, his accounts of different faculties of representation, and his Transcendental Deduction, for example; they do not examine Kant's underlying positive ontology of the self, his more general ontology of substance, powers, and accidents, or the other faculties that Kant ascribes to the self.[4] This traditional orientation reflects a problematic research methodology: these accounts do not consider any of Kant's voluminous recorded thought from before the *Critique of Pure Reason* (1781), even though Kant published the *Critique* when he was already fifty-seven years old, after decades of detailed, systematic, and largely consistent studies of the self. These interpretations instead rely almost exclusively on the *Critique* and rarely even consult Kant's many other post-1781 sources. They accordingly focus on the faculties highlighted in the *Critique,* namely the various faculties of representation and not the faculties of pleasure or desire, and they take the *Critique's* pithy and notoriously cryptic chapter on the Paralogisms as a full statement of the soul's ontology. Unfortunately, the Paralogisms present only a *negative* account of the soul's ontology, not a complete one. These interpretations therefore understand Kant's rejection of

[4] Patricia Kitcher, *Kant's Transcendental Psychology* [*KTP*] (New York: Oxford University Press, 1990); "Kant's Real Self," in *Self and Nature in Kant's Philosophy*, ed. Allen Wood (Ithaca, N.Y.: Cornell University Press, 1984): 113–48; C. Thomas Powell, *Kant's Theory of Self-Consciousness* (Oxford: Clarendon Press, 1990); and Wayne Waxman, *Kant's Model of the Mind: A New Interpretation of Transcendental Idealism* (New York: Oxford University Press, 1991). Andrew Brook's outstanding *Kant and the Mind* (Cambridge: Cambridge University Press, 1994) is also generally and intentionally ahistorical, but while Brook does not delve into questions of Kant's account of substance, powers, and accidents, he does take seriously the possibility that Kant's self is a noumenal substance in his useful survey of the secondary literature on the question of what we are immediately conscious of with immediate consciousness (246–52). Karl Ameriks' earlier *Kant's Theory of Mind: An Analysis of the Paralogisms of Pure Reason* [*Kant's Theory of Mind*] (Oxford: Clarendon Press, 1982), which still sets the standard for studies of Kant's account of the self more than two decades after its first publication, provides, from among these commentaries, by far the most historically sensitive and systematic interpretation of Kant's account of the self, though he of course did not yet have available to him most of the notes on Kant's lectures on anthropology, which have since been published, and he does not focus as much on Kant's *positive* ontology of the self as do I, nor as much on the role of *personal identity* as a catalyst for Kant's account of the self as do Kitcher and I. I addressed Ameriks' interpretation of the Paralogisms in "Kant's Immediatism, Post-1781" (Pacific Division Meeting of the American Philosophical Association, 2001), and argued against his position (*Kant's Theory of Mind*, 67–70) that the second (but not the first) edition of the *Critique* argues for the possibility that the "I" could be a mere accident, and not a noumenal substance. Like both Allison (1983) and Pippin (1987), Ameriks sees the claim that the soul is a noumenal substance as irreconcilably at odds with Kant's critique of rational psychology (*Kant's Theory of Mind*, 67). This essay provides some of the background for my rejection of this view, by presenting Kant's pre-*Critique*, positive arguments for his claims of the soul's noumenal substantiality and simplicity, clarifying these claims, and briefly outlining the contrast that Kant draws between these claims and those of the rationalists, which he rejects. Also, like Ameriks, Allison, and Pippin, James Van Cleve (in his insightful *Problems from Kant* [New York: Oxford University Press, 1999]) recognizes that Kant seems to claim (for him, in the first edition of the First Paralogism) that the soul is a substance, where this is a noumenal, not phenomenal substance. But unlike Ameriks, Allison, and Pippin, who claim that this position is inconsistent with Kant's rejection of rational psychology, Van Cleve argues that it is inconsistent with the First Analogy (*Problems from Kant*, 72–74). For more on Van Cleve's position and its relation to mine, see note 64, below. Finally, Heiner F. Klemme's recent *Kants Philosophie des Subjekts. Systematische und entwicklungsgeschichtliche Untersuchungen zum Verhältnis von Selbstbewußtsein und Selbsterkenntnis* (Hamburg: Meiner, 1996) is a masterly study of Kant's account of the self across his career to which I, unfortunately, do not have space to respond in this essay.

the intemperate *rationalist* arguments—which hold that the conclusions of the soul's simplicity and substantiality imply the soul's permanence, incorruptibility, and immortality—as Kant's rejection of *all* ontologically significant conclusions of the soul's substantiality and simplicity, even more temperate ones that do not imply the soul's permanence, incorruptibility, and immortality.

Although these interpretations—united in their view that Kant rejects the substantiality of the soul in any ontologically significant sense—may at first seem viable if we consider only the Paralogisms, a closer look at the Paralogisms reveals significant problems, a look beyond the Paralogisms shows that these interpretations are repeatedly contradicted, and application of these interpretations when approaching other doctrines in Kant's philosophy repeatedly leads us down dead-ends. Thus Kant states in the Paralogisms, and in many other places after the *Critique*, that the soul is a substance, understood in the most basic ontological sense of the pure category of substance.[5] And Kant claims that the soul has powers and accidents, while at the same time repeatedly making clear that accidents (unlike mere logical predicates) are a form of existence and that all existence is grounded in substance, either being substance or a determination, or mode, of substance. Without Kant's concept of noumenal substance, his concepts of the soul's accidents and powers are rendered incoherent, frustrating our understanding of many other Kantian doctrines, including Kant's critique of rational psychology, his transcendental idealism, his views on the relationship of the noumenal to phenomenal, and his accounts of absolute spontaneity. One more doctrine confused by an incomplete understanding of Kant's metaphysics of the soul is that of personal identity. While the soul's accidents and powers are many, Kant holds that they are united in an identical soul. Still, Kant's basis for this claim of personal identity remains poorly understood, a problem that has undermined analysis of Kant's rejection of rational psychology in the Paralogisms.

Kitcher thinks, as I do, that an appreciation for the historical context for Kant's account of the self in the *Critique* is important to deciphering this often cryptic account. Kitcher's interpretation, again like mine, holds that Kant's concern over the problem of personal identity prior to the *Critique* plays an important role in shaping his account of the self. Nonetheless, our historical interpretations differ radically. Kitcher believes that passages in the *Critique* link Kant beyond doubt to Hume's *Treatise* account of personal identity, and that Kant is therefore responding to Hume's account of personal identity in the *Critique* and doing so on Hume's empiricist terms. Not surprisingly, because she eliminates interpretive options that would have Kant failing to abide by Hume's empiricist guidelines, she interprets Kant to offer an empiricist, reductionist account of the self in the *Critique*. This account reduces the self to nothing but representations and perhaps powers,[6] to

Kitcher's Kant

[5] For example, *CPR*, A 350–51, A 356, A 365–66, A 399–403; *Pro*, 4:334–35; *R* 6001, 18:420–21 [1780–89]; *MMr*, 29:770–71, 772, 904, 912; *MvS*, 28:511; *MFNS*, 4:542–43; *Vorarbeiten zur Schrift gegen Eberhard*, 20:359 [approx. 1790]; *ML2*, 28:590–92; *MD*, 28:681–86; *MK2*, 28:754–56, 759; *RP*, 20:270, 308–09; *What Real Progress, Loße Blätter*, 20:340 [approx. 1793]; and *MVi*, 29:1025–27, 1032.

[6] Kitcher's "Kant's Real Self" appears to argue not only that Kant's self was not a substance, but also that it had no powers, with mental states causing subsequent mental states: "Kant expressed himself in this way to stress that the content of the resultant state is not merely similar to the contents of the earlier states, but is actually produced from the earlier states and their contents" ("Kant's Real Self,"

the exclusion of a non-empirical form of apperception and of a substantial soul that we might apperceive by means of this non-empirical apperception. Kitcher's interpretation of the historical context for Kant's *Critique* is therefore an empiricist, Humean one, and it implies an anti-metaphysical trajectory heading into the *Critique*.

In what follows, I argue that during the two decades preceding the *Critique* Kant concludes that the self is a simple substance and that, from 1772 until shortly before the 1781 *Critique,* Kant argues both that we have an immediate awareness of being this simple substance and also that this immediate awareness and with it our simplicity and substantiality are necessary conditions for thought and personal identity. Part 1 begins by rejecting Kitcher's historical analysis that there is a unique historical link between Kant's *Critique* and Hume's empiricist account of personal identity in the *Treatise*.

Part 2 moves on to Kant's own positive conclusions concerning the ontology of the soul prior to 1781, i.e., to the conclusions that the soul is a noumenal substance and that it is simple. Here I examine Kant's concepts of noumenal substances, powers, accidents, and related concepts, and show that while Kant argues that the soul is a simple substance, he rejects the same conclusions when these are construed, as by the rationalists, to imply the soul's permanence and incorruptibility. Kant charges that the rationalists conflate substance and power; that they accordingly fail to recognize that the simple soul can have more than one power; that they accordingly distinguish in degree and not in kind between understanding and sensibility; that they thus intellectualize appearance; and that they subsequently distinguish in degree and not in kind between noumenal substance and phenomenal substance. Kant then argues that phenomenal substance is a mere analogy to noumenal substance as found in immediate self-consciousness, and that permanence is only tautologically ascribed to phenomenal substance, though it does not hold for noumenal substance. Kant accordingly maintains that the rationalists—failing to register these facts about substance and permanence—infer from the soul's noumenal substantiality and simplicity to its permanence, incorruptibility, and immortality, i.e., to properties that only accrue to something if substantial and simple at the *phenomenal* level.

Part 3 then turns to the first of Kant's two immediatist arguments for his conclusions of the soul's substantiality and simplicity, his pseudo-empirical immediatist argument, which maintains that we have an immediate consciousness of being something that exercises powers to ground accidents, i.e., of being a noumenal substance. Part 4 then covers Kant's second immediatist argument, his transcendental argument, which asserts that our simplicity and substantiality, as well as

---

115). Her later *KTP*, however, makes clear that the self does have powers, which are for Kant merely activated faculties, asserting, "Thinking selves are not merely systems of cognitive states, because some faculty must always be present to synthesize states" (*KTP*, 123). But while Kitcher acknowledges that Kant's self has powers, she argues that it is not a substance, and that Kant's ostensibly wholesale rejection of the soul's substantiality against the rationalists goes back to Hume: "Given the background of Kant's response to Hume, we can understand one way that he thinks Rational Psychologists have been led astray. As Hume notes, we do not have an intuition of the self …" (*KTP*, 188). Also: "Kant tries to prevent these confusions [about the nature of the mind] by repeating and extending Hume's original objection to Descartes …" (*KTP*, 189; see also *KTP*, 194).

our immediate consciousness of this, are necessary for thought and personal identity. Although Kant here infers to his conclusions, he readily concedes that these inferences really do nothing more than draw attention to what immediate consciousness already offers, so that the real work is done by immediate consciousness. I also note in Part 4 another important structural relationship revealed in these pre-*Critique* writings. Kant traces the ideas of reason, like the analogies of experience, back to our immediate consciousness of ourselves as a paradigm. He accordingly recognizes that reason's inferences can reveal nothing about ourselves not already supplied by immediate consciousness; that the rationalists accordingly entertain false hopes when embarking on inferences concerning the self; and that they subsequently arrive at stronger conclusions by means of these inferences than those warranted by immediate consciousness alone only because, along their path of inference, they conflate phenomenal and noumenal properties (as described in Part 2). Part 5 concludes with a summary of my arguments, but it also offers a brief overview of the relation of Kant's positive account of the self and personal identity to the accounts of self and personal identity of his empiricist predecessors, Locke and Hume, and suggests connections to central questions in modern philosophy of mind.

A methodological note: Kant repeatedly lavishes attention on the topic of the ontology of the soul during the two decades between the time that he was thirty-seven and fifty-seven years old, regularly offering highly detailed accounts. These include those found in the approximately one thousand pages of invaluable notes on Kant's lectures on anthropology first published in any language, German, in 1997[7] and so not available to Kitcher, Powell, or Waxman. Other accounts and

---

[7] The new *Akademie* volumes 25.1 and 25.2 contain 1564 pages of notes on Kant's anthropology lectures, but 354 of these belong to the 1781/82 set of notes, *Menschenkunde*, which was previously published in 1831 as *Immanuel Kant's Menschenkunde oder philosophische Anthropologie. Nach handschriftlichen Vorlesungen herausgegeben*, ed. Fr. Ch. Starke (i.e., Johann Adam Bergk) (Leipzig, 1831); this was reprinted in 1976, ed. Giorgio Tonelli (Hildesheim/New York, 1976). Starke also published a shorter set of notes (32,000 words as compared to the 103,000 of the *Menschenkunde*), from 1790–91, nearly a third of which was actually taken from *Menschenkunde*. This set of notes, published as *Immanuel Kant's Anweisung zur Menschen- und Weltkenntniß. Nach dessen Vorlesungen im Winterhalbjahre von 1790–1*, ed. Fr. Ch. Starke (Leipzig, 1831), was also included in Tonelli's 1976 reprint, but was not included in the *Akademie* collection. Finally, Starke also published *Taschenbuch für Menschenbesserung nach Hippel, Wieland, Sterne, Helvetius, Shakespeare und Kant. Mit einer Abhandlung über Menschenkenntniß von Dr. Heinichen*, ed. Fr. Ch. Starke (Leipzig, 1826), which included many passages from *Menschenkunde* and *Anweisung*, and also a few fragments of notes on Kant's lectures on anthropology from the 1770s. These fragments likely come from Theodor Gottlieb von Hippel's *Lebensläufe*, which included traces of notes on Kant's lectures on anthropology that point mainly to the 1772/73 lectures.

In addition to Starke's publications, in 1924, Arnold and Elisabeth-Maria Kowalewski released *Aus Kants Vorlesungen über Anthropologie nach einem ungedruckten Kollegheft vom Wintersemester 1791–1792*, in *Philosophische Kalender für 1925. Im Zeichen Immanuel Kants* (Berlin, 1924). The notebook to which this title refers itself contains the title, *ANTHROPOLOGIE docente Profess: Kant. Heinrich L.A. Gr zu Dohna angefangen den 11ten Sept. 1791 // 10ten Maerz 1792*. The *Akademie* edition does not include this notebook. The *Akademie* editors have concluded that this notebook actually draws from as many as seven different sets of notes: the 1772/73 *Collins*, *Parow*, and *Brauer* (not included in the *Akademie* edition), the ~~1781/82 *Men*~~*schenkunde*, the 1784/85 *Mrongovius*, and the 1788/89 *Reichel* and *Matuszewski* (both not included in the *Akademie* edition). As the introduction to volume 25 of the *Akademie* further explains, approximately 40 percent of the 90,000 word notebook (or what would have comprised about one hundred pages in the *Akademie* edition, had it been included) is nearly identical to the 1772/73 *Collins, Parow* and *Brauer* notes. This material begins with the section, *Vom Witze Scharfsinnigkeit und*

important detail come from Kant's previously available, pre-*Critique*, minor works, correspondence, lecture notes, and personal notes, many of which remain untranslated. Because most of these sources went unpublished during Kant's lifetime and so are not always as reliable as Kant's published sources,[8] and because this period of Kant's work is far less familiar to most readers than that from 1781 onwards, I often quote from a range of these sources to support any given point, and refer the reader to even more. This minimizes the risk of presenting as Kant's considered position from this period something that is an anomaly, something taken out of context, or, in the case of Kant's lecture notes, something reflective of only the textbook being used in Kant's class or student error in recording what Kant said.[9] Moreover, so that this analysis may serve as a self-standing background for interpreting the *Critique*, I generally avoid references to the *Critique* and later recorded thought, other than for the purpose of general observations or to suggest a way of approaching Kant's earlier work.

1. REJECTING KITCHER'S EMPIRICIST HISTORICAL THESIS

Kitcher recognizes the important interpretive value of historical context when approaching Kant's notoriously ambiguous account of the self in the *Critique*'s Deduction and Paralogisms chapters. She accordingly notes that commentary on Kant's account of the self has suffered by ignoring this historical context, and here she cites as the most important failing that of not registering a claimed connection of Kant to Hume on personal identity (see, e.g., *KTP*, 95, 99, 120,

---

Urteilskraft, and so does not include Kant's discussions of the nature of the soul and our manner of epistemic access to it (which come much earlier in the notes) that will figure prominently in the interpretation that will be offered here of Kant's early account of the self. This information comes from the introduction by Reinhard Brandt and Werner Stark, and the bibliography, of volume 25 of the *Akademie* edition. (25:cxxxiii–cli, 1565–1658)

[8] For discussion of the various sources consulted in this paper and their reliability, see notes 7, 9, 18, 19, 20, 48, and 63.

[9] While state law mandated that Kant's lectures follow one of a number of texts, from which Kant chose Baumgarten's *Metaphysica*, which is heavily rationalist, circumstantial evidence speaks strongly against dismissing Kant's lectures as not directly indicative of his own philosophical position. For one, Kant made his willingness to bend or depart from Baumgarten's text clear in many ways. First, in *M. Immanuel Kant's Announcement of the Programme of his lectures for the Winter Semester 1765–6* (1765), Kant announces that until the time when he can lecture from his own book (*Inquiry*, 1763), "I can easily, by applying gentle pressure, induce A.G. Baumgarten, the author of the textbook on which this course will be based—and that book has been chosen chiefly for the richness of its contents and the precision of its method—to follow the same path [as in the *Inquiry*]" (*Pr*, 2:308). Kant continues to assert a minimal connection between his lectures on anthropology and Baumgarten's textbook in 1781/82, when he explains, in his *Menschenkunde* lectures on anthropology, that "Because there is no other book on anthropology, we will accordingly choose as our guiding thread the metaphysical psychology of Baumgarten, a man who was very rich in material, and very short on execution" (*Me*, 25:859). Kant again characterizes the role of Baumgarten's textbook in his lectures as a mere "guiding thread" and nothing more in his 1784/85 *Anthropology Mrongovius*, explaining that "Baumgarten's empirical psychology is because of its ordering the best guiding thread and only the ordering of the materials and chapters is adhered to in this anthropology" (*AM*, 25:1214). For more on his project in anthropology, see note 18; for more on the reliability of these notes, see note 19; for discussion of the importance of the anthropology lectures in relation to the metaphysics lectures, see note 48; and for more on whether Kant presented his own views in his lectures, see note 63.

and 127),[10] blaming this partly on Kant for never explicitly conceding a debt to Hume on this point:

> Whatever the reason, [Kant's] habit of not placing his discussions in their historical context has enormous potential for misleading later scholars. Imagine trying to understand the Aesthetic's metaphysical claims about space and time outside the context of the Newton-Leibniz debate! In the case of Hume and apperception, that potential has been realized. (*KTP*, 97)

Kitcher sees her own interpretation as breaking from this mold with a new focus on the history behind Kant's account of the self in the *Critique*: "My interpretation of the subjective Deduction depends on the historical claim that the doctrine of apperception is, in part, a reply to Hume's skepticism about personal identity" (*KTP*, 97).

The importance of Kitcher's historical thesis to her interpretation of the *Critique* quickly becomes apparent. She views Kant's account of the self in the Deduction as nothing more or less than his "reply" (e.g., *KTP*, 97, 100, 128) or "solution" (*KTP*, 121) to Hume's problem of personal identity, and she accordingly rejects interpretive options that conflict with this historical thesis and selects from the remaining options those that best conform with it. A crucial interpretive principle falling out of this historical analysis is that Kant must accept only those starting premises acceptable to the empiricist Hume: "To answer Hume, [Kant] must appeal to some fact about mental life that his predecessor takes for granted" (*KTP*, 108).[11] Historical sensitivity for Kitcher thus takes us to Kant and Hume's shared

---

[10] Lothar Kreimendahl's *Kant – Der Durchbruch von 1769* (Köln: Jürgen Dinter Verlag für Philosophie, 1990) documents in impressive detail the influence of Hume on Kant. Kreimendahl greets with appropriate skepticism Kitcher's thesis that it was Hume's thoughts on personal identity in the *A Treatise of Human Nature* [*Treatise*] that inspired Kant's views on the self and the transcendental unity of apperception, which for Kreimendahl would in effect place Hume's view on personal identity at or at least near the origin of Kant's critical philosophy. But Kreimendahl, like Allison, Pippin, Ameriks, Kitcher, and many others, assumes that Kant's conclusions that the soul is a simple substance are incompatible with his Paralogisms. On this basis he argues that Kitcher's thesis suffers because neither of the two dates that Kitcher offers for Kant's introduction to Hume's account of personal identity squares with our understanding of the evolution of Kant's views, i.e., neither 1772 (if the introduction came by way of Beattie) nor 1777 (if by way of Tetens). Regarding the former date, Kreimendahl argues that it is too early for what is generally taken to be the date of the beginnings of Kant's views on the Paralogisms, although here Kitcher might reply that Kant need not have actually *read* this translation until a few years later. Moreover, Kreimendahl argues, this early date for the beginnings of the Paralogisms is too early because it comes before Kant's conclusions that the soul is a simple substance in his lectures on metaphysics in the 1770s, which conclusions Kreimendahl sees as inconsistent with the Paralogisms (for my response to this concern, see note 44). Kreimendahl adds to these points another, with which I agree, namely that the later date considered by Kitcher (which would for Kreimendahl avoid his criticisms regarding the earlier date) would be too late because it would have Kant prompted toward his Deduction a few years after he had already been working on this problem (here Kreimendahl is likely referring to Kant's *Reflexionen* in the *Duisburg Nachlass* from 1774–75). Kitcher might counter here that Kant had not yet arrived at the particular formulation of his account of the self that was responsive to Hume's challenge on personal identity, but I think that this argument would ultimately not work for the simple reason that what Kant argues at this earlier point is sufficiently similar to what he argues in the *Critique* that there is no basis for assuming an intervening change of heart prompted by Hume's challenge to personal identity.

[11] On this basis, Kitcher dismisses the analysis of the Deduction offered by Robert Paul Wolff in his "Kant's Debt to Hume via Beattie," *Journal of the History of Ideas* 21 (1960): 117–23, asserting that "[Wolff] offers four different reconstructions of the Deduction argument (on pp. 116, 119, 132, 161) and each one begins with a premise that all my representations are bound up in a unity—a premise that begs the question against Hume" (*KTP*, 249–50 n. 29).

empiricism. This is naturally of central importance to Kitcher's interpretation of Kant's account of the self in the *Critique*, and it is crucial for an interpretation of Kant's account of apperception because empiricism rules out a place for Kant's account of pure, original, or transcendental apperception, which has no place within an empiricist account of the self because it is a type of apperception that for Kant "precedes all data" (*CPR*, A 107) or "preced[es] all particular experience" (*CPR*, A 117 n.) and so by definition is devoid of all matter, and hence of anything that corresponds to sensation. Kant contrasts this *pure* apperception with another type of apperception, *empirical* apperception, or inner sense, which is a consciousness of all of our mental states as they appear to us in time by means of introspection; i.e., in Kant's words, the "consciousness of oneself in accordance with the determinations of our state in internal perception is merely empirical, forever variable; it can provide no standing or abiding self in this stream of inner appearances, and is customarily called *inner sense* or *empirical apperception*" (*CPR*, A 107). Kitcher does not confront this consequence of her historical thesis explicitly, but indeed her extended analysis of Kant's account of apperception—within her interpretation of Kant's account of the self and its identity—never once mentions the former, nonempirical type of apperception (i.e., pure apperception). Kant is accordingly understood to construe personal identity as something attained without recourse to pure apperception, as the result of the contentual interdependence of mental content of empirical apperception alone. On this reading, Kant establishes a relation of existential dependence among mental states denied by Hume by arguing that mental states are nothing more than the effects of earlier mental states, and that mental states are therefore connected if the earlier mental states serve as a necessary and sufficient condition for the later mental states.

Surprisingly, Kitcher's empiricist historical thesis never considers any of Kant's recorded thought from before 1781, but instead limits itself to the questions of whether *Treatise* passages on personal identity were available to Kant before the *Critique* and whether these passages bear a unique resemblance to Kant's account of the self in the *Critique*.[12] Kitcher succeeds in her first aim, demonstrating that the 1772[13] German translation of James Beattie's *Essay on the Nature and Immutability of Truth* contained key passages from Hume's attack on personal identity in the *Treatise*.[14]

---

[12] Kitcher also claims in passing that "[a]t the time Kant was writing the *Critique*, Hume's views on personal identity were known in Germany" (*KTP*, 98). But Kitcher provides no evidence for this assertion, and, even if granted, especially in such vague terms, it would certainly not address whether Kant formulated his account in response to Hume, much less whether he would choose to honor the empiricist starting point of Hume's argument. Kitcher instead relies on her thesis of a unique resemblance between the *Critique* and Hume on personal identity to ground these historical claims.

[13] Kitcher dates the 1772 translation to 1776.

[14] Although translated excerpts from Hume's attack on personal identity in the *Treatise* were available to Kant only as of 1772, the German translation of Hume's *An Enquiry Concerning Human Understanding* (originally published as *Philosophical Essays Concerning Human Understanding*) was available much earlier. Kant's collection of Hume's works, *Vermischte Schriften* (*Collected Works*) (4 vols., Hamburg), which contained a translation of Hume's *Enquiry*, was published in 1754–56. Kant's earliest reference to these works of which I am aware comes in Kant's *Programm*, from 1765–66; but there he mentions only Hume's ethics. In *Dreams of a Spirit-Seer*, from 1766, however, Kant quotes nearly verbatim part of Hume's discussion of mind-body interaction located within a broader discussion of causation in the *Enquiry*. (Compare Kant's "That my will moves my arm is no more intelligible to me

But the more important question is whether Kant's account of the self as it emerges in the *Critique* is *uniquely* connected to Hume's *Treatise* passages on personal identity (as related in Beattie's book). Here Kitcher compares the language of Kant's views in the *Critique* to the Beattie passages, but not to Kant's own pre-*Critique* or pre-1772 (the year of the translation of Beattie's book) sources, or to other philosophers' works. The Beattie passages include Hume's hallmark claims that "[w]hat we call a mind is nothing but a heap or collection of different perceptions (or objects) united together by certain relations …" (*KTP*, 99), and that "[i]f anyone, upon serious and unprejudiced reflection, thinks he has a different notion of himself, I must confess I can reason with him no longer" (ibid.). Having reduced the mind to a heap of perceptions, Hume concludes that "[t]here is properly no simplicity in the mind at one time, nor identity in different (times) … They are successive perceptions alone that constitute the mind..." (ibid.).

Even if we limit ourselves to the passage from the *Critique* that Kitcher repeatedly selects as the best evidence of a unique connection between the *Critique* and Hume on personal identity, we find that this passage is more closely connected to Kant's own earlier material (from the early 1760s) than to Hume's thought on personal identity. Kitcher thus compares these Beattie passages to a passage from *CPR*, A 107 of Kant's Deduction as proof of their unique connection. As translated by Paul Guyer and Allen Wood, the A 107 passage reads as follows:

> The consciousness of oneself in accordance with the determinations of our state in internal perception is merely empirical, forever variable; it can provide no standing or abiding self in this stream of inner appearances, and is customarily called **inner sense** or **empirical apperception**. That which should **necessarily** be represented as numerically identical cannot be thought of as such through empirical data. There must be a condition that precedes all experience and makes the latter itself possible, which should make such a transcendental presupposition valid. (*CPR*, A 107, bold in original)

---

than someone's claiming that my will could halt the moon in its orbit" [*DSS*, 2:370] to the passage in the first edition of Hume's *Enquiry*: "Were we empower'd, by a secret Wish, to remove Mountains, or controul the Planets in their Orbits; this extensive Authority over matter would not be more extraordinary, nor more beyond the Bounds of our Comprehension [than the influence of our soul over our body]" [*Philosophical Essays Concerning the Human Understanding* {Zürich: Georg Olms Verlag, 1986}], 107). See also Kant's famous reference in the *Prolegomena* [*Pro*, 4:260] to Hume's role in awakening him from his dogmatic slumber. Here Kant leaves the date of this awakening unclear, however. For a different, and potentially conflicting account of the reason for his awakening, see Kant's 1798 letter to Christian Garve, 12:257–58.) Although Hume does not broach the problem of personal identity as such in the *Enquiry*, he provides hints of this account when he explicitly draws a parallel between, on the one hand, our disparate thoughts as lacking a necessary connection to one another, and, on the other hand, disparate objects in the world as lacking a necessary connection to one another (Essay 8). It is accordingly plausible that Kant read Hume's *Enquiry* long before 1772 and recognized the implications for the notion of the unity of our thoughts and thus personal identity. Moreover, we cannot rule out the possibility that Kant was introduced to Hume's 1739 *Treatise* attack on personal identity by word of mouth, or some other means, before 1772. It is also worth noting here that Kant was aware of Locke's account of personal identity in 1762–64, as indicated by Kant's explicit references to that account in *Metaphysik Herder* (*MH*, 28:71, 683). This will be discussed in further detail below. Kant also read Leibniz's response to Locke's *An Essay Concerning Human Understanding* [*Essay*] (ed. P.H. Nidditch [Oxford: Clarendon Press, 1975]), the *New Essays on Human Understanding* [*New Essays*] (eds. Jonathan Bennett and Peter Remnant [Cambridge: Cambridge University Press, 1982]), soon after its 1765 publication.

Kitcher sees the wording of this passage as conclusive, uniquely resembling Hume's: it is "strikingly reminiscent of Hume's discussion" (*KTP*, 100), "virtually repeats Hume's description of the absence of a self in introspection" (*KTP*, 102), and uses language that "clearly echoed Hume's."[15] On these grounds, then, she concludes that "Kant knew about Hume's attack on the self" (*KTP*, 98); that he "certainly knew of it" (*KTP*, 108); and that Kant accordingly shaped his *Critique* account of the self mainly in response to this attack by Hume on personal identity.

Had Kant claimed that there is no form of apperception other than empirical apperception, like Hume, this claim at A 107 that there is no self in *empirical* apperception would indeed imply that Kant, like Hume, thought that there was no self in apperception *at all*. But, as noted, the *Critique* speaks of *two* types of apperception—empirical apperception and *pure* apperception—and so Kant's claims about empirical apperception do not yet exhaust his views on what apperception might have to offer. Indeed, even the A 107 passage specifically restricts to one species of apperception, empirical apperception, its claims that we there encounter no self, by no means taking this for an absence of a self in *all* apperception, which was Hume's unique contribution. Here Kitcher's historical analysis is further compromised by her own translations of A 107, which eliminate Kant's note that his views there (regarding the lack of a permanent or continuing self) concern, in particular, his account of *empirical* apperception, or *inner sense*.[16] The

---

[15] This last quotation is from Kitcher's comments on my "Nonreductionism and Personal Identity in Kant's Early Philosophy, 1761–1781" [""Nonreductionism and Personal Identity"] (Central Division Meeting of the American Philosophical Association, 2001), 2.

[16] Here are Kitcher's translations, into which the missing text is reinserted in the place of her ellipses and boldfaced. First:

> The consciousness of the self, according to the determinations of our state in inner perception, is merely empirical, and always changing. There can be no permanent and continuing self in this flux of inner appearances**, and is customarily called inner sense or empirical apperception**. What has necessarily to be represented as numerically identical cannot be thought as such through empirical data. To render such a transcendental presupposition valid, there must be a condition which precedes all experience, and which makes experience itself possible. (*KTP*, 102)

And again, in the same source:

> The consciousness of the self, according to the determinations of our state in inner perception, is merely empirical, and always changing. There can be no permanent and continuing self in this flux of inner appearances**, and is customarily called inner sense or empirical apperception**. (*KTP*, 100)

Next, in her comments on my "Nonreductionism and Personal Identity":

> consciousness of self **in accordance with the determinations of our state** in inner perception is **merely empirical,** always changing. No fixed and abiding self can present itself in **this stream of** inner appearances**, and is customarily called inner sense or empirical apperception.** (2; see also *KTP*, 187)

Kitcher also refers to a passage from *CPR*, B 133–34 once in *Kant's Transcendental Psychology* (100) and once in her comments on my "Nonreductionism and Personal Identity." This passage only repeats in even greater detail the emphasis of *CPR*, A 107 on the fact that it is specifically inner sense, or empirical apperception, where there is constantly flux and no identity. Kitcher's bracketed paraphrase of a skipped section of the passage acknowledges this emphasis on inner sense in particular as the location of the problem of personal identity, although it is worth noting that this skipped section also speaks of an identical consciousness as a prerequisite for the unity of our representations, arguing

resemblance between Kant's and Hume's accounts of the problem of personal identity is thus confined to their shared view that there is no identical self among our mental accidents, which for Kant together comprise empirical apperception, or inner sense.

Is this resemblance sufficient to establish a unique connection historically between Kant and Hume on personal identity? Kant's own pre-*Critique* thought shows us that it is not. Kant instead recognized the flux and lack of an identical self among the mental accidents of empirical apperception about one decade before the translation of Beattie's work. In particular, notes on Kant's lectures on metaphysics from between 1762 and 1764, taken by Johann Gottfried Herder, already contain language that more closely resembles Kant's language, even in this A 107 passage, than does Hume's. In the *Metaphysik Herder* passage, Kant uses the same metaphor as the *CPR*, A 107 passage, namely of a *streaming*, or *flowing*. And more importantly, just like the *Critique's* A 107 passage, and unlike Hume's *Treatise*, *Metaphysik Herder* is very clear to limit the scope of this problem of a streaming or flowing to one *species* of apperception: it is a streaming or flowing specifically of our *accidents* or *determinations*, which, for Kant, comprise the *empirical* self of *empirical* apperception, just as in the *Critique*. Compare the following passages. First the *Metaphysik Herder* passage:

> The soul is always *flowing*, always other *accidents*—(because of the preceding, by larger changes not to be punished) … (*MH*, 28:145, my emphasis)

Next, the A 107 passage:

> The consciousness of oneself in accordance with the *determination* of our *state* in *internal perception* is merely *empirical*, forever variable; it can provide no standing or abiding self in this *stream* of *inner appearances*, and is customarily called ***inner sense*** or ***empirical apperception*** (*CPR*, A 107; bold in original, my italics).

Kant's A 107 claim that the specific location of the problem of flux is in the empirical self is accordingly mirrored by the *Metaphysik Herder* passage from nearly two decades earlier.

The parallels between Kant's accounts of the self in the *Critique* and *Metaphysik Herder* extend to the *solution* to the stated problem of the changing nature of the soul. In *Metaphysik Herder,* Kant makes clear that although the accidents of inner sense change, they can nonetheless be united in one self, and that we find this source of unity of the accidents of inner sense in something *other* than these accidents and their nature. These accidents need to inhere in a common subject, which we are conscious of being even though we do not intuit ourselves in any determinate manner. Kant thus explains that, despite the larger changes that can occur in our constantly flowing accidents, which can limit moral culpability for

---

that "it is only because I can combine a manifold of given representations in one consciousness that it is possible for me to represent the identity of the consciousness in these representations itself" (*CPR*, B 133), and that "only because I can comprehend their manifold in a consciousness do I call them all together my representations" (*CPR*, B 134). Here is Kitcher's quote: "For the empirical consciousness, which accompanies different representations, is in itself diverse and without relation to the identity of the subject....[If inner sense were the only means of cognizing a self], I should have as many-colored and diverse a self as I have representations of which I am conscious to myself [*CPR*, B 133–34]" (*KTP*, 100; Kitcher's brackets; see also *KTP*, 187).

earlier acts, "Still one reproaches oneself because of what has preceded" (*MH*, 28:145). What explains this mistaken sense of guilt for our past acts, according to Kant? It is not some sort of contentual interdependence between our present accidents and those at the time of our moral shortcomings, but instead our awareness of the inherence of these accidents in a common subject, which Kant identifies as a substance (of a sort to be explained below) on the same page (ibid.). We err in focusing on our identity, as this common subject, with the earlier subject, rather than on the fact that this identity obtains *even though* we have changed in a morally significant way: "Here there is a large confusion between the changes of personality and *still the subject remains (just as for each substance the substantial* [as opposed to the accidents] *remains through all changes)*" (ibid., my emphasis). In the *Critique*, in a passage not quoted by Kitcher, but on the same page from which she quotes, Kant clearly underscores precisely this contrast, between the changing states of inner sense, in *empirical* apperception, and the *unchanging* consciousness of ourselves, in *pure*, transcendental apperception, to solve the problem of the changing self in inner sense. This contrast holds even if we grant that in the *Critique* this pure, unchanging apperception of ourselves is not an indeterminate consciousness of a substantial self:

> Now no cognitions can occur in us, no connection and unity among them, without that unity of consciousness that *precedes all data* of the *intuitions*, and in relation to which all representation of objects is alone possible. This *pure, unchanging* consciousness I will now name transcendental apperception. (*CPR*, A 107, my emphasis)

In their attempts to explain how it is possible that our disparate accidents could be unified, Kant's *Critique* and *Metaphysik Herder* thus share a strategy of appealing to a distinction between an unchanging pure apperception (and perhaps an underlying, unchanging substantial self, of which we have an indeterminate consciousness) and a changing self of empirical apperception. At the same time, this distinction between pure apperception and empirical apperception places Kant's account of the self in the *Critique* directly at odds with Hume's account of personal identity. For Hume did not distinguish between empirical apperception and pure apperception, but instead famously reduced the self to a "bundle . . . of . . . perceptions" in empirical apperception, which for Kant is merely the empirical self. The main overlap between Kant's and Hume's accounts is thus the generic recognition of the problem of a lack of identity found in our changing mental states—a recognition shared by Locke and Leibniz as well.

And this brings us to another point weighing against the empiricist historical thesis that there is a unique connection between Kant's account of the self in the *Critique* and Hume's account of personal identity. Both before and after the *Critique*, and indeed in the 1762–74 *Metaphysic Herder* notes just considered, Kant freely acknowledges a connection of his thoughts on personal identity to another philosopher's, but this philosopher is Locke and not Hume. This evidence helps us to make sense, on the one hand, of the closer resemblance between Kant's views on personal identity before and after 1772 than between Kant's and Hume's views on personal identity, but, on the other hand, of the presence of a common, basic

Lockean concern (in both Kant's and Hume's accounts of personal identity) about the identity of a thing over time, despite the streaming of its particular states.[17]

Thus, even a brief consideration of the history of Kant's own recorded thought undermines the view that there is a unique connection of Kant's account of the self in the *Critique* to Hume's account of personal identity, and reveals a stronger connection to Kant's own nonempiricist thought from more than a decade before the translation of Beattie's work, and a stronger connection, as well, to the problem of personal identity as presented in Locke's work. As we will see in the more detailed investigation of these pre-*Critique* sources below, the *Metaphysik Herder* and *Critique* view that we have two distinct types of apperception—one of fluctuating accidents and one which is unchanging—is presented by Kant repeatedly throughout the two decades leading to the *Critique*. But before more fully considering this historical evidence supporting a new, positive historical thesis, we will consider evidence from Kant's pre-*Critique* material, not considered by Kitcher, which might appear to support the sort of empiricist version of Kant's account of the self defended by Kitcher in which Kant's self includes only the particular contents of empirical apperception, to the exclusion of both pure apperception and a substantial self of which this might be a pure apperception.

Consider first two passages from *Anthropology Parow* (1772/73[18]): "…the soul is an object of inner sense …" (*APa*, 25:243); and, "The first thought that arises

---

[17] *Metaphysik Herder* provides the first passage drawing this connection to Locke. There Kant recalls Locke's discussion (*Essay*, II.xxvii.14), in his account of personal identity, of a reputable man (with whom Locke had met) who claimed that his soul had once been Socrates': "In Locke's time an honorable, learned man who had never lied, presented himself with great certainty as Socrates, who can disprove this?" (*MH*, 28:43; also *MH*, 28:71, 683). The second example comes from the notes on Kant's lectures on metaphysics approximately thirty years later. There Kant once again invokes the metaphor of the flowing nature of an identical thing over time, observing that "[t]hus identity is not at all to be doubted morally, but (physically) theoretically one cannot assume it, as little as the water in a river always remains the same" (*MD*, 28:682), and in the margins of these notes is written "Locke Essay" (Locke *Versuch*). What is more, even after the *Critique*, Kant connects his account of the flowing accidents of inner sense in relation to a constant underlying substance not only to Locke's more general and evocative metaphor of the flowing of a river, but also to Locke's account of accidents and substances, which Kant sees as consistent with his own view (to be discussed further below) in that it holds that accidents are not separate existences, but instead merely the states, or modes, of substances as we perceive them. In *Metaphysik L2*, from approximately 1790/91, Kant thus explains that "Accidents are the ways to think of the existence of a thing, and are not different existences; just as Locke says, that the substance is a carrier of accidents, which is why it is also called substratum" (*ML2*, 28:563).

[18] In a late 1773 letter to Marcus Herz, Kant describes the nature and objectives of his current course on anthropology and makes clear his intellectual ownership of this course:

> This winter *I* am giving, for the second time, a lecture course on anthropology, a subject that *I* now intend to *make* into a proper academic discipline. But *my plan is unique*. *I* intend to *disclose* the bases of all sciences, the principles of morality, of skill, of human intercourse, of the method of molding and governing men, and thus of everything that pertains to the practical. *I* shall seek to discuss the phenomena and their laws rather than the possibility of modifying universal human nature. (*C*, 10:138, my emphasis. See also 10:126–27; *Pr*, 2:309.)

Kant held his first course on anthropology only a year earlier, but his *Programm* specifies that his forthcoming 1765/66 course on metaphysics, logic, ethics, and physical geography will commence (in its section on metaphysics) with an investigation of empirical psychology. An account of empirical psychology appears even earlier, in the 1762–64 *Metaphysik Herder*, although these notes pick up abruptly, just beyond the section on self-consciousness that begins both Baumgarten's account of empirical psychology and Kant's later accounts in his anthropology and metaphysics lectures.

for a human *with the use of his inner sense*, is the I" (*APa*, 25:244, my emphasis). The relevance of the first passage is obvious, while the relevance of the second passage becomes apparent only when considered in tandem with notes from a different portion of the same lecture, taken by a different student: *Anthropology Collins* (1772/73) records Kant as saying, "The first thought that confronts us, *when we view ourselves*, is expressed by the I; it expresses the observation of ourselves" (*AC*, 25:10, my emphasis). Comparing the italicized passages, we see that the phrase 'with the use of his inner sense' and the phrase 'when we view ourselves' play the same role, of specifying the occasion of our first thought. Assuming that these notes and their discrepancy reflect Kant's thinking, and not student error,[19]

Kant's anthropology lectures quickly became his most popular course and took on a very public dimension, as he was aware, and so it would have been very odd for him to have nevertheless presented Baumgarten's teachings in a perfunctory manner. For example, in 1778, while lecturing in Berlin on Kant's logic and metaphysics, Herz requested of Kant a copy of notes on Kant's lectures on logic and his lectures on metaphysics, which included a section on empirical psychology, and Kant was more than happy to attempt to accommodate him by acquiring satisfactory student notes (10:240–43). Herz and Moses Mendelssohn were acquaintances who, in 1770, were reported by Mendelssohn to have "almost daily conversations," and were still in contact just before (10:211–14; 10:230–33) and after (e.g., 10:266–67; 10:268–70) Herz's requests for the lecture notes; and Mendelssohn had surprised Kant by attending two of his lectures, on August 18, 1777 (10:211–14), a year earlier. It was, therefore, obvious that Kant's views on empirical psychology might reach Mendelssohn, for whom he had tremendous respect, and yet Kant never distanced himself from these lectures. (And, for that matter, Kant did not distance himself, in his letter to Herz, from the lectures—on an undisclosed subject matter—that Mendelssohn had attended, although he bothered to express regret that they were largely devoted to summarizing what Kant had lectured on at an earlier point, before a vacation. See 10:211–14, and Arnulf Zweig's note in the translation, 164.) Instead, Kant's letter to Herz underscores Kant's commitment to the project (*C*, 10:242). Among others, Johann Fichte (25:cli), (likely) Gottlieb Schlegel (25:cxxxix), and Karl Leonhard Reinhold (25:cxlii) each obtained notes of these lectures. For more on the reliability of notes on Kant's lectures, see notes 9, 19, and 63.

[19] Regardless of whether his lectures reflected his own philosophical position, there is still the additional issue of whether the students' notes that we have are reliable transcriptions. On this point, Kant offers mixed signals. On the one hand, in his response to Herz's request for a set of lecture notes, Kant calls it a "matter of luck whether one has attentive and capable students during a certain period of time" (10:241), and in a later letter Kant complains:

> However, there are many difficulties in carrying out the commission you gave me. Those of my students who are most capable of grasping everything are just the ones who bother least to take explicit and verbatim notes; rather, they write down only the main points, which they can think over afterwards. Those who are the most thorough in note-taking are seldom capable of distinguishing the important from the unimportant. They pile a mass of misunderstood stuff under what they may possibly have grasped correctly. Besides, I have almost no private acquaintance with my auditors, and it is difficult for me even to find out which ones might have accomplished something useful. (10:242)

On the other hand, Kant says of any notes on his lectures on metaphysics that he may be able to procure that it may be *difficult* for a discerning reader of these notes to understand *exactly* what Kant meant in his lectures, even though he recognizes the especially difficult nature of these lectures, due—importantly—to the manner in which he himself has "worked up" these lectures over the years: "metaphysics is a course that I have worked up in the last few years in such a way that I fear it must be difficult even for a discerning head to get *precisely* the right idea from somebody's lecture notes" (10:241, original emphasis). In contrast to his metaphysics, however, Kant earlier had told Herz of his lectures on anthropology that

> Hence the subtle and, to my view, eternally futile inquiries as to the manner in which bodily organs are connected with thought I omit [in these lectures on anthropology] entirely. I include so many observations of ordinary life that my auditors have constant occasion to compare their ordinary experience with my remarks and thus, from beginning to end, find my lectures entertaining and never dry. (10:146)

these passages imply that to use our inner sense *is* to view ourselves. In a letter to Marcus Herz written earlier that same year, on February 21, 1772, Kant appears to draw precisely this conclusion, remarking that "in the case of inner sense the thinking or the existence of thought and my own self are one and the same" (*C*, 10:134). Moreover, in *Anthropology Friedländer* (1775/76) we find Kant similarly maintaining that inner sense in its entirety constitutes an object that is the soul: "the world as an object of inner sense is soul" (*AF*, 25:469). Thus Kant explains that, while physics is the knowledge of objects of outer sense, "knowledge of humans as an object of inner sense, similarly constitutes such a field [of study]" (*AF*, 25:473). Finally, in the *Duisburg Nachlass* from approximately 1774 to 1775, Kant introduces the term 'apperception' and at times links this term with 'self-perception.' Because for Kant perception involves inner sense, this link hints at perception of the self in inner sense.

2. THE POSITIVE THESIS: KANT'S NONREDUCTIONIST CONCLUSIONS IN HIS PRE-1781 ACCOUNT OF THE SELF

While a reductionist, empiricist interpretation of Kant's account of the nature of the soul during the two decades leading to the *Critique* seems plausible based on the passages just quoted about where one intuits the soul, what Kant says *directly* about the nature of the soul quickly makes clear that his is a nonreductionist, non-empiricist account of the self. In *Anthropology Collins* (*AC*, 25:10) and *Anthropology Parow* (*APa*, 25:244–45), Kant claims that our immediate self-consciousness reveals the soul's "simplicity" (*Einfachkeit*), "substantiality" (*Substantialität*), and "freedom" (*Freyheit*), and that the soul is a "rational substance" (*vernünftige Substanz*). Three years later, in *Anthropology Friedländer*, Kant concludes the substantiality, simplicity, and freedom of the soul (*AF*, 25:473–74)—leaving out reference to the obvious rationality of the soul. Two years later, in *Anthropology Pillau*, Kant fails to take a position on the issue, but in *Metaphysik L1* (1777–80),[20] Kant draws these conclu-

In a word, "this pleasant empirical study" (10:146) is, in contrast to Kant's lectures on metaphysics, extremely easy for a notetaker to follow. For more on this topic, see notes 9, 18, 20, and 63.

[20] The history and evolution of *Metaphysik L1* and *Metaphysik L2* are worth reviewing because of the importance of these sources, which provide a broad-ranging and detailed treatment of Kant's empirical and rational psychology. *Metaphysik Pölitz* consists of notes from three courses on metaphysics, probably taught by Kant in the seventies and eighties, and was first published in 1821 by Pölitz (Lehmann, 28:1339–48). In his foreword to that publication, Pölitz explains that these notes derive from two separate sources, thereafter named *Metaphysik L1* and *Metaphysik L2*, respectively, where '*L*' stands for "Leipzig." Pölitz considered the first source, *Metaphysik L1*, older, more thorough, and more instructive than the second source, *Metaphysik L2*. The first source included a prolegomena, an ontology, and a *metaphysica specialis*, comprised of cosmology, psychology, and theology. Pölitz thought that the second source, *Metaphysik L2*, was copied from another, unknown source, in 1788, only to have corrections and marginalia in someone else's handwriting added in 1789 or 1790, probably inspired by a new course by Kant on metaphysics (Lehmann, 28:1511–12); Lehmann, by contrast, tells us that *Metaphysik L2* originated in the winter of 1790/91. The second source, like the first, contained a prolegomena, ontology, and a *metaphysica specialis,* as well as a separate introduction. Pölitz included the *metaphysica specialis* from *Metaphysik L1* in his reader along with the introduction and ontology of *Metaphysik L2*, although Pölitz claims to have edited the material from *Metaphysik L2* with a "thorough comparison and regard for the first, older manuscript [*Metaphysik L1*]" (Lehmann, 28:1512), thereby complicating any attempt to date specific parts of *Metaphysik L2*. The parts of *Metaphysik L1* and *Metaphysik L2* not included by Pölitz in his reader can nonetheless also be found (at least in part) in

sions again not only in his empirical psychology, but also in his rational psychology. Specifically, in his empirical psychology he concludes the soul's substantiality, simplicity, and immateriality (*ML1*, 28:225–26) (and, arguably, the soul's freedom), while in his rational psychology he concludes the soul's substantiality, simplicity, singularity or unity, and freedom (*ML1*, 28:266–68). Even before all of these accounts, in a *Reflexion* on rational psychology from between 1769 and 1770, Kant sets out the substantiality, simplicity, immateriality, and freedom of the soul (*R* 4230, 17:467–69).

Kant's considered view in the 1770s that the soul is simple, unified, free, and substantial, is revealing on the question of the reducibility of the soul to inner sense, because it is also Kant's considered view that inner sense is neither simple, unified, free, nor substantial. First, the complexity, or lack of simplicity, of inner sense is obvious: inner sense is a *collection* of all of our disparate mental states. Moreover, insofar as each of these states is located in time, each is divisible, as Kant explains in *Metaphysik L1*: "… all our given representations happen in time … From this it follows that there is no appearance and no part of a given appearance that could not be divided to infinity; thus there is nothing simple in appearance" (*ML1*, 28:202; see also 204). What is more, without a distinct, simple, unifying subject, these mental states would lack unity (e.g., *APa*, 25:244; *ML1*, 28:266), as we will see later.

Next, if reduced to the states of inner sense, the self would not be free. Instead, invoking an argument that will famously reappear in the second *Critique* (*CPrR*, 5:97 [1788]), Kant tells us in *Metaphysik L1* (*ML1*, 28:267) that if we limit the range of possible causes of a state of inner sense to the world understood as an object of inner and outer sense, we can at best exercise a form of "practical," or "psychological," freedom, in the event that our states of inner sense do not follow with necessity from stimuli, but are instead determined by our antecedent decisions. But even with such practical freedom, if we limit the range of possible causes to objects of inner and outer sense, the "inner principle" that would regulate the flow of events would presuppose an "external principle" that determined this inner principle outside of the subject, in the same manner in which the "inner principle" that regulates a watch's internal movement presupposes an "external

---

Kant's *Gesammelte Schriften*, in the form of Heinze's synopsis of and extracts from the prolegomena and ontology of the now missing *Metaphysik L1,* and the *metaphysica specialis* of *Metaphysik L2*, which follows the extant original.

The date of the lecture course from which the notes of *Metaphysik L1* were taken is much disputed. Benno Erdmann argued for the time around 1774; Emil Arnoldt estimated 1778/79, or more probably between 1779/80 and 1784/85; Heinze the second-half of the seventies; and Menzer 1778/79 or 1779/80 (Lehmann, 28:1340–47). Briefly (and with little argumentation), the 1774 date seems very unlikely because of the maturity of Kant's transcendental deduction, compared to the *Duisburg Nachlass* of 1774/75 and the relative maturity and clarity of Kant's terminology in empirical psychology compared to the 1772/73 anthropology lectures and even the 1775/76 *Anthropology Friedländer*. On the other hand, the post-1780 dates seem all but ruled out by the immaturity of Kant's transcendental deduction—in the Heinze extracts (28:186), Kant specifies the categories of quality, quantity, and relation as the only categories and only thereafter catalogues the judgment and concepts of modality, although note here Kant's recognition of the special status of the categories of modality, even in the *Critique*, at A 74/BB 99–100, even though Kant here recognizes them as categories—and the remaining rationalist sympathies relative to the first *Critique*. Accordingly, it will tentatively be assumed that the *Metaphysik L1* lectures date from between 1777 and 1780.

principle," a watchmaker, who determines the inner principle of the watch from outside the watch. Accordingly, although a practical, or psychological freedom may obtain if stimuli do not dictate our actions, this freedom would be *secundum quid talis* (*ML1*, 28:267), i.e., determined immediately by the earlier states of inner sense according to an "inner principle," and mediately by the "external principle" responsible for the "inner principle." This freedom is accordingly not the freedom that Kant attributes to the soul, which is a *transcendental* freedom (e.g., *ML1*, 28:267), where the inner principle is supplied from within, not from without. Instead, again anticipating his words in the second *Critique* (*CPrR*, 5:97), Kant asserts that this is the freedom enjoyed by a "turnspit" (*ML1*, 28:267).

Finally, and most importantly, the states of inner sense are what Kant terms "predicates," "accidents," or "modes" of the soul that has them (e.g., *ID*, 2:405 [1770]), while in Kant's view the thinking subject can never be reduced to its predicates, accidents, or modes. To understand why—and why Kant also rejects the reduction of the self as substance to its powers, or to a combination of accidents and powers—we must consider more closely Kant's understanding of the concepts of substance, power, and accident. Activity is at the root of Kant's definition of noumenal substance in his pre-1781 (and post-1781) philosophy.[21] A noumenal substance, and the soul as its paradigmatic instance, is what Kant refers to as a "real subject,"[22] "first subject,"[23] "last subject,"[24] or a *substantiale*.[25] A real subject, first subject, last subject, or *substantiale*, unlike an exclusively logical subject, and unlike accidents and powers (*ML1*, 28:261)—with which Kitcher equates Kant's self in some of her works (e.g., *KTP*, 117–18)[26]—serves as the *ground* for the existence of accidents through the activity of its powers:

> The distinction between a logical and real subject is this, that the former contains the logical ground for the setting of a predicate, the latter the real ground (something different and positive), and is thus the cause, while the accident is the effect. This relationship of the cause we pull from our own actions and apply it to that, which is constantly in the appearances of outer things. But we ultimately find that everything about the object [of appearance] is accidents. The first subject is therefore a something, through which the accidents exist … The reality of accidents is something that we judge on the basis of sensation of an object, which [reality] is different from that of the subject. (*R* 4412, 17:536–37 [1771])

While we could use concepts to refer to the accidents of inner sense so that these concepts would be predicates of the subject in question, nonetheless, because this subject actually grounds the existence of the accidents that the predicates refer to (unlike a merely logical subject, which is a mere concept, as in the case of the concept of a triangle, which as such does not ground the existence of it predicates through its activity, and which can, in turn, serve as the predicate of another subject), this subject is not a merely logical subject but instead also a *real*, or *first subject*, and thus a *substance* (see, e.g., *R* 4700, 17:680 [1773–75]).

---

[21] *MH*, 28:26; *DSS*, 2:328; *R* 4056, 17:400 [1769]; *R* 4679, 17:663, 664 [1774–75].

[22] *NM*, 2:202; *R* 4240, 17:474 [1769–70]; *R* 4412, 17:536–37 [1771].

[23] *R* 4412, 17:536–37 [1771]; *R* 4699, 17:679 [1773–75].

[24] *R* 3921, 17:346 [1769]; *R* 4052, 17:398–99 [1769]; *AC*, 25:10.

[25] *R* 4413, 17:537 [1771]; *R* 5292, 18:145 [1776–78].

[26] For more on Kitcher's characterizations of Kant's self, and whether she reduces Kant's self to nothing but representations or to nothing but representations and powers, see note 6 above.

On the one hand, then, Kant holds that as the merely phenomenal manifestations of a substance, accidents are not existences separate from the substance of which they are modes. Kant is accordingly quick to flag the potentially misleading language of the "inherence" of accidents in substance, which might suggest that accidents are existences distinct from the substance in which they inhere: "the accidents are not particular things, which inhere in the subject, but instead predicates of a subject, i.e., ways that the subject exists. The concept of inherence is a logical aid …" (*R* 3783, 17:292 [1764–8]). On the other hand, though, Kant repeatedly emphasizes that substances and accidents are nonetheless distinct, one being the cause, the other the effect, although an effect understood only as the phenomenal appearance. In the section on cosmology in *Metaphysik L1*, Kant makes this point both in reference to the world as the whole of all substances and in reference to humans as particular substances:

> The world is a *totum substantiale*; thus a whole of accidents is no world. Accidents are also not regarded as parts, not as *compartes* of the whole. E.g. *thoughts and motions do not belong to humans; rather those are parts of the condition, but not of the whole.* The world-whole is thus not a whole of all conditions, but rather of all substances. (*ML1*, 28:195, my emphasis)

Our soul as substance is to our thoughts what our body is to our motions and what the substance of the world is to its condition. Substance is thus the underlying agency whose predicates are merely its phenomenal expressions, not its parts (*R* 3921, 17:345 [1769]). Kant summarizes these observations concerning the relation of thoughts as accidents to the thinking subject as the substance in *R* 3787, from between 1764 and 1766: "… thus an *effect* cannot be a part of the *cause* and belong, with the cause, as a part of the same whole. *The thought is not a part of the human, but instead its effect*" (*R* 3787, 17:292, my emphasis).[27]

Just as Kant's claim that the soul is a substance rules out its reduction to its *accidents*, so too it rules out the reduction of the soul to a combination of *accidents and powers*—a form of reductionism that Kitcher at least sometimes attributes to Kant's *Critique*,[28] partly on the basis of her historical analysis. The reason is straightforward: Kant defines powers as nothing more than *relations*, the relations of substances to accidents. This makes powers dependent on the existence of substance for the creation of accidents and for the relation between the substance and its accidents that results, which renders the notion of the existence of powers without substance incoherent. In the following passage from *Metaphysik L1*, Kant summarizes these points and derides as incoherent the notion of powers without underlying substances, linking such a way of thinking about powers as existing without an underlying substance to the *rationalists*—although Kitcher attributes this view to *Kant* on the basis of what she sees as a Humean, empiricist historical impetus:[29]

> Wolff [the Leibnizian] assumes one basic power [of representation] and says: the soul itself is a basic power which represents the universe. It is already false when one says:

---

[27] See also *R* 4052, 17:398–99 [1769–70]; *APa*, 25:244–45; *MH*, 28:25, 26, 52; and *ML1*, 28:225–26.

[28] See note 6 above.

[29] See note 6 above.

> the soul is a basic power. This arises because the soul is falsely defined, as the Ontology teaches. **Power is not what contains in itself the ground of the actual representation, but rather the relation of the substance to the accident**, insofar as the ground of the actual representation is contained in it. *Power is thus not a separate principle, but rather a relation* [*respectus*]. **Whoever thus says: the soul is power, maintains that the soul is no separate substance, but rather only a power, thus a phenomenon and accident**. (*ML1*, 28:261, my emphasis in bold)[30]

Moreover, Kant's definitions of substance and power render coherent the notion that we, as substances, have *more* than one power (insofar as we are not reduced to a power and, as substances, have more than one fundamental sort of relation to our accidents). Indeed, Kant argues that we have three fundamental powers[31]—the powers of representation, pleasure, and desire[32]—explicitly rejecting the rationalists' position here, which maintained that we have just one power, of representation.[33] Kant's conclusion of the *simplicity* of our soul therefore rules out an interpretation of his account of the self that would have him reducing the soul to its powers, or to its powers and accidents, just as it was shown to rule out a reduction of the soul to its accidents, because we have *three* basic powers (or relations to our accidents), not one, so that the soul is not simple in this respect.

Kant's rejection of the rationalist conflation of substances and powers, and in turn his rejection of the conflation of the soul as a substance with one of its powers, is also central to his views on the ontology of the soul because it eventually leads him to yet further distance his account of the soul's substantiality and simplicity from the rationalists' account by leading him to reject their conclusions of the permanence and incorruptibility of the soul as a simple substance. As a consequence, Kant is able to simultaneously hold, without contradiction, the conclusions of the soul's substantiality and simplicity alongside his rejection of the *rationalists'* conclusions of the soul's substantiality and simplicity. Although our focus is on Kant's positive ontology of the soul and his arguments for this positive ontology—which have for so long been overshadowed by Kant's negative ontology as he presents it in his attack on the rational psychologists' ontology of the soul in the Paralogisms of the *Critique*—a brief overview of Kant's negative ontology of the soul in relation to this positive ontology helps to clarify the meaning of these positive conclusions that might otherwise be taken for identical with those of the rationalists.

First, as we have seen, Kant rejects the rationalists' definition of 'substance,' arguing that substance and power are distinct and that the rationalists conflate them. Because powers are, for Kant, nothing more than any irreducible relation of a substance to its accidents, substance can for him have as many powers as there are irreducible relations of this substance to its accidents, even if the substance is

---

[30] See also *MH*, 28:25, 145; *R* 4762, 17:720 [1775–78].

[31] Kant usually refers to these three powers as "faculties." The term 'faculty' refers to our *ability* to exercise a certain power or set of powers, and to accordingly effect a certain type of relation between our substance and our accidents, and for this reason each faculty can also alternately be referred to as a power (e.g., *R* 3583, 17:72 [1776–79]; *R* 3585, 17:73 [1780–89]).

[32] E.g., *ML1*, 28:228–59, 262; *MH*, 28:117, 902, 145; *R* 3630, 17:134 [1762–63]; *APa*, 25:367.

[33] Consider here a passage from *Metaphysik Herder*: "The Wolffians assumed falsely that the soul qua simple is just a power of representation. This arises though a false definition of power: because it is just a respect, thus can the soul have many respects. As many types of accidents there are, that cannot be reduced to another … A representation and desire are basic powers" (*MH*, 28:145 [1762–64]).

simple. The rationalists, by contrast, are forced to assume that a simple substance has only one power, the power of representation. This allows Kant to draw a distinction in kind between understanding and sensibility in the 1770 *Inaugural Dissertation* (2:392–93, 398–406), whereas the rationalists—because they equate substance with power and thus a simple substance with something having only one power—are forced to view understanding and sensibility not as powers that are distinct in kind, but instead as distinct only logically and in degree of clarity as part of one power, of representation. The rationalists accordingly intellectualize appearances, so that sensibility does not for them present to us a phenomenal substance that is different in kind from noumenal substance and so cut off categorically from noumenal substance, but instead merely a confused representation of noumenal substance that, in theory, could be deciphered to yield an understanding of noumenal substance. Kant, by contrast, distinguishes between two kinds of substance: "noumenal substance," which (as we have seen) is for him a thing that exercises powers by means of which it grounds accidents; and "phenomenal substance," which is located in a spatiotemporal framework foreign to the noumenal realm, and which, as such, is fully cut off from the noumenal realm regardless of the distinctness of the representation of this phenomenal substance (*R* 4316, 17:504 [1770/71]).

Already in 1769, Kant had recognized that we must assume the permanence of substance for the possibility of doing philosophy (*R* 4105, 17:416), but beginning with the *Duisburg Nachlass*, from approximately 1774 to 1775, Kant maps this requirement of permanence onto his distinction between noumenal substance and phenomenal substance (*R* 4675, 17:648 [1774–75]). Anticipating the *Critique's* First Analogy (*CPR*, A 181/B 225–A 188/B 231; see also *Pro*, 4:335), Kant now maintains that the assumption of the permanence of substance is necessary in order to temporally situate individual appearances and transform them into experiences (*R* 4675, 17:650 [1774–75]; *R* 4682, 17:669 [1774–75]; *ID*, 2:418–19). The assumption of permanence accordingly only accrues to *phenomenal* substance,[34] which Kant argues is merely what we take to be a spatiotemporal analogy to noumenal substance, whose paradigm comes to us by means of immediate consciousness of ourselves as noumenal substances. This phenomenal analogy of noumenal substance is that *in appearance* that seems to act as a substratum for what is changing in appearance due to subjective factors (e.g., the order in which we view something, the angle from which we view it, etc.),[35] and thus is that which is permanent in space and time (and so in appearance) that underlies our particular appearances. Because Kant distinguished the soul as substance from power as early as 1762–64, he had already (prior to the 1774–75 *Duisburg Nachlass*) concluded that it was incoherent to think of the soul as inhering in God, and had concluded that we need not partake in God's necessary existence in the future, but that we could

[34] *R* 4765, 17:721 [1775–77]; *R* 5358, 18:160 [1776–78]; *R* 5312, 18:150 [1776–78]; *R* 5287, 18:143 [1776–78]; *R* 6001, 18:420–1 [1780–89]; *R* 5454, 18:186 [1776–79]; *R* 4054, 17:399 [1769]; *R* 4060, 17:401 [1769]; *NM*, 2:202; *R* 3879, 17:323 [1764–68]; *DSS*, 2:321 n.; *R* 4137, 17:430 [1769].

[35] *R* 4494, 17:572 [1772–73]; *R* 4499, 17:574 [1772–73]; *R* 4699, 17:679 [1773–75]; *R* 5290, 18:144 [1776–78]; *R* 5312, 18:150 [1776–78]; *MLI*, 28:226, 204, 209.

instead, as dependent substances, be annihilated at any moment by God.[36] Kant recognized that noumenal substance is therefore permanent only relative to its phenomenal accidents, though nothing that we know about it precludes that it could someday cease along with its phenomenal accidents.[37]

Following the *Duisburg Nachlass*, Kant accordingly consistently ties (tautological) permanence to *phenomenal* substance alone,[38] while continuing to hold that noumenal substance, including the soul as noumenal substance, is not permanent.[39] While Kant therefore believes that the soul is a substance in the noumenal sense, he does not think that it is a substance in the phenomenal sense, which implies permanence, and he therefore parts with the rationalist tradition, which does not properly distinguish between noumenal substance and phenomenal substance and thus mistakenly assumes the soul's permanence as noumenal substance. Because the permanence of the soul as substance is required for the rationalists' sole end in the field of psychology—as Kant sees it in *Metaphysik L1* (*ML1*, 28:263), and as he sees it in the Paralogisms (*CPR*, A 349, A 351) and *Prolegomena* (4:334), namely the conclusion that the soul is immortal—Kant already rejects rational psychology when he rejects the permanence of noumenal substance.

The other necessary precondition for the soul's immortality is its incorruptibility, and here Kant's distinction in kind between noumena and phenomena again leads him to reject the rationalists' position, namely their inference to the determinate, phenomenal simplicity of the soul on the basis of its noumenal simplicity. Thus Kant argues against Leibniz and Wolff that the soul's simplicity does not imply its simplicity as a physical point in space[40] because, for one, there is no such thing as a physically, as opposed to mathematically, simple point in space,[41] and, second, we do not know whether or how the soul might appear in space, so that it might appear as a complex body.[42] The incorruptibility accruing to the notion of a simple physical point therefore does not accrue to the soul by virtue of its noumenal simplicity, and in fact the soul may be corruptible in appearance as well as corruptible by means of evanescence at the noumenal level.[43]

Having rejected the rational psychologists' overweened conclusions of the soul's substantiality and simplicity as implying the soul's permanence and incorruptibility, Kant in turn rejects their ultimate conclusion, grounded in these conclusions, of the soul's immortality. He rejects all theoretical arguments for immortality other than a distinct—and non-rationalist—argument from analogy (*R* 4106, 17:417 [1769]), and he has already abandoned even this sole theoretical argument—along with all other theoretical arguments for the soul's immortality—by the time of *Reflexion* 4557 (17:593 [1772–76]), there maintaining that the soul's

---

[36] *R* 4054, 17:399 [1769]; see also *R* 4060, 17:401 [1769].

[37] *R* 5454, 18:186 [1776–79]; *R* 4702, 17:680 [1773–77].

[38] *R* 4765, 17:721 [1775–77]; *R* 5358, 18:160 [1776–78]; *R* 5312, 18:150 [1776–78]; *R* 5287, 18:143 [1776–78]; *R* 6001, 18:420–21 [1780–89].

[39] E.g., *R* 5454, 18:186 [1776–78]; *R* 4702, 17:680 [1773–77].

[40] *ML1*, 28:272–73; *R* 4499, 17:574 [1772–73]; *R* 4500, 17:574–75 [1772–73].

[41] E.g., *ML1*, 28:202, 208; *R* 4066, 17:402 [1769]; *R* 4316, 17:504 [1770/71].

[42] E.g., *R* 4718, 17:685–86 [1773–75]; *DSS*, 2:321; letter to Mendelssohn, April 8, 1766, 10:72; *ML1*, 28:225, 272, 280.

[43] *R* 4239, 17:473 [1769–70]; *R* 4556, 17:592 [1772–76].

immortality can be defended on the basis of practical arguments alone (see also *ML1*, 28:289–90).[44]

In addition to ruling out the reduction of the soul to its thoughts, or to its thoughts and powers, Kant's characterization of the soul as a noumenal substance, simple, unified, and free also explains the passages considered earlier in support of Kitcher's reductionism. Kant's references to the soul as an object *of* inner sense do not imply that the soul is an object *in* inner sense, reducible to the accidents of inner sense. Instead it means that the soul is the object referred to by the accidents of inner sense, serving as their noumenal ground for existing so that, in this respect, the soul is an object of inner sense. A number of Kant's personal notes from the same period support this interpretation. They show that Kant expected no one to interpret his allusions to the soul as an object of inner sense as implying that the soul was nothing but the sum of these accidents in inner sense. Instead, Kant would have expected everyone to assume that the soul could not have accidents without being a substance, so that the assumption would be that Kant's claims here that the soul is an object of inner sense mean that the accidents of inner sense refer to the soul as its object and that this object underlying the accidents or predicates of inner sense is a substance, or a last subject. *R* 3921 maintains that one cannot help but assume that predicates imply not only a predicated subject, but a *last* subject: "predicates without a subject and without a last subject *cannot be thought*" (*R* 3921, 17:346 [1769], my emphasis). Similarly, *R* 4052 asserts the ineluctability of our conclusion of an underlying, last subject: "Because logical subjects can in turn be predicates, so *we are necessitated in accordance with laws of reason*, to think of a last subject; this is substance" (*R* 4052, 17:398–99 [1769–75], my emphasis; see also *R* 5295, 18:145–46 [1776–78]). Given Kant's conviction that we assume the existence of a distinct substance underlying any predicates or accidents, Kant would have taken it for granted that his students or readers would not misinterpret any ambiguous references to the accidents of inner sense as the "I" as the claim that the "I" is *nothing but* the accidents of inner sense. Instead, he would have assumed that they would see it as a reference to the soul—as that object referred to by these accidents and underlying them.

---

[44] As outlined in note 10, Kreimendahl rejects both Kitcher's later (1777) proposed date for Kant's introduction to Hume's account of personal identity (and purported movement from this account toward his own Deduction and Paralogisms), because Kant had already begun work toward the Deduction before then, as well as Kitcher's earlier (1772) proposed date because it precedes *Metaphysik L1*, in which Kant still argued for the soul's simplicity and substantiality. We can now see why there is not necessarily a conflict between these views in *Metaphysik L1* on the nature of the soul and Kant's critique of rational psychology of the sort found in the Paralogisms. Moreover, even if there is a conflict of this sort, one does not avoid it by moving the date of Kant's exposure to Hume's views on personal identity forward to a point in time after *Metaphysik L1* and closer to the first *Critique*. For Kant continues to assert these ontological conclusions regarding the soul in and after the *Critique* (e.g., *CPR*, A 350–51, A 356, A 365–66, A 399–403; *Pro*, 334–35; *R* 6001, 18:420–21 [1780–89]; *MMr*, 29:770–71, 772, 904, 912; *MvS*, 28:511; *MFNS*, 4:543–44; *MD*, 28:682; *MK2*, 28:754–55; *RP*, 20:270; *What Real Progress*, *Loße Blätter*, 20:340; and *MVi*, 29:1032). This does not in itself, of course, rule out that Kreimendahl and other advocates of the traditional reading of the Paralogisms could reconcile such later passages with the traditional understanding of the Paralogisms, but such a reconciliation would likely then extend to *Metaphysik L1*, as well, so that, again, there would be no need to assume that Kitcher's starting point would have to postdate *Metaphysik L1*.

3. KANT'S PSEUDO-EMPIRICAL IMMEDIATISM

Is, then, reason's demand that predicates inhere in a last subject also the basis for Kant's nonempiricist, nonreductionist conclusion that the soul is not just accidents and powers but also a substance? The rationalists took this tack, and, as we will see later, it was in the process of doing so that they embellished (in the manner reviewed above) this conclusion of substantiality as well as that of simplicity with determinate phenomenal predicates. But just as Kant rejects the rationalists' intemperate understanding of the meaning of the conclusions of the soul's substantiality and simplicity (according to which the soul is permanent as substance and incorruptible as a simple substance), and just as he cites the rationalists' blunders of conflating substance and power, understanding and sensibility, and noumenal and phenomenal substance when they infer to the soul's permanence and incorruptibility, so too he rejects their reliance on inference *per se* as a means by which to arrive at stronger conclusions than immediate consciousness alone can provide. In their place, and leading to his own temperate conclusions, are two "immediatist" arguments (as I call them): one pseudo-empirical and the other transcendental. While the former (which we will examine in this section) argues that we simply *have* an indeterminate *immediate* consciousness, *immediate* feeling, or *immediate* intuition of being a simple substance, the latter (which we will look at in the next section) argues transcendentally that both our simplicity and substantiality, as well as our immediate consciousness of this, are in fact necessary for personal identity and thought—although the transcendental argument is carefully limited in its role to illuminating what immediate consciousness already reveals, rather than supplying any new proofs.

Kant thus alludes to the rational psychologists and outright rejects their claim that one *infers* these conclusions regarding the soul, arguing instead that we reach these conclusions on the basis of an *immediate* epistemic access, which he alternately refers to as our immediate consciousness of ourselves,[45] our immediate feeling of ourselves,[46] or our immediate intuition of ourselves.[47] In *Anthropology Collins,* Kant asserts that "that, which many philosophers present as profound conclusions, are nothing more than immediate intuitions of our self" (*AC*, 25:10), clarifying a few pages later that he here has Descartes in mind (*AC*, 25:14); and, in a note from 1769, Kant maintains that "the I … nonetheless is no concept, but an intuition" (*R* 3921, 17:346 [1769]). Kant contrasts these various forms of immediate access to ourselves not only with arguments that draw inferences on the basis of no observations (*a priori* arguments), but also with arguments based on empirical observations which themselves presuppose an inference, to their objects: "for I could have representations of bodies, even if there were no bodies there; but I intuit myself, I am *immediately* conscious of myself" (*ML1*, 28:224).[48]

---

[45] E.g., *R* 175, 15:64 [1769]; *AF*, 25:475, 477; *APi*, 25:736; *ML1*, 28:224, 265.

[46] E.g., *R* 175, 15:63–64 [1769]; *ML1*, 28:206, 265.

[47] *AC*, 25:14, 32; *ML1*, 28:225–26; *R* 3921, 17:346 [1769].

[48] The overlapping content of Kant's anthropology lectures in the 1770s and his *Metaphysik L1* raises the question whether the former really contribute anything new to our knowledge of Kant's pre-*Critique* thought. In his October 20, 1778 letter to Marcus Herz, Kant explains that "I now handle empirical psychology [as part of his lectures on metaphysics] more briefly since I lecture on *Anthropologie*"

As noted, common to our "consciousness," "intuition," and "feeling" of ourselves is the *immediacy* of these forms of access. Here we are aware of *being* a substance, i.e., of *being* a thing that exercises powers that ground accidents. This is the only case in which we stand in this immediate relation to a substance, and this case thus provides us with the *paradigm* of noumenal substance as a thing that grounds accidents through the exercise of its powers. *Reflexion* 3921 (1769) makes both of these points:

> Therefore it is also no wonder, that *we know no subject before all predicates, other than the I, which nonetheless is no concept, but rather an intuition.* Therefore through the understanding we know about bodies not the real subjects, but rather the predicates of extension, solidity, rest, movement, etc. The cause is: through our senses only the relations of things can be revealed, and we can represent the absolute or subject only on the basis of extrapolation from our selves. *The idea of substance actually comes from the representation of our selves* … (*R* 3921, 17:345–46 [1769], my emphasis; see also *R* 5297, 18:146 [1776–78])

R 5294 (1776–78) underscores this point that self-consciousness provides the paradigm of substance:

> It is ridiculous to want to think of the soul in corporeal terms; *for we have the concept of substance only from the soul, and that of the body we model after it* [*bilden wir uns darnach*]. (*R* 5294, 18:145, my emphasis)

And a passage from *Metaphysik L1* also emphasizes, in striking terms, that only in one case—that of our relation to ourselves—do we have access to noumenal

---

(10:242); and indeed the empirical psychology of *Metaphysik L1* is a mere forty-one pages, substantially less than the anthropology notes of the 1770s, which average about 220 pages. These numbers are misleading for our analysis, though, because *Metaphysik L1* packs most of Kant's most important and controversial conclusions on the soul into these mere forty-one pages, as well as into the subsequent thirty-nine pages on rational psychology, making *Metaphysik L1* an invaluable resource for understanding Kant's account of the self despite its brevity. Still, as Karl Ameriks and Steve Naragon point out in their introduction to *The Cambridge Edition of the Works of Immanuel Kant*, *Lectures on Metaphysics*:

> only a few years ago it was something of a risk for a scholar to invoke these lectures [on metaphysics, other than *Metaphysik Mrongovius* and *Vigilantius (K3)*] as backing for major interpretive claims. But unexpected good fortune brought the discovery and publication in 1983 of two additional sets of Metaphysics lectures - *Mrongovius* and *Vigilantius (K3)*. These especially detailed lectures have provided further confirmation of the general independence and reliability of the various sets of notes. (xiv)

The anthropology lectures offer even greater detail regarding the nature of the self, and their great number (seven), and the fact that they span two decades, helps them to not only flesh out our understanding of Kant's theory of the self, but also to corroborate one another and the notes on Kant's lectures on metaphysics. But because *Metaphysik L1* was also the only set of notes, other than the incomplete 1762–64 *Metaphysik Herder* notes, from Kant's philosophical period before 1781, setting it off from the others; because the date of *Metaphysik L1* is highly controversial (beyond the near consensus that they come from before 1781), far more so than that of any other set of notes on metaphysics; and because, in Ameriks' words, *Metaphysik L1* stood out even among the sets of notes on metaphysics for its "especially striking passages...that still show a strong sympathy for rationalist arguments about the soul's substantiality, simplicity, and freedom" (*Lectures on Metaphysics*, xvi), *Metaphysik L1* was not buttressed by the other sets of lectures on metaphysics as they were by one another, leaving its status as a transcription of what Kant had actually said and of what Kant had himself believed as more open to question than the status of the other sets of notes. For this reason, the corroborating evidence now supplied by the four sets of detailed notes on anthropology lectures from the 1770s is especially important for *Metaphysik L1*, and in turn for our understanding of Kant's account of the self before 1781.

substance (or what Kant here refers to as the 'first subject' and the 'substratum'), so that this instance provides us with the paradigm of substance:

> But this I is an absolute subject, to which all accidents and predicates can be attributed, and which cannot at all be a predicate of another thing. Thus the I expresses the substantial; for that substratum, in which all accidents inhere, is the substantial. *This is the only case where we can intuit substance immediately. We can intuit the substratum*[49] *and the first subject of no thing; but in myself I intuit substance immediately. The I therefore expresses not only substance, but also the substantial self. Yes, what is yet more, the concept, that we have of all substances generally, we have borrowed from this I. This is the original concept of substance.* (*ML1*, 28:225–26, my emphasis)

Having already seen how Kant simultaneously concludes the soul's substantiality and simplicity while rejecting the rationalists' definition of substance and their conclusions of the soul's permanence as substance and incorruptibility as a simple substance, we can now consider how Kant is able to offer his immediatism alongside his transcendental idealism, as in his 1777–80 *Metaphysik L1*. In particular, *Metaphysik L1* offers Kant's view that space and time are merely the forms of intuition (a view that goes back to his 1770 *Inaugural Dissertation*[50]); his discursivity thesis (a thesis that also goes back to the *Inaugural Dissertation*, e.g., at *ID*, 2:394, in those respects that are relevant to Kant's views that the conclusions regarding the self are derived not by means of inference of reason but from an immediate consciousness, intuition, or feeling of the self[51]); and his list of the categories (new to *Metaphysik L1*, but excluding those of modality).

The compatibility of these doctrines with Kant's immediatism goes back to the unique place of our immediate relation to ourselves within the constraints of transcendental idealism. The basic thought behind the subtleties of Kant's transcendental idealism are offered by Kant in his *principium generale commerci*, here explained in a *Reflexion* from the mid–1770s: "the *principium generale commerci* is: all influence in the world is in part the effect of the active in the passive, in part the resistance of the latter" (*R* 4704, 17:681). Because they are merely modes or accidents (as construed phenomenally) of the substance and not existences

---

[49] The absence of the term 'noumenon' in reference to ourselves may seem noteworthy coming from Kant's writings from 1781 on, but it is not noteworthy in the context of Kant's early thought: Kant almost never uses the term 'noumenon' prior to the *Critique of Pure Reason*—with his 1770 *Inaugural Dissertation*; *R* 4563, from between 1772 and 1776; *R* 4757, from between 1775 and 1777; and *R* 5553, from 1778/79 (but possibly later) providing the lone exceptions of which I am aware. This dearth of references need not imply a lack of commitment to the contrast between phenomena and noumena during the 1770s, however. Even in the first *Critique*, there is not (to my knowledge) a single reference to "noumenon" or "noumena" prior to the chapter "On the ground of the distinction of all objects in general into *phenomena* and *noumena*," beginning 235 pages into the first edition and 294 pages into the second edition (and there is no reference in either preface, either), and few afterwards. Other than in this section, Kant almost exclusively uses the terms 'thing in itself' (e.g., *CPR*, A 30/B 45, A 360), 'things in themselves' (*CPR*, A 36/B 52), 'substance in itself,' 'anything in itself' (*CPR*, A 30/B 45), 'object in itself' (*CPR*, B 69), 'objects in themselves' (*CPR*, A 30/B 45, A 34/B 51, A 36/B 52), 'soul [understood as] a thinking being in itself' (e.g., *CPR*, A 360), 'soul itself' (e.g., *CPR*, A 22/B 37), and 'thinking being in itself' (*CPR*, A 360). In all of these instances, the context makes it clear that Kant is speaking of something that is not located in space or time. The dominance of the latter terms also prevails throughout the rest of Kant's recorded thought after 1781.

[50] *ID*, 2:392–93, 398–406. See also *R* 4507, 17:577 [1772–73].

[51] As Kant maintains at *R* 3921, 17:345–46 [1769]; *R* 175, 15:64 [1769]; *AC*, 25:10, 14; *AF*, 25:475, 477; *R* 4493, 17:572 [1772/73], *APa*, 25:736; and *ML1*, 28:206, 224, 265.

distinct from the substance, the states of a substance can never simply be shuttled from one substance to another (even by God; see *MH*, 28:52), like a book from one bookshelf to the next. Instead one substance can influence another only by changing its state, not by imparting to it its own state; and here the manner in which the affected substance receives the impression of the affecting substance helps to determine the resulting state of the affected substance. Kant illustrates this point with the example of the action of fire on two different bodies, wax and clay: "e.g., wax held by fire melts, and clay dries. Thus the difference lies here in the bodies, how they are affected" (*ML1*, 28:207). By extension, also in accordance with the *principium generale commerci*, Kant observes of our faculty of sensibility that how we are affected by other substances will in part reflect the sort of resistance that our faculty of sensibility offers:

> *We know nothing more of things than the manner in which we are affected by them; but not what is in the things* … The senses prove only *the manner of the contact* by the appearances in me … I myself intuit myself, but bodies only as they affect me. But this manner does not teach me the properties of things. (*ML1*, 28:206–07; see also *ID*, 2:392)

Our immediate relation to ourselves as substances is thus unique because it obviates the problem inherent in our relation to other substances (that we only experience the substances' effect on us, where our receptive activity intervenes) by being the lone instance in which we are aware of *being* a substance, rather than of being *affected by* a substance; and this privileged perspective on substance thus provides us with the paradigm of substance as something that has powers by means of which it grounds accidents. At the same time, because this underlying *substantiale* is distinct from all of its accidents, it is empty of all accidents. Accordingly, our consciousness of being this predicate-free *substantiale* provides us with no content on the basis of which we could conclude *anything determinate* about this *substantiale*; and so Kant is clear to deny to immediate self-consciousness anything that he will later call 'knowledge,' which demands determinate content.

While reason may demand an end to the regress to the unconditioned cause of accidents, so that we naturally assume that substances underlie accidents, Kant's point is thus that in one case—that of our relation to ourselves—no inference is needed. Instead, this immediate consciousness of ourselves and our activity reveals our substantiality, and in the process it supplies us with the paradigm of such a thing that both starts a causal series (in this respect being a "first subject") and, in turn, ends the regress to the first cause (thus being a "last subject"), by virtue of being a substance, or real subject, with powers that it exercises. *Reflexion* 4412, already quoted in our discussion of the difference between a real and merely logical subject, combines a number of these observations:

> The difference between a logical and real subject is this, that the former contains the logical ground for setting a predicate, the latter the real ground (something different and positive), and is thus the cause, while the accident is the effect. We pull this relation of cause from our own actions and apply it to that, which is consistently in the appearances of outer things. We find, however, finally that everything on the object [of appearances] are accidents. The first subject is thus a something, through which the accidents are. (*R* 4412, 17:536–37 [1771])

4. KANT'S TRANSCENDENTAL IMMEDIATISM

Kant supplements his pseudo-empirical immediatist argument for the nonreductionism of the soul—based on our immediate consciousness of ourselves as simple, identical, noumenal substance—with a transcendental argument. This transcendental argument maintains that rational beings are necessarily simple, identical substances distinct from their thoughts, and that they necessarily have the described immediate awareness of themselves as simple, identical substances distinct from their thoughts and identical over time. Ironically, Kant explicitly ties this transcendental argument for a nonreductionist, nonempiricist account of the self to the same basic concern over personal identity that Kitcher argued must have pushed Kant toward a reductionist, empiricist account, and which concern over personal identity she believes was first introduced to Kant only a decade later (than the first of Kant's responses to this concern) and by Hume (through Beattie).

Of course, before we consider the specifics of this transcendental argument, an obvious question that confronts us regarding Kant's transcendental approach to the same conclusions that he has already approached from another tack, i.e., from immediate consciousness and, generally, from within what he terms 'empirical psychology,'[52] is how this transcendental methodology, which Kant also presents in *Metaphysik L1* as part of his own *rational* psychology, is any different from that of the rationalists. We already saw that Kant's *conclusions* of the soul's simplicity and substantiality stand as a rejection of the rationalists' identical-sounding-but-completely-different conclusions of the soul's simplicity and substantiality, because Kant's conclusions do not imply incorruptibility and permanence. And we saw that Kant undermined the rationalists' arguments toward their inflated conclusions by rejecting their linchpin assumptions along the way (i.e., their conflation of substance and power, their reduction of the soul's powers to one, their intellectualization of sensibility, their conflation of the noumenal and phenomenal, and finally their conflation of noumenal and phenomenal substance). But we also saw that Kant rejected the rationalists' methodology more broadly, dismissing as self-important nonsense their claims that inferences and concepts are needed to arrive at these conclusions, rather than mere immediate consciousness (*R* 3921, 17:346 [1769]; *AC*, 25:10). And yet now we find Kant, seemingly in spite of this broader point regarding methodology, moving toward his own conclusions regarding the soul by means of an exploration of the concept of the "I" within what he classifies as a rational psychology (*ML1*, 28:263–66).

A closer look at Kant's argument eliminates any appearance of impropriety, however, as we will see in what follows. In a nutshell, Kant argues that the ideas of reason come from the categories of relation, and that, because immediate consciousness of ourselves supplies us with the paradigm of the categories of relation, so too it supplies us with the paradigm of the ideas of reason. Kant accordingly recognizes the redundancy of any exercise attempting to arrive at new insights beyond what immediate consciousness has supplied by recourse to inferences of reason from our accidents back to ourselves as substance when, after all, immediate consciousness has supplied reason with its paradigm from the start.

[52] *AC*, 25:10–15; *APa*, 25:243–47; *AF*, 25:473–76; *APi*, 25:735–37; *ML1*, 28:224–28.

Accordingly, Kant makes very clear that his arguments toward his conclusions of the soul's substantiality and simplicity ultimately rest on only what the testimony of immediate consciousness warrants, and not on any inference, with his arguments doing nothing more than focusing our attention on the lack of alternative ways to think about ourselves than what immediate consciousness offers in the form of the conclusions of our substantiality and simplicity. More specifically, Kant's Virgil argument (see p. 523) presses us to recognize that it is impossible for us to think that representations distributed across persons, or across noumenal substances, could nonetheless be united, underscoring that the manner in which we are conscious of ourselves in immediate self-consciousness is the way in which we must think of all thinking things.

We thus begin with the fascinating structural relation of Kant's account of self-consciousness to his account of the ideas of reason. We have already seen that Kant holds that immediate consciousness of ourselves provides us with the paradigm of noumenal substance[53] and so of the relation of substance to predicates. Elsewhere, as Paul Guyer has observed,[54] Kant makes clear that in all that is real, and so where there is substance, the other two relations obtain as well, of causality and community. Thus *R* 5860, from 1780 to 1789, asserts, "In substance there is 1. The relation of inherence (accidents); 2. Causality (power); 3. Commerce (influence)" (*R* 5860, 18:371 [1780–89]), as does *R* 4496, from 1772–73:

> Three principles. 1. In all that is real there is the relation of a substance to accidents (inherence); 2. of ground to consequence (dependence); 3. of part and combination (composition).
>
> There are therefore three premises: of the subject, of ground and of parts; and three real modes: of inherence, of subordination and of composition. Therefore also three first principles: 1. subject, which is no predicate; 2. Ground, which is no consequence; 3. Unity, which is not (in itself) combined. (*R* 4496, 17:573 [1772–73])

Because Kant repeatedly makes this point that in substance, and so in all that is real, all the relations exist,[55] his additional conclusion is not surprising, that just as the soul supplies us with the paradigm of noumenal substance, so too it supplies us with the paradigm of ground and unity. First, *R* 4412, quoted above, described how we serve as the paradigm of the relation of *ground* by virtue of being a real subject that grounds the existence of its accidents rather than serving as a merely logical ground of them, as does a merely logical subject:

> The difference between a logical and real subject is this, that the former contains the logical ground for setting a predicate, the latter the real ground (something different and positive), and is thus the cause, the accident however the effect. This relation of cause we pull from our own actions and apply it to that which is constantly in the appearances of outer things … The first subject is therefore a something through which the accidents exist. (*R* 4412, 17:537 [1771])

---

[53] *ML1*, 28:225–26; *R* 3921, 17:345–46 [1769]; *R* 5294, 18:145 [1776–78]; *R* 5297, 18:146 [1776–78].

[54] Paul Guyer, *Kant and the Claims of Knowledge* (Cambridge: Cambridge University Press, 1987), 212–14. Guyer also mentions a number of the sources that I have cited in support of the view that where one of the relational categories apply, so too do the others.

[55] *R* 4493, 17:571–72 [1772–73]; *R* 5289, 18:144 [1776–78]; *R* 5292, 18:145 [1776–78]; *R* 5860 [1780–89, maybe 1776–79]; *CPR*, A 204/B 249.

Next, in a series of *Reflexionen* in the *Duisburg Nachlass* devoted largely to the categories of relation, Kant repeatedly asserts that our immediate self-consciousness provides us with the paradigm of *all* relations, maintaining in the first, "Apperception is the perception of one's self as a thinking subject in general," then explaining of apperception that it contains the three relations: "Herein are three exponents: 1. the relation to the subject, 2. the relation of consequences under one another, 3. of combination" (*R* 4674, 17:647 [1773–75]). *R* 4675 next adds that these three relations in the mind supply us with the model on which we build the three analogies in appearance: "The three relationships in the mind thus demand three analogies in appearance" (*R* 4675, 17:648 [1773–75]; see also *R* 4058 [1769]). The subsequent *Reflexion* sets out the sense in which the "I" provides these models of relations—first as substance that exists, then as something that acts and so has powers by means of which it grounds accidents, and finally as something in which disparate accidents are united: "I am, I think, accidents are in me. These are in total relations …" (*R* 4676, 17:656 [1773–75]; see also *ML1*, 28:266).

In addition to arguing that the paradigms of all categories of relation (supplied to us by our immediate consciousness of ourselves) are projected by us onto appearance to yield the three analogies in appearance, Kant argues that the categories of relation are, in turn, behind the formation of reason's ideas. The categories of relation, Kant argues in his pre-*Critique* material (from the same time period as *Metaphysik L1* [1777–80], in which Kant presents this rational psychology), as in the *Critique* (*CPR*, A 323/B 379), respectively specify categorical, hypothetical, and disjunctive relations. These relations are those from which we derive the three ideas of reason, which infer to the unconditioned using the classical categorical, hypothetical, and disjunctive syllogisms. Thus, *R* 5553 (1778/79) argues that

> We can call reason the faculty of ideas. There are ideas of sensibility, also those of pure reason. The latter are either practical or speculative; [of these] the latter are transcendental. These are necessary concepts of reason, for which no object in the senses can be given. *As pure concepts, however, they must be taken from the pure categories.* (*R* 5553, 18:228 [1778/79], my emphasis)

This *Reflexion* goes on to state the obvious—that it is the *relational* categories of which Kant is here speaking (*R* 5553, 18:229 [1778/79])—as does *R* 5555, which asserts that "The concept of the totality of synthesis in accordance with the categories of relation is the pure concept of reason" (*R* 5555, 18:231 [1777–89]). Accordingly, immediate consciousness of ourselves provides us with the model of the relational categories and in turn with a paradigm of the completed relationships of conditioned to unconditioned condition that reason will then look to find emulated elsewhere, although reason will then make the mistake of seeing its ideas as constitutive rather than merely regulative and will infer back from known conditioned conditions in appearance to unconditioned conditions as things in themselves, which are still construed in the determinate terms that properly apply only within the realm of appearance. Thus, corresponding to the relational categories of substance, ground, and community are the three ideas of reason regarding the unconditioned condition on which the *Critique* will focus in the "Paralogisms," the "Antinomies," and the "Ideal of Pure Reason," respectively (*CPR*, A 323/B

379).[56] Indeed, Kant's different labels for the soul, considered above, refer to the manner in which the soul serves as the unconditioned condition by terminating any regress from its representations to itself—calling it (as we have already seen) a "first subject,"[57] "last subject,"[58] "real subject,"[59] or "absolute subject"[60]—even while conceding that our consciousness of serving as the end of a regress does not provide us with an *understanding* of how it is possible that the soul does this.[61]

That Kant believes that the ideas of reason have their paradigm in our immediate consciousness of ourselves is crucial for the purpose of understanding the methodology of Kant's Virgil argument, its relation to the rationalist methodology, and thus Kant's reconciliation of the two. For if immediate consciousness of ourselves provides us with the paradigm of the three ideas of reason by virtue of supplying us with the paradigm of the relations of a substance and its predicates, a ground and its consequences, and community among its states by virtue of their inherence in a common substance, then, while reason can for better or worse apply to other things these models by inferring back from predicates to substances, consequences to ground, and community to a common source, its efforts will be strictly redundant if it applies these models back to the *soul*, inferring from the soul's predicates to the existence of an underlying substance. At best, reason here would do nothing more than retrace a relationship (of accidents to accidents and to substance) that immediate consciousness had already laid bare to us in its entirety—in the process providing reason with its regressive model in the first place. At worst, in the process of this redundant inferential exercise, reason will conflate the phenomenal and noumenal realms and attribute phenomenal qualities to noumenal things, as we will see below in the case of conclusions regarding the soul.

To skip ahead, for a moment, to the *Critique*, to suggest a sense in which Kant is using the Virgil argument a few years prior to the *Critique*, what Kant will argue in the *Critique* regarding the status of the Virgil argument will likewise address the limits of inference with regard to the soul. There Kant will also make clear the futility of any attempt to extract from the inference of the Virgil argument any more than immediate self-consciousness already provides—viewing this exercise in inference as a tired rabbit-from-a-hat stunt. Kant tells us in the Second Paralogism that the second paralogism (i.e., the formalized Virgil argument) has as its "*nervus probandi*" (*CPR*, A 352; i.e. "the nub [literally, 'nerve'] of what is to be proved"[62]) "the proposition that *many representations have to be contained in the absolute unity of the thinking subject in order to constitute one thought*" (ibid. my emphasis); and Kant then asserts that this proposition can be proved neither (a) from *concepts* (*CPR*,

---

[56] In their introduction to the *Critique of Pure Reason*, Paul Guyer and Allen Wood note Kant's claims in the *Critique* (15) and in *Reflexionen* 5553–55 (64–65, 713 n.)—which I have also cited—that the ideas of reason have their origin in the categories of relation.

[57] *R* 4412, 17:536–37 [1771]; *R* 4699, 17:679 [1773–75].

[58] *R* 3921, 17:346 [1769]; *R* 4052, 17:398–99 [1769]; *AC*, 25:10.

[59] *NM*, 2:202; *R* 4240, 17:474 [1769–70]; *R* 4412, 17:536–37 [1771].

[60] *ML1*, 28:225–26.

[61] *R* 4493, 17:571–72 [1772–73] and *R* 4496, 17:573 [1772–73].

[62] Editorial note in the *Critique of Pure Reason* (Cambridge: Cambridge University Press, 1998), 418 n.

A352–53), nor (b) from *experience* (*CPR*, A 353). Starting with (b), we find that the *Critique* specifically isolates for rejection claims to *experience* of the "I" as thing-in-itself, while not challenging—and even explicitly maintaining our ability to have—mere *consciousness* of this "I" as thing-in-itself, as in this passage from the Refutation of Idealism, where Kant tells us that the "I am" is that "which expresses the *consciousness* that can accompany all thinking, [and] … which immediately includes the existence of a subject in itself, but not yet any *cognition* of it, thus not *empirical cognition*, i.e., *experience*" (*CPR*, B 277, my emphasis; cf., *Pro*, 4:334, *CPrR*, 5:97–98). Regarding (a), we see that the Paralogisms observe that the simplicity of the soul, like its existence, is not *inferred* from this expression of apperception, as Descartes believed, but is instead already contained in it:

> the simplicity of my self (as soul) is not really **inferred** from the proposition 'I think,' but rather the former lies already in every thought itself. The proposition **I am simple** must be regarded as an immediate expression of apperception, just as the supposed Cartesian inference *cogito, ergo sum* is in fact tautological, since *cogito* [*sum cogitans*] immediately asserts the reality. (*CPR*, A 354–55, bold and italics in original)

Accordingly, Kant goes on to specify that the conclusion that we are a simple substance is accurate only if we do not extend its meaning beyond *the limits of what self-consciousness alone affords*, so that we must confine ourselves to these conclusions taken in the minimalist sense of the *pure, indeterminate* categories, rejecting the rationalists' fortified, determinate versions that also promise the soul's permanence and incorruptibility (*CPR*, A 355–61, A 348).

What we find Kant doing in his own rational psychology, as presented in *Metaphysik L1*, as well as in *Metaphysik Herder* [1762–4],[63] is adhering to precisely

---

[63] See also Kant's post-*Critique MMr*, 29:904–06; *ML2*, 28:590–92; *MD*, 28:679–81; *MK2*, 28:753–56; and *MVi*, 29:1025–29. Could Kant's views have eventually departed from Baumgarten's so radically that Kant abandoned hope of presenting his own views in courses using Baumgarten's *Metaphysica*? Kant's statements on his use of *Metaphysica* (mentioned in note 9) already undermine this view. Kant also remarks that he improves these lectures each year (*C*, 10:242) and that he has "worked-up" his lectures on metaphysics over the past few years (*C*, 10:241) just two years before completing the *Critique* and so at a time when he had already developed most of his critical tenets. This pre-*Critique* evolution continues long after the *Critique*, although consistency dominates. Most notable, perhaps, are developments in his epistemology, including the addition of the categories of modality (in *Metaphysik L1*, as related by Heinze, Kant lists the categories of quantity, quality, and relation, claims that these are all of the categories, and only subsequently adds that judgments can have the three modalities; contrast here, e.g., *MMr*, 29:801–05, although notice how, in the *Critique* [*CPR*, A 219/B 266], Kant still argues that the categories of modality are different from the rest in adding nothing to the concept to which they are applied); his post-1781, increasingly neat characterization of his own system's place between the flawed empiricist and rationalist extremes (e.g., *MMr*, 29:782); his movement away from theoretical to practical proofs for immortality (*R* 5471, 18:192 [1776–79]; *R* 5472–73, 18:192 [1776–81]; *ML1*, 28:288–91, 295–6; *MVo*, 28:443; *MD*, 28:689; *RP*, 309; cf. the early *R* 4670, 17:634 [1773–78]) and rejection of Mendelssohn's argument for immortality on the basis of the simple, substantial nature of the soul, for the same reasons brought forward in the first *Critique*, namely the possibility of the evanescence of the soul (*MMr*, 29:912–13; *MK2*, 28:763–64); his refutations of idealism after 1781 (e.g., *MK2*, 28:771–72); his account of the beautiful; his continued rejection of Leibniz and Baumgarten's equation of substance with one power (e.g., *MMr*, 29:771; *MD*, 28:671); his rejection of Baumgarten's notion of causing as giving (*MMr*, 29:823; *MK2*, 28:759); his continued rejection of inter-substance causation as understood by Descartes, Leibniz, and Baumgarten (*MMr*, 29:866–69; *MK2*, 28:759; cf. *ID*, 2:409 and Leibniz's "A New System of Nature," in *Philosophical Essays*, eds. Roger Ariew and Daniel Garber [Indianapolis: Hackett, 1989], 142–44); and his continued rejection of Leibniz's theory of monads and its extrapolation from ourselves to other substances (*MMr*, 29:930; *MK2*, 28:758,762;

these guidelines as he lays them out just a few years later in the *Critique*. He presents the Virgil argument as instructive, concluding from it the necessity of the simplicity and substantiality of the soul. (Kant also addresses the substantiality of the soul separately, but again by appeal to immediate consciousness [*ML1*, 28:266].) But his sober analysis homes in on the *nervus probandi* of the arguments for the conclusion that the soul is a substance (ibid.) and the conclusion that the soul is a simple substance (ibid.), and here he cleanly hands off the burden of proving this *nervus probandi* to immediate consciousness of ourselves and thus never moves beyond the indeterminate conclusions of substantiality and simplicity warranted by immediate consciousness, to conclude without adequate proof the soul's permanence and incorruptibility (as do the rationalists).[64]

Of course, it may be asked in what sense necessity obtains for these conclusions if, at the end of the day, they are carried by immediate consciousness. Here the Paralogisms eloquently state what is clear in *Metaphysik L1*'s repeated appeals to immediate consciousness to reveal the necessity of the conclusions of the substantiality and simplicity of rational souls: that if we must necessarily ascribe to ourselves as thinking things certain properties, we must likewise ascribe these necessary attributes *a priori* to all other (finite) thinking things. Kant explains:

> But right at the start it must seem strange that the condition under which I think in general, and which is therefore merely a property of my subject, is at the same time to be valid for everything that thinks, and that on an empirical-seeming proposition ['I'] we can presume to ground an apodictic and universal judgment, namely, that everything that thinks is constituted as the claim of self-consciousness asserts of me. But the cause of this lies in the fact that we must necessarily ascribe to things *a priori* all the properties that constitute the conditions under which alone we think them. Now I cannot have the least representation of a thinking being through an external experience, but only through self-consciousness. Thus such objects are nothing further than the transference of this consciousness of mine to other things, which can be represented as thinking beings only in this way (*CPR*, A 346/B 404–A 347/B 405).

MVi, 29:1033; cf. *IC*, 2:293; *DSS*, 2:328; *R* 4718, 17:685–6 [1773–75]; and from Leibniz, *Philosophical Essays*, 79, 104, 180–81). These developments speak clearly to Kant's continued attempt to deliver what he saw as accurate philosophical views in his lectures on anthropology and metaphysics.

[64] As mentioned above (note 4), James Van Cleve argues that the claim that the soul is a noumenal substance is inconsistent with the First Analogy (*Problems from Kant*, 172–74). The problem, he thinks, is that whereas the first edition of the First Paralogism seems to hold that the unschematized concept of substance applies to the soul because this noumenal concept of substance does not imply permanence, the First Analogy states the opposite: that no noumenal substance can ever pass away. But, as indicated here, the First Analogy does not stand in the way of Kant's account in the First Paralogism, but instead does the opposite, grounding Kant's ability to detect the *sophisma figurae dictionis* in the rationalists' fallacious inference to the soul's permanence, because it shows that the absolute permanence that we must assume pertains only to phenomenal substance. Indeed, Van Cleve himself suggests elsewhere (*Problems from Kant*, 120–21) that the substance in the First Analogy might be only phenomenal substance, but he considers only two passages suggesting this interpretation (*Problems from Kant*, 120), remains uncertain of it, and does not acknowledge it as a possible solution in his later discussion of the conflict, as he sees it, between the First Paralogism and the First Analogy. For reasons I cite earlier (in section 3), I believe that Van Cleve also inaccurately portrays Kant's relation to Spinoza on substance when he suggests that Spinoza's view that God is the only substance and that our thoughts are modes of God is for Kant compatible with a person's being the real thinker of these thoughts (*Problems from Kant*, 174–75). Van Cleve also argues that it is possible that for Kant, as for Spinoza, all substances are necessary beings, but here I believe that Van Cleve fails to distinguish Kant's phenomenal substance, as tautologically permanent, from noumenal substance, which, unlike God, and thus by contrast with Spinoza's view that all substance must be independent, is a dependent substance and thus not necessary in the past or future (*Problems from Kant*, 197–98).

What Kant's Virgil argument thus accomplishes is no more than to focus our attention on the necessity of attributing certain indeterminate properties to the soul, i.e., the "properties that constitute the conditions under which alone we think" of the soul.

Before turning to the Virgil argument in Kant's pre-*Critique* work, it is also instructive (by way of contrast) to consider Kitcher's account of the role that concern over personal identity played for Kant in the years leading to the *Critique*. As we saw, Kitcher argues that Kant formulates his theory of the self to counter Hume's attack on personal identity, or Hume's "denial of any real or necessary connection among mental states," which Kitcher characterizes as "a denial of any *relation of existential dependence among such states*" ("Kant's Real Self," 114). Kant succeeds in this attempt in the *Critique*, Kitcher argues, by showing that thoughts are nothing more than the effects of earlier thoughts, and that earlier thoughts are accordingly both necessary and sufficient conditions for later thoughts, which establishes the relation of existential dependence among them denied by Hume. Two disparate thoughts thus belong to one self if, and only if, the earlier thought serves as a necessary condition for the later thought.[65] This empiricist account of personal identity suffers from what Kitcher admits are "serious deficiencies."[66] If we reduce personal identity to a type of causal—or, as Kitcher calls it, *existential*—relations between mental states, then mental states belonging to two persons in conversation might, for example, need to be understood as belonging to one person because of their interdependence. At the same time, mental states belonging to one person might need to be understood as separated among different persons for lack of their existential interdependence.[67]

As we saw above, long before the 1781 *Critique* Kant describes just such a scenario of fractured representations to spur us into recognizing that we can think of the unity of representations, and so of thought and personal identity, only as this unity is revealed to us through immediate self-consciousness, namely as a unity obtaining by means of the common inherence of our thoughts in one simple,

[65] The earlier need not be sufficient for the existence of the later thought, however, as other earlier thoughts may also be necessary.

[66] See note 10 above.

[67] In *Kant and the Fate of Autonomy* (Cambridge: Cambridge University Press, 2000), 256–62, Karl Ameriks rightly points out that the account attributed to Kant by Kitcher, according to which necessary connections between mental states result from their contentual interdependence, is for all intents and purposes what Hume has already considered when he examines the relation of causality between mental states as the basis for positing a necessary connection between them. Just as Hume can deny that a mental state would not have happened but for a particular antecedent cause, so too can the skeptic deny that a mental state would not have existed but for a particular antecedent existence: both claims of a necessary connection are vulnerable to the same implausible, but equally skeptical challenge, that the latter mental state, as Ameriks puts it, "just happened to 'fly into one's head'" (*Kant and the Fate of Autonomy*, 259). Next, even if we assume that Kitcher's concept of contentual interdependence of mental states succeeds against the skeptic where the concept of causal interdependence fails, we see that, as Kitcher's acknowledgment of the problem of interdependence across persons shows, because Kitcher rejects the view that for Kant unity was tied to an underlying metaphysical self, nothing in her account precludes that this unity would obtain across different substances or persons. Ameriks notes this and then rightly questions why Kitcher does not either stick to her guns and simply accept the lack of true unity in her account and thus accept a Parfitian outcome, or, if she wants more from her account, take the option of a "real, enduring self, a substance" more seriously, granted that it "can still be the natural solution" (*Kant and the Fate of Autonomy*, 261–62).

underlying substance. Thus Kant's 1762–64 *Metaphysik Herder* presents the Virgil argument, which will continue to appear throughout Kant's philosophy into the 1790s and which concludes the simplicity of the soul and the distinctness of the soul from its thoughts, as well as the non-material nature of the soul. Kant maintains that, although a simple *element of* matter[68] can think, it does not follow that *matter* can think. First, he states the question: "Can an *Elementum materiae* think: thus can matter also think?" Next, he responds with the Virgil argument:

> [T]hat does not follow. A whole out of many simple substances that are thinking, thinks first when all the thoughts of each simple substance are unified in it. If each of 100 persons knew a verse from Virgil by heart, would they therefore know all of Virgil by heart? If the latter is not so, thus does matter likewise not think. (*MH*, 28:44; see also *DSS*, 2:328 n. for an abbreviated Virgil argument)

Here Kant confronts us with a scenario in which disparate thoughts are not united in one person and asks whether these thoughts could be united anyway while distributed across persons, pressing us to recognize the incoherence of this alternative. Because matter is defined simply as that which takes up space, it is by definition complex, and so we conclude that *it* cannot think, but instead at best *contains* thinking things, in the same sense in which a group of one hundred people does not think but instead contains thinking things.

Kant's point here strikes at the heart of the sort of empiricist, reductionist view towards which Kitcher believes history pushes Kant. This empiricist approach, as we saw, seeks a solution to the problem of personal identity by reference to the disparate mental states alone (absent any pure apperception or underlying substance), arguing that the specific content of mental states alone can suffice to render them connected in the event that one mental state serves as a necessary (and, perhaps along with other mental states, sufficient) prerequisite for the existence of a later mental state. But here Kant specifically constructs an example where the disparate thoughts under consideration are closely related to one another in content and also in origin, stemming from the same author, Virgil, so that these thoughts are all dependent on Virgil. Nonetheless Kant by no means claims that these thoughts are united through their common dependence on Virgil's thoughts. Instead, this common dependence on Virgil is completely irrelevant to Kant. Kant's point is that the content of thoughts alone is insufficient for uniting them if these thoughts are not, to begin, all contained in an identical, substantial, subject.

One could reply that Kant here simply fails to recognize that one person's (Virgil's) thoughts, in fact, serve as the necessary condition for the others' and that these persons' thoughts are accordingly united in one person; had Kant realized this, he would have concluded that in fact the one hundred people could together have a unified thought of Virgil's work. But approximately fifteen years later, when Kant presents the Virgil argument in *Metaphysik L1*, even though he spells out even more clearly than before the interdependence of the thoughts of the different persons, he still forcefully rejects the possibility that the thoughts of

[68] Referred to here is a *truly* simple substance, which is only an element of matter insofar as it helps to effect an appearance from the outside of matter. Not referred to is something which is simply the smallest part of matter as viewed from the outside, which part does not exist according to Kant, as matter is by definition infinitely divisible.

the different persons involved could be understood as the thought of only one person, and he clearly grounds this conclusion in that which the expression of apperception, 'I think,' alone offers. Here is this remarkable passage:

> Many beings taken together cannot constitute an I. If I say: I think, then I do not express representations which are divided among many beings, rather I express a representation that takes place in one subject. For all thoughts can be only simple or composite. One and precisely the same simple thought can take place only in one simple subject. For if the parts of the representations should be divided among many subjects, then each subject would have only one part of the representation, therefore no single subject would have the whole representation. But for the whole representations to be wholly in a subject, all parts of the representations must also be in one subject. For if they are not connected together in the one subject, then the representation is not whole. E.g., if the saying: whatever you do [*quidquidagis*] etc., were distributed among many subjects so that each had a part; that is, if whatever [*quidquid*] were spoken into the ear of one, you do [*agis*] into that of another, so that no one heard the whole saying, then one could not say: the whole thought is together in the many minds, so that each had a part of the thought; but rather the thought is not at all, because each has only the thought of one word, but not a part of the whole representation. Accordingly many beings can indeed have one and the same thought at the same time, but each has the whole thought. But many beings together cannot have one whole representation. Accordingly that subject which has a whole representation must be *simple*. The soul is thus either a simple substance or a composite [*compositum*] of substances. If it is the latter, then it cannot think at all. For even if a part thinks, all parts together still cannot have one thought, thus a composite [*compositum*] of substances cannot think at all; accordingly the soul must be a simple substance. (*ML1*, 28:266)

In addition to the fact that everyone's thoughts of a verse from Virgil depends on Virgil's thoughts, this hypothetical implies that everyone's thoughts depend on the thought of a particular person distributing the verses; moreover, even if not dependent on a particular person, then these different thoughts would appear to depend on a group of persons distributing the verses whose collaboration with one another would itself reduce them to one person on the sort of empiricist interpretation towards which Kitcher sees the history of Kant's account of the self pointing. As we see, however, Kant repeatedly asserts the opposite conclusion: these thoughts are *not* unified.

One might also argue that while Kant adheres to a reductionist account of personal identity, and while he is aware that the thoughts of these persons are existentially interdependent, he is simply blind to the particular counter-intuitive implications that such interdependence has within this reductionist account, i.e., that this interdependence of thoughts would imply that what are commonly perceived to be distinct persons are in fact one person. But here the evidence again points in the opposite direction: Kant simply was not interested in the notion of the existential interdependence of these thoughts as a way to explain their unity. To begin, in exploring whether the thoughts about the different verses of Virgil belong to one self, Kant says nothing about an unfortunate lack of interdependence of these different thoughts, much less that a lack of interdependence and thus unity of sorts between these thoughts is what is *responsible* for the fact that there are *different* people who have these thoughts, in the first place. Instead, Kant reasons in the *opposite* direction: he *starts* with the fact that different people—or (rational)

substances—have these thoughts, and then concludes that these thoughts accordingly cannot be united. And he explicitly grounds this conclusion offered in his rational psychology (i.e., this *nervus probandi* of the necessity of the inherence of united thoughts in a common identical subject, or absolute unity) in an appeal to nothing more than our own immediate consciousness of the relationship of our thinking subject to its thoughts: "If I say: I think, then I do not express representations which are divided among many beings, rather I express a representation that takes place in one subject" (*ML1*, 28:266).

Of course, if Kant is thoroughly nonempiricist and nonreductionist about the self and the unity of its disparate accidents, and if, additionally, he argues that our immediate consciousness of ourselves provides us with the paradigm of a noumenal substance, we would expect to see Kant's account of the self reflected in his account of the unity of the disparate accidents of all noumenal substances. And indeed we do. This broader account, which applies to the self in addition to all other substances and thus further illuminates Kant's account of the self, argues that while disparate accidents of a substance lack unity or identity in themselves, they may still be unified if inhering as modifications of an identical underlying substance or what Kant also refers to as an "absolute unity."[69] Kant underscores these original and derivative roles, respectively, of the identity (or absolute unity) of the substantial self and the (derivative) unity of its representations through inherence in this identical self or absolute unity, throughout the late 1760s and 1770s. As early as 1769, seven years before the translation of Beattie's book, Kant offers his view of the unity of accidents in a substance that he has also expressed, and will continue to express, in the Virgil argument: "Thus if the same subject x should successively be a and non [a], thus must the subject not be changed; otherwise x would first be a and then y non a. The *substantiale* is unchanging. For in

---

[69] Kant understands the notion of "identity" in a strict sense, whereby two things are identical only if they are identical in all of their properties. This is consistent with the strict sense of identity presented in Leibniz's notion of the "identity of indiscernables" (E.g., G.W. Leibniz, *New Essays*, 230–31); with Locke's strictest sense of identity (which Locke applies to inanimate masses only) whereby a mass is identical over time only if it neither loses nor gains a single atom (*Essay*, II.xxvii.3); and with Hume's all-purpose account of identity, whereby we have sufficient reason to call an object identical across time only if our impression of it is "uninterrupted and invariable." Employing this concept of identity, Kant is quick to point out instances in which we fail to have identity over time. Thus in a passage from his 1777/78 *Anthropology Pillau* that recalls the *Metaphysik Herder* passage (*MH*, 28:145) considered in section 1 above, Kant remarks that

> The identity of the self is very incomplete. Someone can, if he has done something evil, and afterwards he has improved himself, for which no small time is necessary, no longer for this reason have vengeful punishment applied to him, because he is no longer the same (but still [can] have punishment [applied] as an example). (*APi*, 25:735–36; see also *Religion within the Limits of Reason Alone*, 7:47, 71–75 [1793]; *Anthropology from a Pragmatic Point of View*, 7:294; and *R* 5646, 18:295 [1785–88])

Similarly, Kant rejects the identity of our bodies over time, where the material comprising our bodies changes over time, or (as he puts it in *Anthropology Collins*, in language reminiscent of his discussion of our fluctuating mental states) *flows* over time: "Within ten years is the body composed of different material, just as a current *flows* with different water" (*AC*, 25:13, original emphasis; see also *APa*, 25:246). Kant explicitly contrasts the lack of identity of our person, taken here to mean our body, with the identity of our psychological self, or our "I": "*The real I is something substantial, simple, and permanent*; *by contrast, one views the I as person as changeable*; one says for example, I was big, I was small. The I would not even change, if one were in another body" (*AC*, 25:13; my emphasis).

the succession of the accidents it is always the same" (*R* 4060, 17:401 [1769]). As is consistently the case,[70] in this passage Kant construes the noumenal substance (or *substantiale*) as something distinct from its accidents. But he also makes perfectly clear that whether the accidents *a* and *non a* are unified is not a function of whether or not *a* acted as a necessary and/or sufficient condition for *non a*. Instead, *a* and *non a* are unified only if there is a common noumenal substance for which *a* and *non a* are both accidents. Kant's 1770 *Inaugural Dissertation* confirms this: "Furthermore, the possibility of all changes and successions...presupposes the continued duration of a subject, the opposed states of which follow in succession" (*ID*, 2:410), as does *Reflexion* 3578, probably from the same year as the *Inaugural Dissertation*: "Actually it is not a *modus*, but rather the substance that changes. For that, which is changed, remains; the change is only the succession of its determinations" (*R* 3578, 17:70 [1770–71]). Finally, *Reflexion* 4413, from around 1771, leaves no doubt that Kant does not think that unity between states of a substance can serve as a substitute for inherence in a common substance, but instead that inherence in a common substance serves as the precondition for the unity of the states of the substance: "The connection of many (different) things in a whole presupposes the dependence on a common ground and flows out of it" (*R* 4413, 17:537; see also *R* 4418, 17:539 [1771]).

We have thus seen that Kant's account of the self during the two decades leading to the *Critique* is fundamentally nonreductionist and nonempiricist. Kant argues that we have an immediate consciousness of being a simple substance and that the unity of our disparate accidents depends upon their inherence in an identical, simple substance. While this requirement of our simple substantiality is at the foundation of Kant's account of the unity of our representations, Kant recognizes, unlike Leibniz, that it is not necessarily the case that all substances are rational,[71] and so his nonreductionist, nonempiricist account of the soul and the unity of its accidents is not to the exclusion of additional requirements for unity that Kant recognizes apply solely to the soul as a rational substance.

To begin, a rational substance (or "intelligence") has a "higher power of knowledge" (*AC*, 25:10 [1772/73]), or is a "being that thinks and wills" (*ML1*, 28:226 [1777–80]). Kant views these higher capacities, and so our intelligence, as dependent on our capacity for personality (or self-consciousness): "as intelligence he [a human] is conscious of himself" (*AF*, 25:475 [1775/76]), and, "The I is the

[70] E.g., *R* 4412, 17:536 [1771]; *ML1*, 28:261; *MH*, 28:25, 145; *R* 4762, 17:720 [1775–78]; *R* 5290, 18:144 [1776–78].

[71] While Kant rejects Leibniz's assumption that all substances are rational, he nonetheless expresses some empathy for Leibniz, rebuking those who mock Leibniz while themselves failing to recognize that substances do need an inner state in addition to an outer state and that we simply have no other model of an inner of a substance than our own, making this assumption understandable if not justifiable (*DSS*, 2:328, 328 n.; *R* 4718, 17:685–86 [1773–5]; and *ML1*, 28:207–08). When it comes to animals, Kant again recognizes our lack of certainty regarding the inner state of these substances and whether they are rational, making clear that while the capacity for self-consciousness makes us rational and that, "If a horse could grasp the thought I, I would climb down and need to view it as my equal [*Gesellschaft*]" (*Me*, 25:859; see also *A*, 7:127; *AM*, 25:1215; *MoMr*, 27:1572; *MD*, 28:689–90; and *MVi*, 28:735), he thinks that there is nothing about animal behavior that compels us to think that instinct and lower faculties could not account for all of their actions (*MH*, 28:899–902), whereas with other humans this is not the case and we thus do project our I onto them (e.g., *CPR*, A 347).

foundation of the capacities for understanding and reason and the entire higher power of knowledge, for all of these capacities rest on the fact that I can observe and inspect myself and what transpires in me" (*AC*, 25:10 [1772/73]).[72] Accordingly, if a substance were to lose its power of self-consciousness, or personality, its spiritual existence would end—be it temporarily, as a "spiritual slumber," or permanently, as a "spiritual death"; Kant explains in a discussion of the possibility of life after death:

> It is first asked whether the soul will be conscious of itself in its future state or not. If it were not conscious of itself, then this would be *spiritual death* … But if it is not conscious of itself, although its lifepower is still there, then this is *spiritual slumber*, in which the soul does not know where it is, and cannot rightly adapt itself in the other world" (*ML1*, 28:295–96).

Finally, the unity of the earlier and later mental states within an identical consciousness (or personality) does not require merely that these mental states inhere in the same substance, nor that they inhere in the same substance that is self-conscious at these earlier and later moments, nor that we remember these earlier mental states of our substance at the later time, but also that we *remember having been* the earlier substance that *had* the earlier mental states that we now remember. Kant thus continues: "personality, the main matter with the soul after death, and *the identity of the personality of the soul, consists in this*: that it is conscious to itself that it is a person, and *that it is also conscious of the identity; for otherwise the previous state would not at all be connected with the future one*" (*ML1*, 28:296, my emphasis). Thus, according to Kant, while we can imagine changes in our thoughts, in our bodies, and in our guiding moral maxims over time,[73] it is impossible to imagine this occurring with our personality, or I. Kant accordingly observes in *Anthropology Collins* that "There is no human who would not like to trade with another, their face, their whole body, yes in fact with the properties of the soul; but no one decides to trade their whole I; it is in itself a contradiction: therefore it is actually not at all obscure" (*AC*, 25:15, 248 [1772/73]). Thus while Kant's account is clear in its nonreductionism—requiring that accidents inhere as modifications of the same

---

[72] Three years later, in *Anthropology Friedländer*, Kant declares in like manner that "There is no thought that serves as the ground for others other than the thought of the I" (*AF*, 25:473). In the anthropology lectures from the winter of 1777/78, *Anthropology Pillau*, Kant offers these superlatives: "No thought is greater or more important than that of our I...This expression I, or the capacity to represent oneself to oneself, does not simply constitute the excellence of human nature; rather it constitutes the entire worth of humans" (*APi*, 25:735–36; see also *ML1*, 28:275–78). As note 71 observed, four years after this, and five months after the publication of the first *Critique,* in the anthropology lectures published in 1831 as *Menschenkunde*, Kant's awe for our capacity to say 'I' is undiminished, expressing itself both as pride and humility: "The I contains that which distinguishes humans from all animals. If a horse could grasp the thought I, I would climb down and need to view it as my equal [*Gesellschaft*]" (*Me*, 25:859). In the *Anthropology Mrongovius* of 1784/85, Kant continues: "Among *all beings* on the face of the earth, only the human has a representation of his I or of his person. This is also what makes him into a rational being" (*AMr*, 25:1215). Kant's book, *Anthropology from a Pragmatic Point of View*, prepared by Kant in 1796 or 1797 (and available since 1907 in the *Akademie* edition of Kant's complete works) rounds out this chronology of acclaim, asserting that "The fact that man is aware of an ego-concept raises him infinitely above all other creatures living on earth. Because of this, he is a person" (*A*, 7:127; see also, *MoMr*, 27:1572; *MD*, 28:689–90; and *MVi*, 28:735).

[73] See note 69 above.

substance in order to be united—further requirements can piggyback on this one, as Kant's account of rational substances here makes clear.

5. CONCLUSION

This investigation of Kant's account of the soul thus reveals the third path blazed by Kant's immediatism between rationalism and empiricism, so that Kant's conclusions of the soul's substantiality and simplicity does not place him in the company of the rationalists, nor does his rejection of rationalism land him by default in the company of the empiricists. Indeed, it is worth noting here how Kant effects distance between his account and those of Locke and Hume, first in terms of methodology and the place afforded immediate consciousness, and next by means of his nonreductionist conclusions of the inherence of accidents in noumenal substances. Regarding the former, we see that Locke, in agreement with Kant, holds that consciousness is two-fold, both of our ideas and of ourselves as the substance that exercises powers to ground these ideas (*Essay*, II.xxvii.9). But Locke's account is hidebound by its own empiricist epistemology, which affords no place for the latter, reflexive awareness of our own substantiality, for the simple reason that this awareness of being something with powers is divorced from all empirical content, which is found only in the former type of consciousness, of our representations. Hume famously calls Locke to task on this point, and accordingly eliminates the nonempiricist concept of reflexive self-awareness from his own account of consciousness, thereby reducing the self to what he proudly assays as a mere bundle of individual representations (*Treatise*, I.vi). Only subsequently does Hume recognize, however, in the Appendix to the *Treatise*, that this reduction of the self to a fractured heap of representations not only successfully refutes Locke on internal empiricist grounds, but at the same time renders incoherent the unity that in fact characterizes thought (*Treatise*, Appendix), even though we know thought exists, leaving Hume to lament: "Did our perceptions either inhere in something simple and individual, or did the mind perceive some real connexion among them, there wou'd be no difficulty in the case" (*Treatise*, Appendix, 678).

It is here that Kant's transcendental methodology breaks radically from Locke's and Hume's empiricism, in a sense working backwards from the sort of *reductio ad absurdum* that empiricism here faces on the issue of our capacity for thought and personal identity, demanding a place, at the cost of any conflicting epistemology and its conclusions, for those conclusions of which we are immediately aware and without which thought and personal identity are rendered incoherent. These conclusions for which Kant demands a place are roughly those that Hume ruefully notes would solve the problem but which have no place within an empiricist epistemology, although Kant adds the stipulation that not only need these perceptions "inhere in something simple and individual" (ibid.), but that this simple and individual substance be immediately conscious of itself as the common owner in which these perceptions inhere, an immediate consciousness that is a necessary though not sufficient part of what allows for there to be "a real connexion among [the perceptions]" (ibid.)—the categories of the understanding serving as yet another necessary ingredient for forging this unity, an account which Kant first formulates around 1774–75 in his *Duisburg Nachlass* (*R* 4674–4684, 17:643–74).

Kant thus creates room for Locke's position on the two-fold nature of consciousness where an empiricist epistemology precludes it, carving out a place for non-empirical, immediate consciousness of oneself, and of oneself as something that exercises powers (i.e., as something that is a noumenal substance), as a necessary condition for thought and personal identity. Kant's transcendental epistemology thus recognizes, (a) within its negative component, of transcendental idealism, that our representations of other noumenal substances (assuming that they are not the products of our imagination) are merely the *effects* of these noumenal substances on ourselves as noumenal substances, which effects are *per se* colored by our own necessary activity in registering the activity of these noumenal substances on us.[74] But his transcendental epistemology also recognizes, (b) within its positive component, of transcendental immediatism, that in the one case in which we, as rational substances, do not relate to a substance merely as it affects us, but instead relate to a substance by being it, we are not limited to the mediated effects of a substance on ourselves (although we can also perceive the effects of our activity on ourselves in the contents of inner sense), but instead also need to be immediately aware of being the common owner of our representations.[75] Without the transcendental approach of Kant, which starts from the conditions for thought and personal identity, Locke and Hume thus build epistemologies that generalize from, and only make room for, our relationship to *other* substances and not to ourselves as substance. As a result, Locke's account suffers from inconsistency and Hume's from inadequacy in explaining personal identity and thought. Kant accordingly does not beg the question of personal identity against Hume, or any other empiricist, by asserting an immediate consciousness of ourselves as substances, as Kitcher claims he would if he were to hold this view. Instead he offers this assertion as part of a deeper reform that eliminates the empiricist doctrine that otherwise rules out this immediate consciousness.

Second, Kant's nonreductionist account represents a departure from Locke's precedent because, whereas Locke is open to the possibility that our consciousness and our thoughts might move across different substances, Kant rejects this possibility. Of course, Kant readily concedes that there is a succession of phenomenal substances that makes up our physical bodies over our lifetimes, and, even more directly, recognizes the possibility that there could be a succession of any phenomenal substances that might manifest our noumenal substance in space and time.[76] But he is clear that all of our thoughts, as accidents, inhere as modes of a noumenal substance, understood, flexibly, as that which has powers by means of which it relates to its accidents, and that they accordingly need to inhere in the same substance in order to even possibly be united. It makes no sense to speak of divorcing a mode from its substance as though it were a separate existence; for Kant, thoughts are not like books on a bookshelf that can conceivably be trans-

---

[74] *R* 4704, 17:681 [1773–77]; *ML1*, 28:206–07; *MH*, 28:52; *R* 3581, 17:71 [1769–70]; *R* 3783, 17:297 [1764–68]; and *R* 5290, 18:144 [1776–78].

[75] *AC*, 25:14 [1772/73]; *R* 4718, 17:685–86 [1773–75]; *R* 5297, 18:146 [1776–78]; *R* 3921, 17:345–46 [1769]; *MH*, 28:102; *ML1*, 28:206, 207, 225–26, 265.

[76] *IC*, 2:293; *DSS*, 2:321–22; *R* 4757, 17:704 [1775–77]; *ML1*, 28:272–73 [1777–80]; *R* 6004, 18:421 [1780–89].

ferred to another bookshelf (*R* 3783, 17:292 [1764–68], cf. *MMr*, 29:770). Thus, for the same reason that Kant argues that even God could not place an accident in another substance—be this transfer one of an occasionalist variety (as for Descartes or Malebranche) or a pre-established harmony variety (as for Leibniz or Wolff) (*MH*, 28:52 [1762–64]; *R* 3581, 17:71 [1769–70])—Kant also *ipso facto* rules out the transfer of accidents from any one substance—be it a rational or nonrational substance—to another, by any other means, arguing that substances merely affect other substances' powers, but cannot transfer to them their accidents (*ID*, 3:409; *ML1*, 28:214–15). This allows, at best, for the *resemblance* of accidents across different substances, as in the case where people's thoughts about a Virgil poem resemble one another. And, of course, Kant's view that our memories are limited to the previous mental states of our own substance does not imply that our memories are always correct: memories obviously can be false. But they would not be false because they fail to register our migration across noumenal substances, but instead because they fail to accurately recall earlier mental states of the noumenal substance. Thus, whereas Kant, like Locke, does not think that the identity of an underlying substance is sufficient for personal identity (*ML1*, 28:295), he does think that this substantial identity is necessary, along with our connection (through memory) to our past mental states, which include a reflexive component in the form of an immediate self-consciousness.

In departing from the competing models of self offered by the empiricists, rationalists, and skeptics of his day, Kant's immediatism also offers new perspectives on some of the thorny issues in modern philosophy of mind. Because Kant understands noumenal substance simply as that which has powers by means of which it relates to its accidents, temporally located mental states, on the one hand, and temporally and spatially located material states, on the other, are both expressions of underlying noumenal substances, and as such there is no reason to assume that the mental or material could not at bottom issue from the same sort of substance.[77] This eliminates the intractable problem (within dualist models) of interaction between the mental and material, where the mental and material are viewed as fundamentally different kinds of substance. At the same time, by not treating matter as substance in itself, but rather as the appearance of substance in itself, Kant's account addresses the problem of the possibility of the multiple realizability of thought in matter. Because noumenal substances underlie the appearance of matter, and because it is conceivable that such underlying noumenal substances could be of the rational sort (who have among their powers the power of self-consciousness and thought, as we do), the fact that such thinking substances could be realized or manifested in space and time in any number of ways presents no contradiction, because Kant does not wed noumenal substance to any particular phenomenal manifestation over another. On the question of other minds, Kant simply recognizes that, conscious of others only as phenomena, we must simply project to others what we experience in ourselves on the basis of

[77] Although both would have noumenal substance underlying them, this is not to imply that these noumenal substances would have the same sorts of powers, for example that they would all have the power of representation, much less the higher powers of representation, including reason, judgment, understanding, and some instances of imagination.

whether the exhibited behavior can only be explained by appeal to the assumption of self-conscious, rational beings. It is accordingly always possible that we may be wrong, and this is why Kant admits the possibility that other animals might be rational, in which case we would need to treat them as equals,[78] although their behavior does not yet, in Kant's view, appear to require our assumption that they have immediate self-consciousness and that they can accordingly recognize their representations as representations (*FS*, 2:58–61; *MH*, 28:115–19, 899–903; *ML1*, 28:274–78). Additionally, Kant's account saves the phenomenon, so to speak, by recognizing that mental states and material states are fundamentally different and cannot be reduced to one another, because they are simply phenomena that are located in irreducibly distinct phenomenal frameworks. Finally, as we have seen, Kant's immediatism also aligns itself with our own experience of self-consciousness by resisting a reductionism that eliminates our reflective awareness of our representations as such and of being something that exercises powers to ground these representations.

In sum, this study of Kant's recorded thought during the crucial two decades leading to the *Critique* dispels the belief that Kant's pre-*Critique* history finds Kant advancing on a trajectory toward an empiricist account of the self in the *Critique*, and defies the traditional understanding of the Kant-rationalism conflict whereby Kant either accepts or rejects the conclusions of the soul's simplicity and substantiality on the rationalists' terms. In many detailed accounts, including those in the recently published lecture notes on anthropology, Kant asserts the soul's noumenal substantiality and simplicity, our immediate consciousness of these conclusions, and the necessity of these conclusions. But revealed in the magnified detail of these positive accounts of the soul are the hard and fast contrasts between the *meaning* of these conclusions and those of the rational psychologists. Kant registers the rationalists' conflation of substance and power and their resulting failure to draw distinctions in kind, not merely in degree, between powers, most notably between understanding and sensibility. Taking note of this distinction and the categorical divide it effects between phenomena and noumena, Kant identifies (against the rationalists) a subsequent barrier in meaning between, on the one hand, those pure concepts of a noumenal substance, ground, community, and simplicity (the paradigms of which we acquire by means of our immediate consciousness of being such a thing in general, distinct from and absent any determinate qualities at all), and, on the other hand, these concepts as they apply by analogy to the determinate realm of appearance, as necessary conditions of experience. Whereas the soul's noumenal substantiality and simplicity say nothing about the soul's permanence and incorruptibility, phenomenal substantiality and simplicity tautologically imply permanence and incorruptibility. Unaware of this categorical divide in meaning between noumenal and phenomenal substance, the rationalists infer across it. They accordingly impute to these indeterminate conclusions of our noumenal substantiality and simplicity the determinate attributes of permanence and incorruptibility that rightly apply only in the distinct, phenomenal realm of our representations. Kant identifies and so avoids these mistakes, and therefore

[78] For more on this, see note 71 above.

concludes only what is warranted by our immediate consciousness of ourselves as things that exercise powers by means of which we ground accidents, namely the soul's indeterminate, noumenal substantiality and simplicity, which imply neither permanence nor incorruptibility.[79]

[79] For his extensive comments and help, many thanks to Paul Guyer. Thanks also to Andrew Brook, Gary Hatfield, Rolf-Peter Horstmann, Rae Langton, Alison Laywine, and Werner Stark for comments on earlier drafts of this paper. I am grateful also to Patricia Kitcher and Karl Ameriks for their respective roles as commentator and chair of the session at the 2001 Central Division Meeting of the *APA* at which I presented an earlier version of this paper, to the audience at that session, and to the audience at the 2000 Midwest Study Group of the North American Kant Society meeting, at which I also delivered a version of this paper. Finally, my thanks to an anonymous reviewer at the *Journal of the History of Philosophy* for excellent feedback.