

Kant on personal identity and moral obligation

Paul Tulipana

UCLA

ABSTRACT: In this essay, I offer a new interpretation of Kant's theory of personal identity. According to Kant, I claim, human agents are, in their awareness of moral obligation, also aware of themselves as free and responsible creatures with careers involving moral vocations to be pursued over time, and therefore as cross-temporally identical persons.

§

It is a surprising and remarkable fact that Kant never offers an explicit account of personal identity. The only direct mention of the notion he makes during the critical period is in the discussion of the third paralogism of pure reason, in the *Critique of Pure Reason's* Transcendental Dialectic.¹ In that discussion, however, he is exclusively concerned to refute the earlier rationalist idea that the personhood of the self is constituted by an a priori theoretical awareness of one's cross-temporal

¹ References to the *Critique of Pure Reason* are given by pagination in the first (A) and second (B) editions. In quotations, I follow Paul Guyer and Allen Wood's translation in *Critique of Pure Reason*. Cambridge: Cambridge University Press, 1998. References to other works by Kant are given by volume and page number in the Berlin Akademie edition. In quotations, I follow the translations in *The Cambridge Edition of the Works of Immanuel Kant*, notably Mary J. Gregor's translations of the *Groundwork of the Metaphysics of Morals*, the *Critique of Practical Reason*, and the *Metaphysics of Morals in Practical Philosophy*. Cambridge: Cambridge University Press, 1996, and George di Giovanni's translation of the *Religion within the Boundaries of Mere Reason* in *Religion and Rational Theology*. Cambridge: Cambridge University Press, 1996.

identity as a thinking substance, and no explicit alternative is formulated. Near the end of the 1781 A-edition version of the paralogism, he remarks that the theoretical concept of personhood “can remain” as it is “necessary and sufficient for practical use” (A365-6). But, although he reiterates this suggestion in the rewritten 1787 B-edition of the Paralogisms chapter (B430-31), it is never explicitly developed in the intervening or subsequent practical philosophy, which, although deeply concerned with persons in various ways, makes little mention of the self’s identity through time.

As a consequence of this vexing circumstance, Kant’s interpreters have been left to search for an implicit statement of his account of personal identity in what scattered remarks of his appear to them pertinent to the subject. This search is an unusually difficult one, as Kant’s lack of synoptic discussion leaves it uncertain which part or parts of his account of human corporeal, psychological, and volitional life we should expect him to deploy in elaborating the notion of personhood. Is our personal identity to be established, that is, through the awareness we have of our own empirically accessible bodies? By appeal to what of ourselves is revealed in inner sense? Or perhaps through our awareness of practical requirement in one or more of its forms? The interpretive difficulty is exacerbated by the fact that at various points, Kant suggests that the identity of the thinking self “in everything manifold of which [it is] conscious” is an analytic truth (B408), leaving it unclear what role any form of self-consciousness besides, perhaps, the consciousness of ourselves as the mere subjects of apperception, could have to play in establishing it.

These last suggestions, along with the widely held sense that there is no room in Kant’s theoretical philosophy for a form of self-consciousness by means of which we could recognize ourselves as persons in any substantial way, may help to explain the fact that philosophers looking for an account of personal identity in Kant have in the last handful of decades lavished an enormous amount of interpretive attention on the notion of the subject of apperception. Although a great many examples of this interpretive tendency might be cited here, I briefly mention only three.²

² To get a sense of just how widespread this trend has been in the recent past, consider a 1993 review essay in which Günter Zöllner, canvassing a cohort of seven interpreters who published on

First, Patricia Kitcher argues that where we can impute a particular relation of causal dependence between the contents of a set of diachronically distributed mental states—where such a collection is “synthesized”—a coherent Kantian account will find a cross-temporally identical person. In this way, acceptable Kantian accounts of personhood will be more sophisticated versions of Hume’s “bundle” theory of the self.³ Colin Marshall, similarly, argues that Kant gives the identity conditions for the self in terms of apperceptive subjectivity, although he claims that Kant identifies the self with whatever entity is responsible for synthesis rather than with the synthesized representations themselves.⁴ Arthur Melnick, finally, also looks to apperceptive subjectivity in accounting for the self’s identity, although he identifies the self of apperception neither with a collection of representations nor with the entity responsible for combining representations, but

Kant’s theory of the self between 1986 and 1991, describes all seven as engaged in a “search for Kant’s ‘third self’—different from the empirical self of introspection and the noumenal self of moral personhood”. See his “Main Developments in Recent Scholarship on the Critique of Pure Reason.” *Philosophy and Phenomenological Research*, 53: 2 (1993): 445-466, here p. 460-61.

³ For the initial statement of this view, see Kitcher, Patricia. “Kant on Self-Identity.” *Philosophical Review*, 91:1 (1982): 41-72; here pp. 54-55. The view is subsequently developed in her *Kant’s Transcendental Psychology* (New York: Oxford University Press, 1990), and in *Kant’s Thinker* (New York: Oxford University Press, 2011). According to the original essay, the “relation of causal dependence” is left inadequately explained by Kant (p. 55). In later work, she suggests that the relation is one that is unmediated by outer sense (*Kant’s Transcendental Psychology*, p. 121).

⁴ Marshall, Colin. “Kant’s Metaphysics of the Self.” *Philosopher’s Imprint*, 10:8 (2010), pp. 1-2. For Marshall’s discussion of the relation of his view to Kitcher’s and Melnick’s, see pp. 10-12. Marshall himself thinks that there is an important distinction in Kant between the identity of the self and the identity of the person (cf. p. 9, n. 35). In this essay, I take personality to be a possible property of the self, namely, its awareness of itself as identical through time. When a self instantiates this property, I say, it’s a personal self, or, simply, a person.

instead with an ongoing *capacity* “for being an identical intellect accompanying inner attending”.⁵ Although these interpreters thus disagree about Kant’s metaphysics of the self, they each attempt to explain the self’s identity or personality wholly in terms of apperception.

Although the considerations mentioned above do give interpreters excellent reasons to reflect on Kant’s idea of apperceptive subjectivity in the context of thinking about his account of personal identity, the interpretive tendency to reflect exclusively on the subject of apperception in this context is a surprising one, given Kant’s repeated indications that practical thought is central to the cognition of the personal self. Part of the explanation of this tendency to exclusivity is probably innocuous enough: some interpreters simply do not find Kant’s discussions of the self in the practical philosophy relevant to their particular projects. Thus although Andrew Brook makes a point in *Kant and the Mind* to emphasize that for Kant a substantial account of personal identity must ultimately be practical, he does not discuss this aspect of Kant’s view in any detail, as, in his view, Kant’s discussions of the self in the *Critique of Pure Reason* are more important to understanding his theory of mind.⁶ I myself suspect, however, that the tendency is due in largest part not to anodyne editorial doubts of this sort, but rather to deep and prevalent suspicions about the conception of the self that features in Kant’s moral thought.

Interpreters whose preferred tack is avowedly revisionist tend to be admirably candid on this point. In *Kant’s Transcendental Psychology*, for example, where Kitcher describes her account of mental unity (and thus of personal identity) as “deliberately revisionist”, she also reports that

⁵ Melnick, Arthur. *Kant’s Theory of the Self*. New York: Routledge, 2009, here p. 131.

⁶ Brook, Andrew. *Kant and the Mind*. Cambridge: Cambridge University Press, 1994. The claim about relative importance is from the Preface (p. xi) and the points about the role of practical reason in establishing personal identity are from the chapter on the Third Paralogism (p. 207). Brook shares Kitcher’s view that Kant adopts something like Hume’s account of personal identity in the first *Critique*, adding only the claim that as subjects of “apperceptive self-awareness” we have “a unified consciousness, and it reaches backwards and forwards in time” (p. 193).

she feels compelled to adopt this interpretive posture on the basis of what she sees as Kant's "staggering ambivalence" about whether the 'I' of apperception designates a phenomenon or a noumenon. She takes this supposed ambivalence to show that Kant's account of the apperceptive self is ultimately "incoherent in his own system", and thus that certain elements of that system—namely, the noumenal self, and with it the attendant possibility that the 'I' of apperception might refer to it—should be excised.⁷ Béatrice Longuenesse makes a similar claim about the idea of the noumenal self in "Kant and the Identity of Persons", although she defends it by different means. In treating personal identity as a moral phenomenon, she argues, Kant surreptitiously rehabilitates the very sort of metaphysics that he is so concerned in the first *Critique* to reject by locating its basis in practical reason. Having convincingly undercut the notion that theoretical reason could establish conclusions which would extend human knowledge beyond the bounds of possible experience, Kant ultimately succumbs, in his account of reason's practical domain, to the same

⁷ See p. 121 for the claim that her account is deliberately revisionist. The subsequent quotations are from pp. 139-40. The passage that suggests to her that Kant's account of the apperceptive self is incoherent is at B429: "In this way [as the subject of pure apperception] I represent myself to myself neither as I am nor as I appear to myself, but rather I think myself only as I do every object in general from whose kind of intuition I abstract". Kitcher's discussion of this passage seems to suppose that the 'I' of apperception refers to a determinate object, and thus that what Kant says in it amounts to the claim that the relevant object is neither a phenomenon nor a noumenon. As she recognizes however, the distinction Kant explicitly makes here is not between the kinds of object that the apperceptive self might be (if this way of talking makes sense), but rather between ways that one might grasp oneself in thought (p. 257, n. 54). One way to respond to the felt threat of incoherence might thus be to abandon the idea that the 'I' of apperception refers to a determinate object. Doing so might also seem to make natural sense of Kant's repeated claims that this 'I' is "wholly empty" (A346/B404) that it "belongs to thinking in general" (B423n), etc. See pages 15-18 below for further textual defense of this idea.

natural but misleading temptations that he argues in the Transcendental Dialectic had seduced his less sophisticated rationalist forbears. Thus, like Kitcher, Longuenesse is moved to offer what she takes to be an appropriate Kantian alternative to Kant's own supposedly untenable account of personal identity.⁸

Whatever the reasons for it ultimately are, the lack of interpretive attention to Kant's moral metaphysics of the self is, I think, unfortunate, both as an interpretive and as a philosophical matter. First, as I argue in some detail below, any interpretation of Kant's theory of personal identity that does not make reference to his moral thought is at worst mistaken and at best seriously incomplete. Kant's positive metaphysics of the personal self has first and foremost to do with the subject of moral obligation, and not with the subject of transcendental apperception, with the empirically accessible subject of inner sense, or with the human body. Second, on a more careful look, Kant's moral account of personhood is not the metaphysical or logical disaster that some have supposed it to be. In the first place, the kinds of self-consciousness that Kant thinks are germane to establishing personal identity are far from generating incoherence in his system; rather, they fit harmoniously into a set of the forms of self-awareness that he reasonably believes that finite rational beings enjoy. Further, Kantian moral self-consciousness is not itself a previously unimagined practical version of the immediate awareness of oneself as a substantial denizen of the intelligible world that some earlier rationalist philosophers supposed characteristic of first-person thought *simpliciter*. It is rather the awareness of oneself as a finite being subject to moral obligation, and accordingly as the subject of a temporally extended moral vocation. Thus unless we are ready to give up on metaphysical principle the idea that we sometimes genuinely stand

⁸ Longuenesse, Béatrice. "Kant on the Identity of Persons." *Proceedings of the Aristotelian Society*, New Series, 107 (2007): 149-67, here esp. pp. 160-63. Longuenesse's account of personal identity invokes more cognitive resources than just those characteristic of apperceptive subjectivity. She does hold that the unity of consciousness is criterial to personal identity, but she thinks that bodily continuity is as well. More about her view in the essay's third part below.

under moral obligations and know it, we have no obvious reason, logical or metaphysical, to reject Kant's moral account of persons and their identity out of hand.

1. Criteria for Personhood

Kant uses the term “person” differently at various points in order to emphasize various aspects of the being to which it applies. In the *Critique of Pure Reason*, the term is strictly used to discuss identity; there, personhood is associated with the consciousness of the numerical identity of one's thinking self in different times (A361). In some contexts, Kant indicates that this idea is at issue by using the more specific term “psychological personality” (*psychologische Persönlichkeit*). He does this most often when he is contrasting it with what he calls “moral personality” (*moralische Persönlichkeit*).

Throughout the practical philosophy, he associates the latter with the connected notions of the autonomy of the will and transcendental freedom. In the second section of the 1785 *Groundwork of the Metaphysics of Morals*, he associates it with autonomy: persons are “lawgiving beings”, beings whose nature is such that their maxims are formally fit for giving universal law (4:438). Since autonomy and transcendental freedom are reciprocally implicating concepts (4:446-7, 5:28-9), we should expect that persons can also be characterized in terms of the latter. And indeed, Kant offers this characterization in the 1788 *Critique of Practical Reason*, where he associates personhood with “freedom and independence from the mechanism of the whole of nature, regarded nevertheless as also a capacity of a being subject to special laws — namely pure practical laws given by his own reason” (5:87).⁹

⁹ See also the following *Reflexionen*: R655 15:289, R6305 19:307, R6713 19:139, and R6861 19:183.

In the 1797 *Metaphysics of Morals*, finally, Kant connects the ideas of freedom and autonomy with the idea of being an appropriate subject of action imputation:

A *person* is a subject whose actions can be *imputed* to him. *Moral* personality is therefore nothing other than the freedom of a rational being under moral laws (whereas psychological personality is merely the ability to be conscious of one's identity in different conditions of one's existence). (6:223)¹⁰

The fact that Kant regularly distinguishes psychological from moral personality can tempt one to think that he understands them as essentially distinct phenomena that, perhaps because of the deep connection between identity and imputability in the work of earlier philosophers, share a name. With this thought in mind, it becomes natural to also think that Kant simply doesn't have a univocal notion of personhood with a single correlated set of identity conditions. One is rather tempted to think that he understands personhood in two essentially different ways, one theoretical

¹⁰ Although this passage is the first time Kant explicitly specifies personhood in terms of imputability in his published work, he had been thinking about it in these terms since at least the mid-1770s, and he continues to do so throughout the 1790s. In his lectures on metaphysics from the earlier decade (*Metaphysik L1*), he characterizes moral persons as those beings to which "free actions are ascribed" (28:296), and accordingly as those beings to whom "everything can be imputed" (28:277). Similarly, in a note dated somewhere between 1776 and 1778, he writes, that "[m]oral personality is the capacity for motivating grounds of mere reason, by means of which a being is capable of laws and thus also of imputation" (R5049, 18:73). In the critical elucidation of the Analytic in the second *Critique*, he asserts that transcendental freedom forms the basis of all moral laws and that only in virtue of such moral laws is action imputation appropriate (5:96-97). And in lectures given between the second *Critique* and the *Metaphysics of Morals*, collected under the heading *Metaphysik Dohna* (1792-1793) he defines persons as those beings "capable of an ascription (imputation of action)" (28:683).

and one practical, with each way corresponding to its own distinct set of identity conditions—the natural candidates being conditions requiring consciousness of one’s own cross-temporal identity on the theoretical side, and conditions requiring freedom on the practical side.¹¹ It is important to resist this temptation, however. Kant is explicit that the very same notion of personhood at stake in the paralogisms, clearly the notion that he calls psychological personality, is also at stake in the practical philosophy, clearly moral personality. As I have already mentioned, near the end of his discussion of the A-edition third paralogism, he writes:

Meanwhile, the concept of personality, just like the concepts of substance and the simple, can remain...and to this extent the concept is also necessary and sufficient for practical use. (A365-6)¹²

¹¹ Longuenesse, for example, thinks this (“Kant on the Identity of Persons”, p. 160) as, apparently, do Melnick (who describes *Kant’s Theory of the Self* as “a book about the ontological status of the thinking self” that does “not attempt to incorporate an account of the practical-moral self”, p. viii) and Zöller (“Main Developments”, p. 461). It may be another reason why relatively so much attention has been paid to Kant’s theory of apperception and relatively so little toward the moral philosophy in accounting for his theory of personal identity (see pages 2-6 above).

¹² My ellipsis replaces the following text: “(insofar as it is merely transcendental, i.e., the unity of the subject which is otherwise unknown to us, but in whose determinations there is a thoroughgoing connection of apperception)”. There has been much scholarly controversy about the sense in which the merely transcendental concept of personality is supposed to be *sufficient* for practical use. See, for example, Powell, C. Thomas. *Kant’s Theory of Self-Consciousness*. Oxford: Oxford University Press, 1990, pp. 165-73; Longuenesse, “Kant on the Identity of Persons”, p. 163; Kitcher, *Kant’s Thinker*, pp. 186-87; and Dyck, Corey W. *Kant and Rational Psychology*. Oxford: Oxford University Press, 2014, pp. 168-171. My own way of thinking about this question depends on the idea that in the *practical* cognition of myself under the concepts of substance,

Only in the fourth section of this essay will I be able to fully explain the connection between the two notions of personality and the univocal character of the concept of a person. For now it suffices to note that Kant thinks that the concept of person as it appears in the paralogisms is also the concept of person as it appears in the practical philosophy. Accordingly, there is presumptive reason to treat psychological and moral personality as facets or aspects of some more general concept of a person. On my preferred way of thinking about this concept, the various associations that Kant makes—between personhood and consciousness of cross-temporal self-identity, personhood and freedom, and personhood and imputability—provide *criteria* that are individually necessary and jointly sufficient for personhood.

Given this understanding of these matters, and insofar as this essay's central concern is with personal identity, the freedom and imputation criteria for personhood should be understood in this context to provide constraints on the way that the identity criterion can be satisfied. A satisfying account of the consciousness of numerical identity through time that constitutes personal identity will also show that consciousness to be the self-awareness of a free being as such, a being whose actions can thereby be appropriately imputed to it. Explaining Kant's view of the relevant sort of consciousness, however, requires first answering some prior questions about how he understands the nature of the identity in personal identity.

simple, and identical, I equip those concepts with the special kind of body and meaning that they need to be fully significant—or objectively real—without appeal to the schematizing connections to sensible intuition that would be required to establish their objective reality in their theoretical use. In this sense, the principled application of merely “logical” concepts in practical thought is sufficient to demonstrate their practical applicability: the latter does not require a proof that they appropriately comprehend material given in intuition. For a similar suggestion about the categories of freedom in the second *Critique*, see 5:65-6.

2. The Identity Criterion

In modern discussions about personal identity preceding Kant, the phenomenon tends to be understood in one of two ways. On one prominent picture, most commonly associated with Locke, being conscious of one's identity through time involves being empirically aware of one's own psychological continuity. Because on theories of this type an awareness of psychological continuity is constitutive of personhood, we can call them *psychical* theories of personal identity.¹³ On psychical theories, personal identity is *psychical identity*. On another prominent picture, this one most commonly associated with Descartes and Leibniz, being conscious of one's identity through time involves being aware of oneself as a temporally persisting thinking substance. We can call theories of identity of this type *substance* theories, and call their correlate notion of personal identity *substantial identity*.

According to Locke's *Essay concerning Human Understanding*, a person is "a thinking intelligent Being, that has reason and reflection, and can consider it self as it self, the same thinking thing in different times and places" (II.xxvii.9).¹⁴ His theory of personal identity there is a

¹³ I adopt the unusual term "psychical" rather than the more natural "psychological" for reasons of clarity. "Psychical", as I use the term, has a stronger sense than does "psychological", a term which I hereafter reserve for the broader meaning that correlates with Kant's „psychologische“, as in "psychological personhood". For Kant, psychological personhood is merely the ability to be conscious of one's numerical identity in different times. This use of the term is agnostic with respect to the question of whether the identity of which psychological persons need be aware is itself *of the psyche*, i.e., involves the cross-temporal continuity of the person's inner life. In fact, as I note below, the identity of psychological persons is for Kant not "psychical" in this way, but rather "substantial".

¹⁴ References to Locke's *Essay* are given inline by book, chapter, and section number. Quotations are taken from the version edited by Peter H. Nidditch. Oxford: Oxford University Press, 1975.

psychical one, since for him the self-consideration characteristic of persons is a first-personal, backward-looking empirical awareness of a certain sort of inner continuity, along the lines of memory:¹⁵

For as far as any intelligent Being can repeat the *Idea* of any past Action with the same consciousness it had of it at first, and with the same consciousness it has of any present Action; so far it is the same *personal Self*. (II.xxvii.10)

A person x , then, existing at time t_1 , is identical to a person y , existing at a later time t_2 , if y can at t_2 first-personally recall elements of x 's consciousness at t_1 . Locke also explicitly denies in the *Essay* that personal identity is or involves substantial identity: "it being the same consciousness that makes a Man be himself to himself, *personal Identity* depends on that only, whether it be annexed only to one individual Substance, or can be continued in a succession of several Substances" (*Ibid.*).

In the *New Essays on Human Understanding*, Leibniz rejects Locke's psychical account.¹⁶ He does not deny there that psychical identity is an important marker of personal identity, but thinks of it merely as such—as a reliable indicator of an underlying substantial identity:

I also hold this opinion that consciousness or the sense of *I* proves moral or personal identity...You seem to hold, sir, that this apparent identity could be preserved in the absence of any real identity...I should have thought that, according to the order of things, an identity which is apparent to the person concerned – one who senses

¹⁵ There may be reason to think that the right attitude isn't memory, exactly. See Behan, David. "Locke on Persons and Personal Identity." *Canadian Journal of Philosophy* 9:1 (1979): 53-75.

¹⁶ References to Leibniz's *New Essays* are given using the volume and page numbers in the 1962 *Akademie-Verlag* edition edited by André Robinet and Heinrich Schepers. My quotations of Leibniz follow the translation in *New Essays on Human Understanding*, trans. and eds. Peter Remnant and Jonathan Bennett, Cambridge: Cambridge University Press, 1996.

himself to be the same – presupposes a real identity obtaining through each immediate [temporal] transition accompanied by reflection, or by the sense of *I*; because an intimate and immediate perception cannot be mistaken in the natural course of things. (6:236)

For Leibniz, psychological identity thus presupposes substantial or “real” identity. In defense of this idea, he notes the supposed manifest absurdity of the result, predicted by psychological theories, that an individual struck by amnesia at a certain time would be a different person before and after that time (6:114). If this result is indeed manifestly absurd, then the appearance of psychological identity without an underlying substantial basis might be seen, as Leibniz suggests, to “disrupt the order of things for no reason”, insofar as it “would divorce what can come before our awareness from the truth” (6:242). In this way, personal identity is substantial identity, which, Leibniz thinks, must be demonstrated a priori if it is to be demonstrated at all. Since in his own system, persons are essentially immaterial substances—monads—their identity through time can be established a priori by appeal to their intrinsic natures or complete concepts as substances. Substantial and therefore personal identity is thus settled independently of all empirical self-awareness by God’s choice to actualize this possible world rather than another. Leibniz suggests these ideas in a letter to Arnauld dated July 14, 1686:

Now, since people suppose that it is the same individual substance which perdures...it must needs be that there should be some reason why we can veritably say that I perdure...To be sure, my inner experience convinces me a posteriori of this identity but there must be also some reason a priori. It is not possible to find any other reason, excepting that my attributes of the preceding time and state are predicates of the same subject.¹⁷

¹⁷ Quoted in Scheffler, Samuel. "Leibniz on Personal Identity and Moral Personality." *Studia Leibnitiana* 8:2 (1976): 219-40. See esp. pp. 220-226 for further helpful discussion of this point.

Kant follows Leibniz in thinking that personal identity (“psychological personhood”) is a kind of substantial identity. This allegiance is made explicit in the B-edition Third Paralogism, where Kant identifies psychological personhood with “the consciousness of the identity of [one’s] own *substance* as a thinking being in all changes of state” (B408; my emphasis).¹⁸ It is also present in the longer A-edition version of the text, where Kant considers, in the abstract, a Locke-style body-switching case:

An elastic ball that strikes another one in a straight line communicates to the latter its whole motion, hence its whole state (if one looks only at their positions in space). Now assuming substances, on the analogy with such bodies, in which representations, together with consciousness of them, flow from one to another, a whole series of these substances may be thought, of which the first would communicate its state, together with its consciousness, to the second, which would communicate its own state, together with that of the previous substance, to a third substance, and this in turn would share the states of all previous ones, together with their consciousness and its own. The last substance would thus be conscious of all the states of all the previously altered substances as its own states, because these states would have been carried over to it, together with the consciousness of them; and in spite of this it would not have been the very same person in all these states. (A363-4)¹⁹

¹⁸ This passage, and the connected ones quoted below, provide strong reason to doubt that any attempt to attribute to Kant a theory of real personal identity that is not also a theory of the cross-temporal identity of a substance can ultimately be accepted. See Melnick, *Kant’s Theory of the Self* for such an attempt. See below for the distinction between real and logical identity.

¹⁹ Kant’s thought experiment here recalls Locke’s better-known one in which a prince’s consciousness is transferred to a cobbler’s body (*Essay* II.xxvii.15), although of course Kant and

Unlike Leibniz, however, Kant regards it as impossible to theoretically establish that selves are substances in the sense of temporally persisting existents, and therefore that they are persons. This is particularly clear in the discussion of the First Paralogism in B:

However, that the I that I think can always be considered as subject, and as something that does not depend on thinking merely as a predicate, must be valid – this is an *apodictic* and even an *identical proposition*; but it does not signify that I as *object* am for myself a *self-subsisting being* or *substance*. The latter goes very far, and hence demands data that are not encountered at all in thinking, and thus (insofar as I consider merely what thinks as such) perhaps demands more than I will ever encounter anywhere (in it). (B407)²⁰

Confusion about Kant's views on this matter can arise here if one does not adequately distinguish the idea, which Kant clearly rejects, that it can be shown theoretically that selves are substances in the sense of temporally persisting things, from the idea, which he clearly accepts, that it can be shown theoretically that they are substances in some other sense. In particular, Kant thinks that a certain sort of first-personal theoretical self-awareness establishes that one is a substance in a very thin sense, which he calls "logical" or "subjective": it establishes that the self is the subject of each of its possible judgments, "and hence cannot be used as the predicate of another thing" (A348).²¹ (Note that it is no determination of a thing to say that I think it.) He rejects

Locke draw opposing conclusions from it.

²⁰ See also the A text: "For the I is, to be sure, in all thoughts; but not the least intuition is bound up with this representation, which would distinguish it from other objects of intuition. Therefore one can, to be sure, perceive that this representation continually recurs with every thought, but not that it is a standing and abiding intuition, in which thoughts (as variable) would change" (A350).

²¹ Note the connection between Kant's characterization of logical substantiality and Leibniz's idea, quoted above, that personal identity is to be established by means of showing that "my attributes

the idea that this or any other sort of theoretical consciousness can establish that one is a substance in the thicker sense discussed above.

From the distinction between these two notions of substance, and having seen that Kant takes personal identity to be substantial in character, we can infer that personal identity is for Kant realizable in two ways, corresponding to the weaker (logical) and stronger (real) senses of the substantiality of the self.

The logical substantial identity of the self is established a priori in what Kant calls pure apperception. Because the representation “I think” is universally and necessarily combinable with all of my other representations, Kant argues, it is to be understood as the product of this pure self-conscious activity (B131-2). The “I think” itself is an empty or contentless representation: no manifold in intuition is comprehended by it. Nevertheless, it is conceptually articulated, borrowing its internal structure directly from the concept of an object in general internal to the understanding, characterized in §§17-19 of the Transcendental Deduction. Accordingly, my representing myself through it when attaching it to my thoughts involves attributing to myself certain subjective conceptual determinations.²² We have just seen one of these: logical substantiality, or the property of the subject of thought correlated analytically with the necessary ability of the ‘I’ to attach to any

of the preceding time and state are *predicates of the same subject*”. For Kant, Leibniz’s mistake here is thus not that he fails to give an appropriate a priori conceptualization of the self, it is rather a fallacy of transcendental subreption, or of mistaking a logical feature of the self for a real one (cf. A402).

²² See again B429: “In this way [as the subject of pure apperception] I represent myself to myself neither as I am nor as I appear to myself, but rather I think myself only as I do every object in general from whose kind of intuition I abstract. If I here represent myself as **subject** of a thought or even as **ground** of thinking, then these ways of representing do not signify the categories of substance or cause, for these categories are those functions of thinking (of judging) applied to our sensible intuition, which would obviously be demanded if I wanted to **cognize** myself.”

of my representations as subject to predicate. A second such determination is logical identity. Just insofar as the ‘I’ of pure apperception must be able to attach to any of my complex representations, I must be able to represent any of the elements that constitute any one of those representations (any element in that representation’s manifold) as related to a single subject. Thus I must represent myself, as a mere thinking thing, as single or numerically identical, for “the manifold representations that are given in a certain intuition would not all together be my representations if they did not all together belong to *a* self-consciousness” (B132; my emphasis).²³ In this way, the *logical* substantial identity of the self is guaranteed by pure rational self-awareness as such.

We have already seen that Kant understands *real* substantial identity on Leibniz’s model, as the numerical sameness of a really existing object temporally persisting through changes of state. The main task of the third paralogism, in both the A and the B editions, is to argue that we cannot infer our real substantial identity from the awareness of our logical substantial identity involved in pure apperception. The more perspicuous B-edition version of this argument begins from Kant’s distinctive theory of human theoretical cognition, according to which the cognition of an object requires contribution from both the intellectual and the sensible faculties: concepts and intuitions. For human discursive intellects, a determinate-object-referring theoretical self-representation would require the combination of material given in intuition, and pure apperception does not provide anything of the sort:

The proposition of the identity of myself in everything manifold of which I am conscious is equally one lying in the concepts themselves, and hence an analytic proposition; but this identity of the subject, of which I can become conscious in

²³ For this reason, the principle of the transcendental unity of apperception makes explicit reference to the logical identity of the subject: “all my representations in any given intuition must stand under the condition under which alone I can ascribe them to the *identical* self as my representations, and thus can grasp them together, as synthetically combined in an apperception, through the general expression ‘I think’” (B138; my emphasis).

every representation, does not concern the intuition of it, through which it is given as an object, and thus cannot signify the identity of the person, by which would be understood the consciousness of the identity of its own substance as a thinking being in all changes of state; in order to prove that what would be demanded is not a mere analysis of the proposition “I think”, but rather various synthetic judgments grounded on the given intuition. (B408)

The claim that I am identical “in everything manifold of which I am conscious” provides a kind of recapitulation of the principle of the transcendental unity of apperception.²⁴ Kant emphasizes here, as he does in the Deduction, that the principle is analytic (cf. B138). The necessary combinability of the “I think” with any given representation does not depend on the idea that the self is spatially or temporally (i.e., objectively) connected to the manifold in its thoughts. It is rather, on Kant’s analysis of the faculty of theoretical cognition, a conceptual truth: the very idea of the subject of a complex representation presupposes that it, the subject of each of the manifold parts of that representation, is numerically identical. The notion of identity involved in pure apperception is thus “logical” in the sense that it depends merely on the analysis of the concept of a subject of representations. Kant’s point in the passage is that this analysis of the concept of the self cannot give us theoretical knowledge that there is a temporally persisting existent to which the concept refers. For that we would need given intuition.²⁵

²⁴ See note 23 above.

²⁵ On the account provided here, then, the self-awareness of the apperceptive subject does establish a form of personal identity, albeit a metaphysically uninteresting one. This part of Kant’s theory of the self, alongside the discussion in the Paralogisms emphasizing that the self’s logical identity cannot be used to establish its real identity, jointly comprise the “negative” phase of his discussion of personhood, whose aim is to clear the ground for his positive metaphysics of persons in the moral philosophy. For a similar take on Kant’s overall strategy with respect to these matters, see Brook, *Kant and the Mind*, p. 206.

3. Strawson and Longuenesse on Personal Identity

What remains to be settled is how Kant thinks that the real substantial identity of persons is established. One thought that has occurred to some of Kant's interpreters, in particular to those who hope to find the resources for a philosophically attractive account of substantial personhood in his theoretical philosophy, is that Kant should appeal to his ordinary criteria for the identification of empirically determinable bodies—in this case, human bodies—in establishing personal identity.

The tradition of interpretation that I have in mind begins with Peter Strawson's discussion of the paralogisms in the third part of *The Bounds of Sense*.²⁶ Strawson claims there that Kant's picture of cognition commits him to the idea that the criteria for the identity of persons must essentially involve ordinary criteria for the identity of outer objects, but that, in his discussions of personhood, Kant simply fails to clearly recognize and explicate this commitment:

It is quite clearly implicit in Kant's position that any use of the concept of a numerically identical subject of experiences persisting through time requires empirically applicable criteria of identity, and that none such are supplied merely by the kind of connectedness of inner experiences provided for by the necessary unity of apperception....On the first point we must remark now that it is one of the weaknesses of Kant's exposition that he barely alludes to the fact that our ordinary concept of *personal* identity does carry with it empirically applicable criteria for the numerical identity through time of a subject of experiences (a man or human

²⁶ Strawson, P.F. *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*. London: Routledge, 2007. In what follows, references to this book are given inline by page number.

being) and that these criteria, though not the same as those for bodily identity, involve an essential reference to the human body. (pp. 163-4)

Later in the same discussion, Strawson notes that Kant does at one point discuss the concept of a thinking thing as it is established empirically in outer sense. The relevant passage is in Kant's refutation of Mendelssohn's proof of the persistence of the soul in the B-edition Paralogisms:

Thus the persistence of the soul, merely as an object of inner sense, remains unproved and even unprovable, although its persistence in life, where the thinking being (as a human being) is at the same time an object of outer sense, is clear of itself; but this is not at all sufficient for the rational psychologist, who undertakes to prove from mere concepts the absolute persistence of the soul even beyond life. (B415)

Kant is thus clearly aware that there are identity conditions for the thinking subject as an object of outer experience. Noting this, Strawson argues that although Kant was confused in his understanding of personal identity by the misguided metaphysics of transcendental idealism, he had nevertheless marshalled the argumentative resources to adopt an attractive view of the matter. According to that view, the conditions for personal identity are largely constituted by the conditions for persistence in life. The 'I' in the use of 'I think' that Strawson calls "criterionless" and that Kant calls "problematic", which the rational psychologist mistakenly takes to refer to the subject of thought as a real (simple, temporally persisting, immaterial) substance, does not, as Kant thinks, refer to a "transcendental subject" instead, but rather results from a kind of abstraction from the ordinary third-personal use of 'I' in referring to an empirically determined body (p. 166).

Whatever the philosophical merits of this view, it is clearly not Kant's.²⁷ Kant himself thinks that no empirical self-representation could deliver evidence of the self's real substantiality,

²⁷ In "Kant on the Identity of Persons", Longuenesse, commenting on Strawson, helpfully emphasizes that the "I" in the problematic use of "I think" is meant by Kant to provide a condition

simplicity, or identity. He also notes that, even if it could, the evidence it provided would merely support a claim about the contingent real identity (or simplicity, etc.) of a particular thinking thing. It would not reveal something about the nature of thinking things in general:

If more than the *cogito* were the ground of our pure rational cognition of thinking beings in general...[that empirical ground] could never serve to reveal such properties as do not belong to possible experience at all (as properties of the simple), nor could it serve to teach *apodictically* about thinking beings in general something touching on their nature... (B405-6)²⁸

A more nuanced and textually sensitive development of Strawson's idea is on offer in Béatrice Longuenesse's 2007 essay, "Kant on the Identity of Persons". I begin my exposition of her view there by noting again that it diverges from my own in taking the notions of psychological and moral personality to indicate not aspects of a univocal concept of personhood, but rather distinct notions.²⁹ On Longuenesse's reading, Kant thinks that something is a psychological person just in case it has both the consciousness of logical identity characteristic of the subject of pure apperception and the consciousness of substantial identity characteristic of empirical knowledge

for, rather than an abstraction from, the empirical concept of oneself—it is "a concept referring a priori to whoever thinks 'I think' whatever the empirical or ontological criteria of that thinker might be" (p. 165). In what follows, references to this essay are given inline by page number.

²⁸ This line of thought also explains why Kant rejects the empirical aspect of the Lockean psychical conception of personal identity. In fact, Kant is directly concerned in the passage to reject the Lockean idea that we might use representations of *inner sense* to advance the pure rational cognition of the self (cf. B413). It is however obvious that his point holds, *mutatis mutandis*, for representations of outer sense: the limitations of inner sense self-awareness he discusses in the passage stem from its empirical character, not from its being, as it were, directed inward.

²⁹ See note 11 above.

of the persistence of one's body through time (p. 159). Moral personality, on the other hand, is constituted by a consciousness "of a different nature": consciousness of one's own metaphysical freedom (p. 161).

According to Longuenesse, Kant obliquely suggests the view of psychological personality just described in the body of the A-edition Third Paralogism:

But if I consider myself from the standpoint of another (as an object of his outer intuition), then it is this external observer who originally considers *me* as *in time*; for in apperception *time* is properly represented only *in me*. Thus from the I that accompanies – and indeed with complete identity – all representations at every time in *my* consciousness, although he admits this I, he will still not infer the objective persistence of my Self. For just as the time in which the observer posits me is not the time that is encountered in my sensibility but that which is encountered in his own, so the identity that is necessarily combined with my consciousness is not therefore combined with his consciousness, i.e., with the outer intuition of my subject. (A362)

Kant explicitly argues in this passage, as he does in the B-edition passage already quoted,³⁰ that knowledge of myself as a temporally persisting object neither arises in apperception directly nor can it be inferred from the self-representation characteristic of that form of consciousness, and thus that appeal to apperception cannot by itself establish personhood. However, Longuenesse suggests that this passage, along with the rest of the discussion of the Third Paralogism in A, might also be read as evidence that for Kant, when a subject's third-personal empirical representation of herself as a body is conjoined with that subject's first-personal representation of herself as the numerically identical logical subject of thought, the criteria for psychological personality are satisfied.

I take it that the reasons already adduced to reject Strawson's interpretation of Kant's view of personal identity also suffice to reject this account of Kantian psychological personality: for

³⁰ See pages 17-18 above.

Kant, no empirical form of self-awareness could prove the self's substantiality, simplicity, or identity (see again B405-6).³¹ Unlike Strawson, however, Longuenesse proceeds to offer an (avowedly revisionist) interpretation of Kant's *moral* theory of personhood as well, here again invoking empirical criteria. As I mentioned near the beginning of the essay,³² part of what motivates Longuenesse to suggest this revision on Kant's behalf is the belief that the concept of a person that Kant deploys in his moral thought—that of a “self-conscious thinking substance, numerically identical through time and conscious that all its determinations originate within itself” (p. 162)—is the result of a deep error in his thought. Although it is not clear how exactly Longuenesse thinks that the awareness of ourselves as moral persons was thought by Kant to be realized in moral life, it is clear enough that she thinks that the basic critical picture should disallow the possibility of its realization altogether. In suggesting that such a form of awareness exists, she indicates, Kant puts himself at odds with his own best insights. The first *Critique* is to have demolished the metaphysical artifice of rational psychology only for the second to build it again on supposedly firmer, practical ground. What Kant should have done instead was present a moral theory of personhood that paralleled his supposed psychological theory—one that involved, that

³¹ A bit of reflection on the concept of substance can help make this idea clearer. Kant follows the Leibnizian tradition in holding that substances neither arise nor perish (cf. B224-5, A348-50). Insofar as he takes personal identity to involve the cross-temporal identity of such a substance, there is a constraint on his view of persons: namely, he must somehow be able to legitimate the judgment that they endure infinitely. This is of course something that appeal to experience could never accomplish. Thus although Kant thinks that it is true and knowable a priori that the *matter* constituting our bodies is substantial (cf., e.g., *Metaphysical Foundations* 4:541-3), there is (happily) nothing in his oeuvre that suggests that our bodies themselves neither arise nor perish.

³² See pages 5-6 above.

is, both the empirical consciousness of one's own body and a thin sort of practical consciousness as a being committed to "establishing a selective hierarchy among the maxims of my actions".³³

Not only do I doubt that Kant could accept an empirically-based moral theory of personal identity like this one, however, I believe that the call for revision that motivates it is premature. I reject the notion that Kant's idea of the consciousness of oneself as a temporally persisting thing in practical thought, and thus his moral account of personal identity, undercuts the best version of his critical picture. Indeed, I think that his way of establishing real identity in moral consciousness is not only comfortably at home in the critical philosophy, it is quite philosophically striking. I now proceed to explain.

4. Kant on Obligation and Identity

In the general remark concerning the transition from rational psychology to cosmology in the B-edition Paralogisms, Kant begins by emphasizing that the empirical consciousness of the self in inner sense cannot provide us with an awareness of ourselves as we are in ourselves. He then suggests the possibility of another sort of consciousness that might serve to do so:

But suppose there subsequently turned up – not in experience but in certain (not merely logical rules but) laws holding firm a priori and concerning our existence – the occasion for presupposing ourselves to be *legislative* fully a priori in regard to our own *existence*, and as self-determining in this existence; then this would disclose a spontaneity through which our actuality is determinable without the need of conditions of empirical intuition; and here we would become aware that in the consciousness of our existence something is contained a priori that can serve to determine our existence, which is thoroughly determinable only sensibly, in regard

³³ See section IV of her essay, pp. 161-63. The quotation is from p. 163.

to a certain inner faculty in relation to an intelligible world (obviously one only thought of). (B430-31)

This passage makes clear allusion to what Kant calls in the *Critique of Practical Reason* the “Fact of Reason”. In recognizing the Fact of Reason, one is aware of oneself as morally obligated, or as subject to a self-legislated moral law. This awareness is what vindicates the notion that one has the capacity to freely determine one’s power of choice, and thus that one is self-determining in a strong sense.³⁴ The passage goes on to suggest that, in recognizing our freedom through the awareness of obligation, we also become aware of one or more objective determinations of our noumenal selves.³⁵ This last claim needs unpacking, and to unpack it, we need to attend to certain elements of Kant’s analysis of the awareness of moral obligation.

For Kant, that awareness involves both a consciousness of the moral law as a criterion of right action and a consciousness of the moral law as a possible motive of action.³⁶ The latter aspect

³⁴ There is much debate about what the Fact of Reason is, about its role in Kant’s argument in the second *Critique*, and about its significance for his moral philosophy generally. The claims about the Fact that I rely on here, however, are meant to be quite uncontroversial: namely, whatever it or (if it is not itself a certain form of awareness) the awareness of it amounts to, in that awareness, we recognize that we stand under moral obligations, and therefore that we are free (cf. 5:29-33).

³⁵ Kant regularly insists that any such objective determination—including personality—would be an essentially practical matter. It bears mention in this connection that, in treating (real) personal identity as a practical phenomenon, Kant finds himself squarely within the Lockean tradition. Locke’s primary concern in locating an interesting psychological continuant to identify with the person in the *Essay* is a practical one. He even claims there that the term “person” is itself a “Forensick” one, used for “appropriating Actions and their Merit” (II.xxvii.26).

³⁶ In the 1793 *Religion within the Boundaries of Mere Reason*, Kant indicates that the former need not imply the latter: “From the fact that a being has reason does not at all follow that, simply by virtue of representing its maxims as suited to universal legislation, this reason contains a faculty

of the awareness of obligation is our most immediate concern. For Kant, this element in human moral awareness is explained by the interaction of two essential features of human psychology, each stemming from a deep aspect of human nature. As living beings, humans can be moved to act by mere animal inclination, or by persistent sources of pleasure-occasioned desire arising from the needs and instincts that we have simply as living. As rational beings, humans are beholden to the moral law as a self-legislated criterion of right action. Thus, as both rational and living, we can be induced by mere inclination to act against the prescription of the moral law, which we are in turn required to heed (cf. 6:379). In this way, the human awareness of obligation takes the form of the consciousness of *duty*:

[T]he moral law is for [beings affected by needs and sensible motives] an *imperative* that commands categorically because the law is unconditional; the relation of such a will to this law is *dependence* under the name of obligation, which signifies a *necessitation*, though only by reason and its objective law, to an action which is called *duty* because a choice that is pathologically affected (though not thereby determined hence still free) brings with it a wish arising from *subjective* causes, because of which it can often be opposed to the pure objective determining ground and thus needs a resistance of practical reason which, as moral necessitation, may be called an internal but intellectual constraint. (5:32)

Obligation—necessitation by reason—is thus experienced as duty only where choice is both free and pathologically affected, precisely because in these cases the agent’s faculty of choice will have multiple and possibly conflicting determining grounds. In humans, one sort of ground is occasioned by feelings of pleasure or aversion. The ground arises when a human agent finds the

of determining the power of choice unconditionally, and hence to be ‘practical’ on its own; at least, not so far as we can see. The most rational being of this world might still need incentives, coming to him from the objects of inclination, to determine his power of choice” (6:26n).

representation of the reality of a certain possible effect of her doing something (e.g., the representation of a beautiful, newly constructed house) to be connected in herself to one of these feelings. When this connection holds and the agent is aware of it, the feeling attendant to the represented effect can both cause the agent to form a desire (e.g., to want to draw up the plans) and can provide that desire with a certain degree of motivating force. If the agent acts on a desire caused in this way, the represented possible effect determines her faculty of choice to action. Since humans are also free, the representation of a law can act for them as a determining ground of choice as well. In choice involving recognized moral obligation, the representation of the moral law—the recognition that it enjoins or forbids a particular action—acts as such a ground. In moral choice, then, certain human desires can be formed independently of any felt pleasure or aversion. When an agent acts on one of those desires, her representation of the moral law determines her faculty of choice to action.³⁷

Now it is a familiar fact that desires of the first sort can conflict with desires of the second sort. But in the awareness of moral obligation, reason presents the moral law as not just a possible but as a normatively necessary ground of the determination of one's power of choice. Since the feelings that provide the motivating force for pathologically-settled desires are not directly responsive to normative necessity, humans need a "resistance of practical reason" (5:32): in order to act as morality commands, we are required to overcome the force of our pleasure- or aversion-occasioned desires by rational means. In the awareness of obligation, Kant thinks, we are aware that we should and that we can, but also that we will not necessarily do so. Consider in this connection the famous example of the prince's subject:

But ask [somebody] whether, if his prince demanded, on pain of...immediate execution, that he give false testimony against an honorable man whom the prince would like to destroy under a plausible pretext, he would consider it possible to

³⁷ For the material in this paragraph, see the discussion in the *Metaphysics of Morals* on the relation of the faculties of the human mind to moral laws (6:211-14).

overcome his love of life, however great it may be. He would perhaps not venture to assert whether he would do it or not, but he must admit without hesitation that it would be possible for him. (5:30)

In this way, the awareness of obligation involves an implicit awareness of oneself as needing what Kant calls *virtue* (5:32-3): “the capacity and considered resolve to withstand...what opposes the moral disposition within us” (6:380).

Virtue, the ability to overcome the motivational force of one’s pleasure-occasioned desires in the name of rational requirement, thus amounts to the capacity to mobilize a countervailing motivational force. The source of this force, since it cannot simply be more pathological feeling, must be pure practical reason itself in the form of the moral law.³⁸ Kant explains how this latter is possible in the second *Critique*’s chapter on the incentives of pure practical reason.

Every determination of choice, Kant thinks, “proceeds from the representation of a possible action to the deed through the feeling of pleasure or displeasure” (6:399). Thus, considered as an incentive, the moral law is itself a cause of feeling in an agent. This surprising piece of causation

³⁸ Despite a significant amount of recent interest in the topic of Kant’s account of virtue, these claims enjoy widespread interpretive assent. In her recent overview essay, for example, Lara Denis summarizes Kantian virtue as “a continually cultivated capacity to master [one’s] inclinations so as to fulfill all her duties, a capacity whose cultivation and exercise is motivated by respect for the moral law” (“Kant’s conception of virtue” in Paul Guyer, ed. *The Cambridge Companion to Kant and Modern Philosophy*. Cambridge: Cambridge University Press, 2006, pp. 505-37; here p. 513). Similar claims are defended in works as varied in emphasis as Allen W. Wood. *Kant’s Ethical Thought*. Cambridge: Cambridge University Press, 1999, pp. 329-31; Paul Guyer. *Kant on Freedom, Law, and Happiness*. Cambridge: Cambridge University Press, 2000, pp. 303-23; and Adam Cureton and Thomas E. Hill Jr. “Kant on Virtue: Seeking the Ideal in Human Conditions”, in Nancy E. Snow, ed., *The Oxford Handbook of Virtue*. Oxford: Oxford University Press, 2018, pp. 263-274, here esp. pp. 263-265.

takes the form of the representation of the moral law infringing on two persistent forms of self-regarding inclination, each of which Kant takes to be intrinsic to pathologically affected reasoning agents: self-love and self-conceit. Self-love is the propensity to incorporate the ends proposed by one's own inclinations into one's maxims because they are the ends proposed by one's own inclinations. It is the tendency to treat oneself, understood as the subject of all of one's own agreeable experiences, as the object of one's own benevolent concern (5:73). Self-conceit goes a step further: it is a propensity to treat oneself, as the subject of one's own inclinations, as worthy of "esteem", and therefore as providing an *unconditional* principle of action (*Ibid.*):³⁹

[We] find our nature as sensible beings so constituted that the matter of the faculty of desire (objects of inclination, whether of hope or fear) first forces itself upon us, and we find our pathologically determinable self, even though it is quite unfit to give universal law through its maxims, nevertheless striving antecedently to make its claims primary and originally valid, just as if it constituted our entire self. The propensity to make oneself as having subjective determining grounds of choice into the objective determining ground of the will in general can be called *self-love*; and if self-love makes itself lawgiving and the unconditional practical principle, it can be called *self-conceit*. (5:74)

³⁹ For a helpful discussion of these issues see Stephen Engstrom. "The *Triebfeder* of pure practical reason", in Andrews Reath & Jens Timmerman, eds. *Kant's Critique of Practical Reason: A Critical Guide*. Cambridge: Cambridge University Press, 2010, pp. 90-118, here esp. pp. 101-118. See also Andrews Reath. "Kant's Theory of Moral Sensibility: Respect for the Moral Law and the Influence of Inclination", in his *Agency and Autonomy in Kant's Moral Theory*. Oxford: Oxford University Press, 2006, pp. 8-32, here p. 15. For an enlightening treatment of the concept of esteem in Kant's moral thinking, see Allen W. Wood. *Kant's Ethical Thought*, pp. 27-33.

According to Kant, the representation of the moral law as a necessary determining ground of one's power of choice "restricts" self-love—restricts, that is, the appropriate exercise of the propensity by ruling out adopting some but not other maxims on grounds of self-love as inconsistent with morality—and "strikes down" self-conceit, effectively rejecting as rationally unacceptable the exercise of the propensity altogether (5:73). It is in the striking down of self-conceit that the representation of the moral law has its most important effect on feeling. The recognition of the moral law as the necessary determining ground of choice in the awareness of obligation, Kant says, "infringes without end" on the tendency of the agent to treat her pathologically affected self as a source of unconditional value. He continues:

[W]hat in our own judgment infringes upon our self-conceit humiliates. Hence the moral law unavoidably humiliates every human being when he compares it with the sensible propensity of his nature. If something represented as a determining ground of our will humiliates us in our self-consciousness, it awakens a *respect* insofar as it is positive and a determining ground. (5:74)

Since the consciousness of the moral law in the awareness of obligation is in part an awareness of its unconditional applicability (5:31), in that awareness, the agent implicitly compares the idea of herself as the source of unconditional value inherent in her standing propensity to self-conceit with the idea of the moral law as the source of unconditional value. This unflattering comparison engenders a feeling that, from the perspective of the pathologically affected self of self-conceit, registers as a form of humiliation or self-contempt. This same feeling, considered from the perspective of the self of autonomous moral agency, registers as the feeling of respect for the moral law as a necessary determining ground of one's power of choice.

To whatever extent the representation of the moral law incites this "moral feeling" (5:75) in an agent, and thus acts as the subjective determining ground of that agent's will, it therefore

constitutes her virtue, or the “resistance of practical reason” that we have seen Kant suggesting human moral agents are conscious of needing in the awareness of obligation.⁴⁰

Now, although Kant thinks that the feeling of respect for the moral law is a kind of basic endowment of human beings, he also thinks that it can and must be cultivated and strengthened. He explains in the *Religion within the Boundaries of Mere Reason*:

Since any consciousness of obligation depends on moral feeling to make us aware of the constraint present in the thought of duty, there can be no duty to have moral feeling or to acquire it; instead every human being (as a moral being) has it in him originally. Obligation with regard to moral feeling can only be to *cultivate* it and to strengthen it through wonder at its inscrutable source. (6:399-400)

A perfectly virtuous being would possess what Kant calls a “holy will”, a will incapable of adopting any maxim conflicting with the moral law (5:32). The notion of a holy or perfectly virtuous will or is an idea of reason in Kant’s technical sense—a standard generated within reason itself against which the objects of experience are to be measured, but to which no adequate such object can ever be given (A327/B384).⁴¹ Because in the awareness of obligation, an agent is

⁴⁰ See the quotation from 5:32 on page 26 above. Compare 6:23n and 6:395.

⁴¹ “Whoever would draw the concepts of virtue from experience, whoever would make what can at best serve as an example for imperfect illustration into a model for a source of cognition (as many have actually done), would make of virtue an ambiguous non-entity, changeable with time and circumstances, useless for any sort of rule. On the contrary, we are all aware that when someone is represented as a model of virtue, we always have the true original in our own mind alone, with which we compare this alleged model and according to which alone we estimate it. But it is this that is the idea of virtue, in regard to which all possible objects of experience do service as examples (proofs of the feasibility, to a certain degree, of what the concept of reason requires), but never as archetypes. That no human being will ever act adequately to what the pure

conscious that she needs virtue, and because she is also conscious in that awareness that she is not perfectly virtuous—because her consciousness of the moral law as an *imperative* contains an implicit awareness of the possibility of her failing to abide it—Kant thinks that the awareness of obligation not only makes her conscious that she is required to be virtuous, it also makes her conscious that she must *cultivate* her virtue. It does this latter, he thinks, by presenting her with idea of a holy will as a standard toward which she must perpetually strive:

In the supremely self-sufficient intelligence, choice is rightly represented as incapable of any maxim that could not at the same time be objectively a law, and the concept of holiness, which on that account belongs to it, puts it, not indeed above all practical laws, but above all practically restrictive laws, and so above obligation and duty. This *holiness of will* is nevertheless a practical idea, which *must necessarily serve as a model to which all finite rational beings can only approximate without end and which the pure moral law, itself called holy because of this, constantly and rightly holds before their eyes*; the utmost that finite practical reason can effect is to make sure of this unending progress of one's maxims toward this model and of their constancy in continual progress, that is, virtue; and virtue itself, in turn, at least as a naturally acquired ability, can never be completed, because assurance in such a case never becomes apodictic certainty and, as persuasion, is very dangerous. (5:32-3; my emphasis)

idea of virtue contains does not prove in the least that there is something chimerical in this thought. For it is only by means of this idea that any judgment of moral worth or unworth is possible; and so it necessarily lies at the ground of every approach to moral perfection, even though the obstacles in human nature, as yet to be determined as to their degree, may hold us at a distance from it” (A315/B371-72).

Having recognized these connections, we are finally in position to understand the advertised relation between moral obligation and personal identity. The human awareness of obligation is the consciousness of one's pure practical reason providing a constraint on one's action. The awareness of this constraint involves both the awareness through feeling of certain persistent and deep-seated motivational propensities being constrained, and the intellectual awareness of the necessity with which reason prescribes their constraint. Together, the awareness of the necessity of acting in accordance with the moral law and the awareness of the countervailing motivational tendencies amount or give rise to an awareness of oneself as required to cultivate one's virtue. This, in turn, amounts to the awareness of oneself as required to undertake a temporally extended practical project, which project Kant thinks that the agent rightly recognizes will continue throughout her entire agential career. Finally, implicit in that last sort of self-awareness is an awareness of oneself as a temporally extended thing. Ought, after all, implies can. In this way, pure practical reason endows the human being with a certain temporal structure a priori. In the awareness of obligation, the agent finds that she treats herself as the numerically identical subject of a self-legislated, temporally extended moral vocation. This is a kind of consciousness of the numerical identity of one's thinking self in different times. It is a satisfier of the identity criterion for personhood.⁴²

In this way, the awareness of obligation includes not merely the awareness of one's moral personality or freedom, and therefore one's status as an appropriate subject of action imputation, but also the awareness of one's psychological personality or identity through time. Its character

⁴² There is some reason to think that Kant had this idea in mind at least as early as the writing of the first edition of the first *Critique*. In a note dated somewhere between 1769 and 1770, he writes that all "sensation of *personhood*, namely of oneself as an *active principle*, is active"—i.e., does not involve the passivity characteristic of representing oneself as an object of inner or outer sense (R655 15:289; my emphasis). See also R6861, dated either from the late 1770s or from the 1780s: "The apperception of sensation is substance; that of self-activity is the person" (19:183).

accordingly provides a source of deep connection between Kant's notions of psychological and moral personality. The fact that the very same consciousness of ourselves in the awareness of obligation shows us both that we are identical through time and free means that it amounts to a consciousness of ourselves as persons *simpliciter*.

5. The Immortality of the Soul

On the account just offered, Kant thinks that human personhood is recognizable in the awareness of moral obligation and that the requirement that persons be aware of themselves as identical through time is satisfied in that awareness by humans' consciousness of themselves as finite beings necessitated by reason, and thus as both required to undertake a temporally extended project of moral self-development and as capable of so doing. The awareness that we have of ourselves as persons in moral life is thus neither an empirical perception of ourselves as persisting in time nor is it a purely intellectual grasp of ourselves as the disembodied denizens of a timeless noumenal realm. It is rather an awareness of ourselves as the subjects of moral demand, one that arises a priori in the distinctively human form of the awareness of obligation, the consciousness of duty. Because we are finite beings with the capacity to cultivate virtuous dispositions over time and with a recognizable responsibility for doing so, humans are persons.

Although it is from Kant's perspective extremely striking that in the awareness of moral obligation we can know ourselves as persons in this way, the knowledge nevertheless has important and characteristic limitations. One way in which it is limited—a particularly salient one from the perspective of the metaphysical tradition of which Kant is part—is that nothing in the human awareness of obligation can settle whether or not its subject will persist *indefinitely*. We cannot know, that is to say, even as a practical matter, whether our souls are immortal, or whether they continue beyond this life.

The question of the immortality of the soul is of paramount importance to Kant. He thinks that establishing the soul's immortality, along with establishing the existence of God and the freedom of the will, is the "final aim to which in the end the speculation of reason in its transcendental use is directed" (A798/B826). And although the critical philosophy ensures that the impulse to settle these matters in theoretical life must remain unsatisfied, Kant nevertheless thinks of that impulse itself as a dignified, urgent, and unavoidable one (A3/B7). Part of what lends dignity to this impulse, Kant thinks, is a practical need of reason (Bxxix-xxx), and this need is what ultimately gives us a kind of practical warrant to assert that our souls are immortal.

In the section of the second *Critique* on the dialectic of pure practical reason, Kant argues that, in contexts of deliberation and choice, we must accept the possibility that the highest good—happiness distributed in exact proportion to virtue (5:110)—can be achieved through the pursuit of virtue itself. He then argues that since we must for practical purposes assume that it is possible to achieve the highest good, we must also assume that it is possible for us to achieve its precondition, the holiness of the will:

The production of the highest good in the world is the necessary object of a will determinable by the moral law. But in such a will the *complete conformity* of dispositions with the moral law is the supreme condition of the highest good. This conformity must therefore be just as possible as its object is, since it is contained in the same command to promote the object. Complete conformity of the will with the moral law is, however, *holiness*, a perfection of which no rational being of the sensible world is capable at any moment of his existence. Since it is nevertheless required as practically necessary, it can only be found in an *endless progress* toward that complete conformity, and in accordance with principles of pure practical reason it is necessary to assume such a practical progress as the real object of our will. (5:122)

For time-bound, sensible beings such as ourselves, the achievement of holiness is in one sense recognizably impossible: it cannot be completed in any amount of time. However, it must also, in some other sense, be recognizably possible—we cannot, after all, reasonably pursue an end we have no reason to think we can accomplish. For this reason Kant thinks that we must conceive of ourselves as engaged in an “endless progress” toward holiness, that we must take the project of cultivating our virtue over time to continue without end, forever more closely approximating the hoped-for goal.⁴³ However if we are required to understand our moral vocation

⁴³ Kant expands on this point in the *Religion* in the course of interpreting the Christian thought that we can by subduing the evil in our nature become new persons (6:48; cf. Colossians 3:9-10). This Pauline idea fits naturally with Kant’s metaphysics of persons. Suppose that a human being were to achieve the end of subduing the evil in her nature, as the apostle enjoins us to do, becoming holy and thus well-pleasing to God thereby. For Kant, she would in so doing become a different person. After all, the vocation whose normative necessity had theretofore been apparent in her consciousness of duty would have come to its successful completion, and thus would no longer bear on her. Absent the ability to rightly recognize this vocation as her own, she would not be the same Kantian person. She would, through her infinitely extended course of effort, have literally made herself anew. These points about the metaphysics of conversion in disposition do not of course suggest that we might know ourselves as now one, now another person. For Kant, the “abandonment of evil” that constitutes conversion “is one and the same act” as our “progress in the good” (6:75n), and since this act cannot be completed in finite time, we can only regard it “as an ever-continuing striving for the better, hence as a gradual reformation of the propensity to evil” (6:48). In this way, we can know ourselves only as single, cross-temporally identical beings required to continually strive toward holiness, or as individual persons required to continually “put off” our “old selves” (6:75n). There is, however, room to hope that from the perspective of eternity, this infinite progression might be seen in its totality, and “judged by him who scrutinizes the heart (through pure intellectual intuition) to be a perfected whole even with respect to the deed (the life

in this way, then we are also required to treat ourselves, as the subjects of such a vocation, as continuing on without end:

This endless progress is, however, possible only on the presupposition of the existence and personhood of the same rational being continuing endlessly (which is called the immortality of the soul). (5:122)

Since we can have no theoretical basis for denying the claim that the soul is immortal—the appropriate theoretical attitude is agnosticism—we are entitled to assume it on the basis of a need of pure practical reason:

[W]e cannot affirm that we *cognize* and *have insight into* – I do not merely say the reality but even the possibility of [the ideas of God and immortality]. But they are, nevertheless, conditions of applying the morally determined will to its object given to it a priori (the highest good). Consequently their possibility in this practical reason can and must be *assumed*, although we cannot theoretically cognize and have insight into them. (5:4)

Throughout the second *Critique*, Kant is at pains to emphasize that the assumption of the soul's immortality is a mere postulate or presupposition, rather than a cognition of an immortal being. As we have already seen Kant emphasize in the Paralogisms, in the absence of any possible manifold of intuition corresponding to the immortal soul, its cognition is impossible for us (A350, B407). Nevertheless, Kant thinks that the assumption of the soul's immortality for practical purposes is enough to give the rational idea of immortality a kind of *practical* objective reality:

conduct)” (6:67). In this way, religiously-minded people might reasonably hope to be judged worthy for inclusion into the kingdom of God (cf. 6:72). For a different take on these passages, see Kitcher, *Kant's Thinker*, pp. 185-87. Thanks to Jonathan Ettl for discussion.

Now [the ideas of God, freedom, and immortality] receive objective reality through an apodictic practical law, as necessary conditions of the possibility of what it commands us *to make an object*, that is, we are instructed by it *that they have objects*, although we are not able to show how their concept refers to an object, and this is not yet cognition *of these objects*; for one cannot thereby judge synthetically about them at all or determine their application theoretically; hence one can make no theoretical rational use of them at all, in which use all speculative cognition of reason properly consists. (5:135)

The reality of the moral law, presented to us as a fact in the awareness of moral obligation, ensures that we are practically required to assume that the soul's immortality is also real. This latter idea achieves body and meaning, then, not on the familiar basis of being demonstrably applicable to appearances, but rather because the assumption of its reality is required as a practical matter in order for us to understand the ultimate aim of our practical endeavor under the moral law as achievable. Although this is less than the rational psychologist had hoped for, it is also rather more than Kant thinks that he can deliver. Kant returns to this theme in the second *Critique's* Dialectic:

[Presupposing the objective reality of the highest good] leads to...the problem in the solution of which speculative reason could do nothing but commit *paralogisms* (namely the problem of immortality) because it lacked the mark of permanence by which to supplement the psychological concept of an ultimate subject, necessarily ascribed to the soul in self-consciousness, so as to make it the real representation of a substance; this mark practical reason furnishes by the postulate of a duration required for conformity with the moral law in the highest good as the whole end of practical reason. (5:133)

6. Conclusion

Near the beginning of the *Religion*, Kant identifies three essential “constituent parts” (6:28) of the human being: animality, coordinated with the idea of a human as a living being; humanity, coordinated with the idea of a human as a rational being; and personality, coordinated with the idea of a human as a responsible being. With this framework in mind, we can recognize that for Kant, there is room for more than one set of identity conditions applying to the human being, with each set corresponding to one of the distinct conceptual categorizations under which humans fall. First, there are the empirical conditions for bodily identity, discussed under the heading “persistence in life” in the first *Critique*, and corresponding to the idea of the human as an animal or a living thing. Second, there are the conditions for the logical identity of humans as thinking things, met in pure apperception, corresponding to the idea of the human as a rational or thinking thing. Third, there are the conditions for the real identity of the subject of a temporally extended moral vocation, met with a priori in the consciousness of duty, corresponding to the idea of the human as responsible or as a person. The interpretive tradition over the last fifty years has been largely focused on the second set of conditions, secondarily on the first. My interest in this essay has been with the third.

The central idea of Kant’s positive metaphysics of personhood, as I have presented it here, is that in the awareness of moral obligation, we recognize ourselves as the subjects of temporally extended moral vocations and therefore as persons. Kant supplements this claim with one that we are practically required to suppose that the subject of that vocation is immortal. Whatever may ultimately be made of either thought, it seems clear enough that neither constitutes a regression to an implausible rationalist metaphysics of personhood. On the contrary, it is hard to think of a more characteristically Kantian notion than the idea that, in the recognition of our finite natural condition, we are endlessly compelled by reason to strive to be more and better than we are.

Kant himself was inclined to view this notion of personhood as one of the *Critique of Practical Reason's* central insights. He famously concludes the book with reference to the “starry heavens above and the moral law within”. The latter, he says,

...infinitely raises my worth as an *intelligence* by my personality, in which the moral law reveals to me a life independent of animality and even of the whole sensible world, at least insofar as this may be inferred from the purposive determination of my existence by this law, a determination not restricted to the conditions and boundaries of this life, but reaching into the infinite. (5:162)⁴⁴

⁴⁴ Special thanks are due to Michael Friedman, David Hills, and Allen Wood for much helpful discussion about earlier drafts of this essay.