

Трансцендентальный журнал 2013-2024

ISSN 2079-8784

URL - <http://ras.jes.su>

Все права защищены

Выпуск 3 Том 3. 2022

## Essentially Embodied Kantian Selves and The Fantasy of Transhuman Selves

Ханна Роберт

*США*

### Аннотация

By “essentially embodied Kantian selves,” I mean necessarily and completely embodied rational conscious, self-conscious, sensible (i.e., sense-perceiving, imagining, and emoting), volitional or willing, discursive (i.e., conceptualizing, judging, and inferring) animals, or persons, innately possessing dignity, and fully capable not only of free agency, but also of a priori knowledge of analytic and synthetic a priori truths alike, with egocentric centering in manifestly real orientable space and time. The basic theory of essentially embodied Kantian selves was spelled out by Kant over the course of slightly less than two decades, between 1768 and 1787, but above all, it flows from an empirical realist and metaphysical reading of the “Refutation of Idealism” that Kant inserted into the Postulates of Empirical Thought section in the 1787 edition of the first Critique. In my opinion, all rational but also “human, all-too-human” creatures like us are, synthetic a priori necessarily, essentially embodied Kantian selves. Let’s call that the essentially embodied Kantian selves thesis, or for short, EEKST. If EEKST is true, then it’s synthetic a priori impossible for the selves of creatures like us to exist independently of our own living organismic animal bodies or beyond the deaths of those bodies, whether temporarily or permanently, by any means whatsoever. Indeed, the very ideas of disembodied selves, their survival after death, and of human immortality, while minimally logically consistent, are in fact conceptually empty and incoherent, even over and above the synthetic a priori impossibility of such things, since the term “myself” indexically picks out an essentially embodied Kantian self, all of whose core features require grounding in a particular living organismic animal body. According to the recent and contemporary movement of transhumanism, the selves of creatures like us can not only exist independently of our bodies, as functional systems of representational content

that are inherently able to be implemented or realized in digital-mechanical technology and uploadable to servers, but also to survive accidental or natural human death in server-limbo, then be downloaded into technologically enhanced partially mechanical humanoid bodies or even into wholly artificially-created completely mechanical non-humanoid bodies, survive in these new implementations or realizations for an indefinitely long time, repeat that process, and possibly even become immortal. Transhumanism is in fact metaphysically equivalent to Swedenborgianism, which Kant so effectively criticizes and wittily derides in his 1766 book, *Dreams of a Spirit-Seer Elucidated by Dreams of Metaphysics*. Moreover, and more importantly, if EEKST is true, then, just like Swedenborgianism, transhumanism is not only conceptually empty and incoherent, but also synthetic a priori impossible. And what's more, it's also existentially and morally reprehensible. In short, then, the belief in transhuman selves is nothing but a reprehensible noumenal fantasy or Hirngespinnst.

**Ключевые слова:** Immanuel Kant, Selves, Embodiment, Swedenborgianism, Transhumanism

**Дата публикации:** 29.12.2022

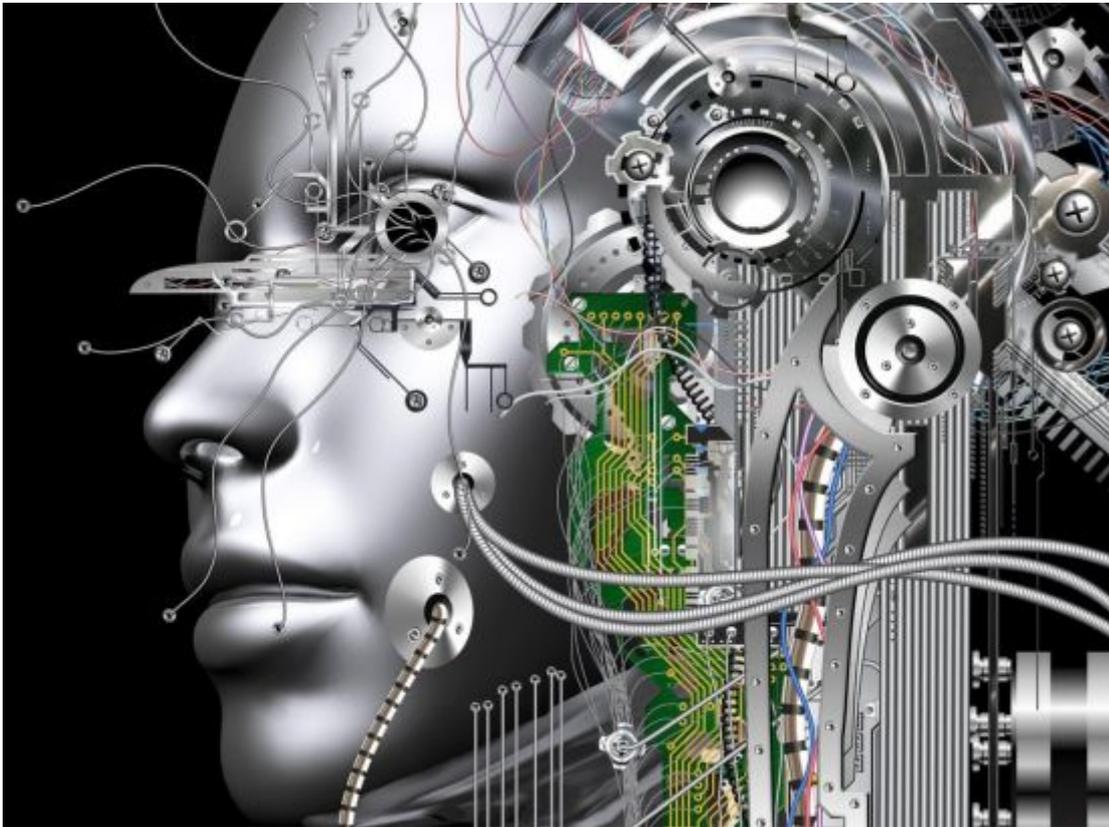
**Ссылка для цитирования:**

Ханна Р. Essentially Embodied Kantian Selves and The Fantasy of Transhuman Selves // Трансцендентальный журнал – 2022. – Том 3. – Выпуск 3 [Электронный ресурс]. URL: <https://transcendental.su/S271326680021060-6-1> (дата обращения: 20.05.2024). DOI: 10.18254/S271326680021060-6

1



*“Who Was Swedenborg?” (Swedenborg Foundation, 2022)*



*“Face of the Future” (Guardian, 2018)*

<sup>3</sup> In Stockholm there dwells a certain [Swedenborg], a gentleman of comfortable means and independent position. For the last twenty years or more he has, as he tells us, devoted himself exclusively to cultivating the closest contact with spirits and with the souls of the dead, and, in exchange, to giving them information about this present world, to composing hefty volumes devoted to his discoveries, and periodically travelling to London in order to supervise their publication.... [Swedenborg] distinguishes between the outer and inner memory in humankind. A person has outer memory as someone belonging to the visible world, whereas a person has inner memory in connection with the spirit-world. [Swedenborg’s] own superiority consists in the fact that, already in this life, he sees himself as a person who belongs to the community of spirits and that he is recognized as someone belonging to that community. It is also in this inner memory that everything, which has vanished from outer memory, is conserved, none of a person’s representations ever getting lost. After death, the memory of everything which had ever entered his soul and which had so far remained concealed from him, goes to make up the complete book of his life (*DSS 2: 354, 362*).

<sup>4</sup> As humans, we are defined by, among other things, our desire to transcend our humanity. Mythology, religion, fiction and science offer different versions of this dream. Transhumanism—a social movement predicated on the belief that we can and should leave behind our biological condition by merging with technology—is a kind of feverish amalgamation of all four. Though it’s oriented toward the future, and is fueled by excitable speculation about the implications of the latest science and technology, its roots can be glimpsed in ancient stories like that of the Sumerian king Gilgamesh and his quest for immortality.... Transhumanism represents a desire to obliterate the

boundary between human bodies and machines, and a confusion in the first place as to the distinction between the two (Guardian, 2018).

5

## Introduction

6 By “essentially embodied Kantian selves,” I mean necessarily and completely embodied rational conscious, self-conscious, sensible (i.e., sense-perceiving, imagining, and emoting), volitional or willing, discursive (i.e., conceptualizing, judging, and inferring) animals, or persons, innately possessing dignity, and fully capable not only of free agency, but also of a priori knowledge of analytic and synthetic a priori truths alike, with egocentric centering in manifestly real orientable space and time. The basic theory of essentially embodied Kantian selves was spelled out by Kant over the course of slightly less than two decades—between 1768 and 1787—in the proto-Critical 1768 essay, “Concerning the Ultimate Ground of the Differentiation of Directions in Space,” in the *Transcendental Aesthetic* and the *Transcendental Analytic*, especially including the Deduction of the Pure Concepts of the Understanding, in the *Critique of Pure Reason* in 1781 (i.e., the first or A edition) and 1787 (i.e., the second or B edition), and in the 1786 essay, “What is Orientation in Thinking,” but above all, it flows from an *empirical realist* and *metaphysical* reading of the “Refutation of Idealism” that Kant inserted into the Postulates of Empirical Thought section in the 1787 edition of the first *Critique*—henceforth, for convenience, “the RI.” This basic theory is also supplemented and supported by *practical* accounts of essentially embodied Kantian selves that can be found in the *Critique of Practical Reason*, the *Critique of the Power of Judgment*, *Religion Within the Bounds of Mere Reason*, and the *Anthropology from a Pragmatic Point of View*.

7 Or as the pre-Critical Kant much more compactly puts it in 1766: “*My soul is wholly in my whole body, and wholly in each of its [organic] parts*” (*DSS 2: 325 italics* in the original, bracketted word taken from the original source quoted by Kant, cited in the editorial note on p. 449, n. 11).

8 In my opinion, all rational but also “human, all-too-human” creatures like us—for example, the readers of this essay—are, synthetic a priori necessarily,<sup>1</sup> essentially embodied Kantian selves. Let’s call that *the essentially embodied Kantian selves thesis*, or for short, EEKST. If EEKST is true, then it’s synthetic a priori impossible for the selves of creatures like us to exist independently of our own living organismic animal bodies or beyond the deaths of those bodies, whether temporarily or permanently, by any means whatsoever. Indeed, the very ideas of disembodied selves, their survival after death, and of human immortality, while minimally logically consistent, are in fact conceptually empty and incoherent, even over and above the synthetic a priori impossibility of such things, since the term “myself” indexically picks out an essentially embodied Kantian self, all of whose core features require grounding in a particular living organismic animal body.

9 Now, the recent and contemporary movement of *transhumanism* is a

10 social and philosophical movement devoted to promoting the research and development of robust human-enhancement technologies. Such technologies would augment or increase human sensory reception, emotive ability, or cognitive capacity as well as radically improve human health and extend human life spans. Such

<sup>11</sup> More specifically, transhumanism claims that the selves of creatures like us can not only exist independently of our bodies, as functional systems of representational content that are inherently able to be implemented or realized in digital-mechanical technology and uploadable to servers, but also to survive accidental or natural human death in server-limbo, then be downloaded into technologically enhanced partially mechanical humanoid bodies or even into wholly artificially-created completely mechanical non-humanoid bodies, survive in these new implementations or realizations for an indefinitely long time, repeat that process, and possibly even become immortal. Now, it's true that the belief in and desire for digital-mechanical immortality isn't strictly required for transhumanism: as a recent survey showed, as many as 23.8% of contemporary transhumanists *don't* actually desire digital-mechanical immortality (Wikipedia, 2022). But of course, that also implies that as many as 76.2% of contemporary transhumanists *do* actually desire (and therefore, presumably, also believe in) digital-mechanical immortality.

<sup>12</sup> Leaving aside for a moment (although I will come back to this, in section III below) transhumanism's inherently questionable metaphysical appeal to what I call *the mechanistic worldview* (Hanna, 2022), including *the strong thesis of artificial intelligence*, aka strong AI, and also its all-too-trendy, breathless, and high-gloss futurological appeal to the Promethean wonders of digital technology, transhumanism is in fact metaphysically equivalent to *Swedenborgianism*, which Kant so effectively criticizes and wittily derides in his 1766 book, *Dreams of a Spirit-Seer Elucidated by Dreams of Metaphysics*. Moreover, and more importantly, if EEKST is true, then, just like Swedenborgianism, transhumanism is not *only conceptually empty and incoherent*, but also *synthetic a priori impossible*. And what's more, it's also existentially and morally *reprehensible* (Gare, 2016). In short, then, the belief in transhuman selves is nothing but a reprehensible noumenal fantasy or *Hirngespinnst*.

### <sup>13</sup> **Transcendental Idealism, The RI, and EEKST**

<sup>14</sup> Kant's metaphysical doctrine of *transcendental idealism*, as it's presented in the *Critique of Pure Reason* and the *Prolegomena to Any Future Metaphysics*, says that (i) necessarily, all the actual and possible proper (i.e., directly intuitionally accessible and/or meaningfully conceptualizable) objects of human cognition are manifestly real *appearances* or *phenomena*, which are also therefore *knowable* by us, at least to some non-trivial extent, and not inherently non-manifest or hidden *things-in-themselves* or *noumena*, which are therefore completely *unknowable* by us (let's call that Kant's *cognitive idealism*), and (ii) necessarily, the basic *ontological* structures of all actual and really possible things in the manifestly real world isomorphically correspond to the innately specified basic *mentalist* structures of our rational human cognitive capacities, and *not* conversely (let's call that Kant's *transcendental conformity thesis*) (see, e.g., Hanna, 2001: esp. chs. 1-2, 2006: esp. chs. 1-4). Transcendental idealism in this sense differs sharply not only from Berkeley's *metaphysical or dogmatic idealism*, which says (i) that matter is impossible, (ii) all the actual and possible proper objects of human cognition are ideas, and (iii) that necessarily, for actual and possible object of any kind, call it X, X is either an idea in a conscious mind or X is a conscious mind, but also from

Cartesian *skeptical or problematic idealism*, which says that possibly, nothing exists outside my own conscious states.

<sup>15</sup> More specifically, in sharp contrast to Berkeleyan metaphysical idealism, (ia) Kant's cognitive idealism does *not* say that matter is impossible, (ib) Kant's cognitive idealism does *not* say that all the actual and possible proper objects of all human cognition are nothing but *ideas* (i.e., objects existing merely in inner sense), and (ic) Kant's cognitive idealism and transcendental conformity thesis do *not* apply to *all* actual and possible objects of any kind whatsoever, including things-in-themselves or noumena, but instead *only* to manifestly real appearances or phenomena. And in equally sharp contrast to Cartesian skeptical or problematic idealism, Kant's cognitive idealism does *not* say that it is possible that nothing exists outside my conscious states (i.e., inner sense): on the contrary, Kant's cognitive idealism implies *that necessarily, something that's manifestly real actually exists outside my conscious states (i.e., inner sense) in space*. And this is in fact the conclusion of the "Refutation of Idealism" in the B or 1787 edition of the first *Critique*. Indeed, Kant regards both of these views—i.e., the Berkeleyan and Cartesian views alike—as inherently skeptical.

<sup>16</sup> Moreover, in the B Preface Kant famously says of his *anti-skeptical* philosophical predecessors, most notably Descartes himself and Thomas Reid, that

<sup>17</sup> it always remains a scandal of philosophy and universal human reason that the existence of things outside us (from which after all get the whole matter for our cognitions, even for our inner sense) should have to be assumed **on** [the basis of] **faith** (auf Glauben), and that if it occurs to anyone to doubt it, we should be unable to answer him with a satisfactory proof (*CPR* Bxxxix n.).

<sup>18</sup> Why do these purported refutations of idealism presented by Kant's anti-skeptical predecessors all fail? Kant's diagnostic insight, which he works out at length in the Paralogisms of Pure Reason (*CPR* A341-405/B399-432), is that the purported refutations all presuppose *the Cartesian model of the mind*, which effectively generates the very worries the refuters are trying so hard to refute. According to the Cartesian model of the mind, the inner world of conscious experiences and the outer world of material objects are at once (i) essentially different and ontologically distinct from one another, in that their basic natures are incompatible (because the inner or mental is intrinsically immaterial and non-spatial, whereas the outer or physical is intrinsically material and extended in space), so it is metaphysically possible for one to exist without the other (i.e., *ontological dualism*), and also (ii) epistemically mutually independent of one another, in that from the veridical cognition or knowledge of the one, no veridical cognition or knowledge of the other can ever be directly accessed or immediately inferred (i.e., *indirect epistemic realism*). Hence the pre-Kantian anti-skeptic is driven by a sort of philosophical despair to rely upon either a rationally undemonstrated "faith" in the existence of a non-deceiving God (as, e.g., in Descartes's classical Rationalist epistemology) or in the dictates of common sense (as, e.g., in Reid's common sense realism). In other words, one basic aim of the RI is to provide a fundamental critique of the Cartesian model of our mind, and to reject both its ontological dualism and its indirect realism.

<sup>19</sup> More specifically, and most importantly for my purposes in this essay, Kant believes that if we critically liberate ourselves from the Cartesian model of the mind, then we can be both "transcendental idealists" *and* "empirical realists" (i.e., manifest realists, including direct perceptual realists). So the RI is intended not only to refute

Berkeley's metaphysical idealism and also Descartes's skeptical idealism, and correspondingly to criticize the Cartesian model of the mind, but also and above all to offer a radically new anti-Cartesian anti-dualist, direct realist, transcendental idealist, and empirical or manifest realist model of the rational human mind, in order to ground EEKST.

20 The nerve of the RI can be found at B275-276 in the Postulates of Empirical Thought section in the first *Critique*; but I will also take into account the three "Notes" that immediately follow it in the text at B276-279, as well as a crucial footnote that Kant added at the last minute to the B or second edition Preface (*CPR* Bxxxix-xli). The view that Kant is aiming to refute is what he officially calls "skeptical idealism" or "problematic idealism":

21 [T]he **skeptical idealist** [is] one who **doubts** [the existence of matter], because he holds [matter and its existence] to be unprovable (*CPR* A377, boldfacing in the original).

22 Problematic idealism ... professes only our incapacity for proving an existence outside us from our own [existence] by means of immediate experience (*CPR* B275).

23 Skeptical or problematic idealism (the Cartesian skeptic's view) says that *possibly* the external world does not exist. This is to be sharply contrasted with what he officially calls "dogmatic idealism" (Berkeley's view), which takes the modally stronger position that the external or material world "is false and impossible," i.e., that the external or material world both actually and necessarily does not exist (*CPR* B274). Since Kant takes on the modally weaker form of skepticism, he has of course given himself a heavier burden of proof than would be required to refute Berkeley alone, since it is always harder to show that something is impossible than to show merely that its denial is actual or possible. But on the other hand, if Kant can show that skeptical or problematic idealism is false, then that will also suffice to show that dogmatic idealism is false, and more generally that "material idealism"—which is the inclusive disjunction of skeptical or problematic and metaphysical or dogmatic idealism (*CPR* B274)—is false. So if the RI is sound, then it will kill three skeptical birds with one argumentative stone.

24 Let's now look at the RI in detail. For each step I will offer a short commentary. Then I'll develop some criticisms of the overall argument, and finally I'll present a charitable interpretation that—in my opinion—turns the RI into a sound argument.

25 (1) "I am conscious of my existence as determined in time" (*CPR* B276).

26 *Commentary on step 1.* Obviously, this step can be directly compared and contrasted with Descartes's *cogito*: Necessarily, "I am, I exist" is true whenever I say or think it. Kant begins with what he elsewhere in the first *Critique* calls "empirical apperception" (*CPR* A107). Empirical apperception is empirical self-consciousness, or empirical reflective consciousness. So what Kant is saying here is that I have an empirical reflective consciousness *of* myself, as I consciously exist in "inner sense." Inner sense for Kant is the subject's intuitional awareness of a temporal succession of representational contents (*CPR* A22/B37, A107, B152-155, A357-359, A361-363, B420, B422-423 n.). Intuitional awareness, in turn, is (i) immediate or directly referential, (ii) sense-related, (iii) singular, (iv) object-dependent, and (v) logically prior to thought or nonconceptual (*CPR* A19/B33, A51/B75, B132, B146-147, A320/B377) (*Prolog* 4: 281-282). Occasionally in the first *Critique* Kant confuses inner sense and empirical

apperception by calling them both “consciousness.” But when he is being careful, we can see that he invokes a distinction between (i) a first-order unreflective reflexive consciousness of the phenomenal contents (whether objectively representational or merely sensory) of one’s own mental state, and (ii) a second-order reflective consciousness of first-order consciousness. In one of the *Reflexionen* and in the *Prolegomena* he says this of inner sense:

27 (The inner sense) Consciousness is the intuition of its self (*R* 5049; 18: 72).

28 [The ego] is nothing more than the feeling of an existence without the slightest concept and is only the representation of that to which all thinking stands in relation (*Prol* 4: 334 n.).

29 By contrast, he says of empirical apperception that it is “**one** consciousness of myself” through which “I can say of all perceptions that I am conscious of them” (*CPR* A122, boldfacing in the original). And in the *Anthropology* he distinguishes usefully between what he calls “taking notice of oneself” (*das Bemerken*), that is, an unreflective reflexive consciousness of oneself in inner sense at a given time, as opposed to “observing oneself” (*Beobachten*) (*A* 7: 132), that is, the introspective function of empirical apperception, which is repeatable over time and directly accessible via memory.

30 This difference between two levels of consciousness is crucial to Kant’s argument against problematic idealism. To use some non-Kantian terminology borrowed from William James and Thomas Nagel, inner sense is not only a “stream of consciousness” (James, 1950: vol. 1, ch. IX, pp. 224-290), but also captures “what it is like to be, for an organism” (Nagel, 1979: p. 166): inner sense is a constantly-changing succession of unreflectively reflexive egocentric phenomenal states in a human or nonhuman animal cognizer. In other words, inner sense is the *phenomenal consciousness* of an animal cognizer. Empirical apperception, by contrast, is a second-order judgmental consciousness of myself as a singular or individuated first-order stream of unreflectively reflexive representations. The propositional element in empirical apperception makes it imperative that we further distinguish it from what Kant variously calls “pure apperception,” “transcendental apperception,” or “the original synthetic unity of apperception.” This is an a priori or empirically underdetermined, spontaneous (i.e., unconditioned or unprecedented, creative), innate capacity for anonymous content-unification and for propositional and conceptual self-representation in general: more precisely, it is a universal capacity for attaching the cognitive prefix “I think” to any concept-involving representational content of the mind whatsoever (*CPR* B131-139, A341-348/B399-406) (see also Senderowicz, 2021). Empirical apperception, which presupposes transcendental apperception, is perhaps best regarded as the realization of that innate spontaneous capacity under concrete empirical conditions. Through empirical apperception, by carrying out an introspective judgment, I become conscious of my own first-order consciousness *as* constituting a determinate conscious human individual self: “**I**, as a thinking being, am an object of inner sense, and am called ‘soul’” (*CPR* A342/B400, boldfacing in the original).

31 Kant’s idea in this first step of the RI, then, is that even the most refractory skeptic would have to allow for the bare fact of such empirical introspection. To deny it would entail either (i) that we are always unconscious, or (ii) that even if we are sometimes conscious, then we are never conscious of our own consciousness (i.e., self-

conscious or meta-conscious), or (iii) that even if we are sometimes self-conscious or meta-conscious, then we are never able to make first person psychological reports. There may well be living biological human creatures that are always unconscious (e.g., humans in persistent vegetative states), or animals that have consciousness without self-consciousness or meta-consciousness (e.g., newborn human infants and cats), or animals who have self-consciousness or meta-consciousness without the capacity for carrying introspective discursive judgments or self-referring psychological reports (e.g., human toddlers and adult apes): but these are not creatures sharing our rational human cognitive constitution.

<sup>32</sup> (2) “All determination in time presupposes something **persistent** in perception” (*CPR* B276, boldfacing in the original).

<sup>33</sup> *Commentary on step 2.* For Kant, to “determine” something X, is either (necessarily or contingently) to ascribe or apply some definite attribute (a quality or property) to X, or to show how X enters (necessarily or contingently) as a relatum into some definite relation, and thereby takes on the attribute of belonging to that relation, or to show how X (necessarily or contingently) supports some definite relation. That all time-determination presupposes “that which persists,” is a direct consequence of the arguments given by Kant in support of the first Analogy of Experience, the “principle of the permanence of substance” (*CPR* A182-189/B224-232). In the first Analogy Kant asserts

<sup>34</sup> that which persists, in relation to which alone all temporal relations of appearances can be determined, is substance in the appearance, i.e., the real in the appearance, which as the substratum of all change always remains the same (*CPR* B225).

<sup>35</sup> The rationale behind this is the plausible thought that every change of attributes or relations in time requires something which remains the same throughout those changes. Now when we apply Kant’s reasoning to strictly psychological phenomena, it grounds the conclusion that every determinate sequence of successive changes of conscious mental contents in time requires some or another unchanging substratum (something which persists) to which those changes are directly ascribed or applied. We need not, for our purposes, accept Kant’s questionable further thesis—which seems to reflect a quantifier shift fallacy—to the effect that therefore there exists one and only one unchanging substratum to which every change of attributes or relations whatsoever is ascribed or applied, in order to buy into his original point. Nor need we, for our purposes, accept his questionable assumption that the unchanging substratum is either absolutely or even sempiternally persistent, rather than only relatively or temporarily persistent.

<sup>36</sup> In any case, the crucial point Kant is driving at in step 2 has to do with psychological “determination in time.” This phrase could be read as referring merely to the application of temporal predicates to my experiences. But I think that by using this phrase Kant is instead invoking something slightly stronger than this, namely, the *individuating determination* of my stream of experiences. This seems to be clearly implied by his use of the unusual phrase “my existence (*meines Daseins*) as determined in time”—as opposed to, say, “my experiences as determined in time”—and by his telling remark in the B Preface footnote to the effect that

<sup>37</sup> this consciousness of my existence in time is thus bound up identically (*identisch verbunden*) with the consciousness of a relation to something outside of me (*CPR* Bxl, italics added).

38 Kant's idea is that if I am to exist in inner sense as a constantly changing yet individuated stream of consciousness, and as an object of empirical apperception, an individual empirical self, then that stream must be essentially discriminable or uniquely identifiable, in the sense that it is distinguishable from any other such flow. To individuate my stream of consciousness is to confer upon all the separate elements of that stream—sensations, conceptions, images, judgments, etc.—a contingent yet *particular* ordering. This ordering activity occurs primarily through what Kant calls “the synthesis of reproduction” (*CPR* A100-102), which I think is best construed as our cognitive capacity for (short-term, long-term, semantic, episodic, and procedural) memory. In any case, what reproductive synthesis does is to convert that otherwise undifferentiated stream of mental contents into a single personal history or autobiography, whereby my inner life takes on a definite psychological shape or profile. Now according to Kant the individuating determination (through reproductive synthesis) of any such flow of changing mental contents, requires a relatively fixed underpinning or matrix, that is, a psychological persistent factor which “as the substratum of all [psychological] change always remains the same.”

39 It's hard to know precisely what Kant means by this, but I think that an analogy taken from physical geography is quite illuminating. A given river can be individuated only in relation to a spatially fixed material underpinning or matrix that includes its banks and riverbed, its beginnings and its terminus, and more generally the total path or locus it follows in getting from one end of the river to the other. Let us call this total path or locus its “geophysical route.” The Mississippi, for example, flows south along a certain route from northern Minnesota to the Gulf of Mexico, and could not be the self-same river unless it did so. Since the actual quantity of water in the Mississippi at any given time is always changing and running off into the Gulf of Mexico, the Mississippi would then seem to be uniquely defined by three physical factors: **first**, the fact that it is always water that is flowing in it, and not (say) beer or oil, **second**, the actual history of all the water that has already flowed through it over the years, and **third**, its route. This geophysical route can of course vary slightly within certain parameters, due to flooding or erosion, but those defining parameters continue to exist in a fixed way all the same. Like water in the Mississippi, which is always changing and running off into the Gulf, the contents of my stream of consciousness are always changing and running off into the past. So, analogously, my own individual empirical self's psychological life would seem to be uniquely defined and distinguished from all other such “streams of consciousness,” or conscious psychological processes, by three analogous psychological factors: **first**, the fact that only conscious human sensations, memories, concepts, etc., are flowing in it, and not (say) either non-sensory or “intellectual” intuitions or completely alien sorts of sense perceptions (*CPR* B71-72), **second**, the actual history of the various conscious mental contents that have already occurred in my psychological life, and **third**, its “psychological route”: a fixed underpinning or matrix that remains invariant in relation to the constantly changing flow of my sense-qualities and representations in time. All the psychological changes in my inner life as an empirical self must be changes *of*, or changes ascribed or applied *to*, this particular fixed something, which in turn functions as a source of unity for my otherwise ever-changing stream of consciousness. But just as a river cannot be individuated without its geophysical route (its underlying geophysical substratum or defining parameters—that which geophysically persists in relation to it), so too the individuation of my stream of consciousness requires a psychological route

(its underlying psychological substratum or defining parameters—that which psychologically persists in relation to it). And also by analogy we can predict, as in the case of the river’s geophysical route, that small variations within my individuating psychological determining substratum will also be permissible, so long as they always remain within certain fixed parameters.

40 (3) “But this persisting element cannot be an intuition in me. For all the determining grounds of my existence that can be encountered in me are representations, and as such they themselves need something persisting distinct from them, in relation to which their change, and thus my existence in the time in which they change, can be determined” (*CPR* Bxxxix n.).

41 *Commentary on step 3.* This is the first of the two most crucial steps of the proof. From step 2 we know that every changing conscious individuating determination of myself, as an individual empirical self in time presupposes something that persists, in relation to which I can uniquely determine the conscious stream of contents in my inner sense. But this persistent thing must be outside my own conscious mental states, and not merely inside me. For if it were merely inside me, it would then *belong* to the ever-changing stream of consciousness, and so could not provide a uniquely determining substratum for the mental modifications I experience directly. Hence it must fall outside the proper domain of my inner sense, that is, outside the series of first-order phenomenally conscious representational states that I am directly aware of via my second-order introspective consciousness or empirical apperception.

42 Now at this point it might well occur to us that something else in inner sense might plausibly function as the “persisting element,” namely the *form* of inner sense, as opposed to its contents. And indeed according to Kant the form of inner sense always remains the same, since it is invariantly presupposed by any actual or possible inner experience (*CPR* A22-23/B37, A31/B46). But the form of inner sense is nothing other than the representation of *time*. And it is incoherent to suggest that either the representation of time or time itself could be a persisting or enduring thing *in* time. Either the representation of time or time itself is a necessary formal precondition for the series of changes in my stream of consciousness. Now, to hold that the representation of time occurs in time, would be to confuse properties of the psychological *vehicle* of a representation (which does indeed occur in inner sense, hence in time) with semantic properties of its representational *content*. And also it would plainly be conceptually incoherent to hold that time itself occurs in time. So neither the representation of time nor time itself could also function as an enduring substance or substratum *to which* my changing conscious representational states are ascribed or applied. Hence nothing in either the content or the form of inner sense can function as the persistent element or substratum that is required for the individuation or unique determination of my stream of consciousness.

43 (4) “Thus the perception of this persistent thing is possible only through a **thing** outside me and not through the mere **representation** of a thing outside me. Consequently the determination of my existence in time is possible only by means of the existence of actual things that I perceive outside myself” (*CPR* B275-276, boldfacing in the original).

44 *Commentary on step 4.* This is the second crucial step of the RI. In order uniquely to determine or individuate my own successive existence in time I must presuppose the existence of outer things perceptually represented by me, and not merely the existence of my internally flowing conscious representations of those outer things. The radical nature of what Kant is saying here cannot be overemphasized. He is saying that any individuating temporally determinate introspective awareness of myself as an individual empirical self is necessarily also a direct non-conceptual veridical representation of some real material thing existing *outside* my stream of conscious experiences and *at a distance from me in space*. The latter factor is especially to be noted. In the Transcendental Aesthetic Kant argues that “in order for certain sensations be referred to something outside of me” they must be referred to “something in another place in space from that in which I find myself” (*CPR* A23/B38). My unique individuality at the level of first-order phenomenal consciousness is therefore *inherited* from the world of distal physical objects. In this way, despite the fact that via empirical apperception in a loose and everyday sense we introspect “an object of inner sense [which is] called ‘soul,’” there is strictly speaking for Kant no independent “inner object” of inner sense:

45 inner sense, by means of which the mind intuits itself, or its inner state, gives, to be sure, no intuition of the soul itself, as an object (*CPR* A22/B37).

46 That is, what is truly inner for Kant cannot be reified: it is neither a *noumenal* inner thing nor a *phenomenal* inner thing. And as he puts it in the first Note concerning the Refutation, “inner experience itself is ... only mediate and possible only through outer experience” (*CPR* B277). So, insofar as I am aware of myself in empirical apperception as a uniquely determined psychological being, an individual empirical self, then I must directly and non-conceptually ascribe or apply the changing contents of my mental states to the objective furniture of the distal material world.

47 This crucial point needs further emphasis. Far from having the problem of escaping from a “Cartesian box” into the outer world, *Kant’s* problem in the first *Critique* is instead that of distinguishing himself from various surrounding material objects in the outer world! This problem comes out clearly if we put it in non-Kantian terminology, this time borrowed from G.E. Moore and Jean-Paul Sartre. Kant’s view of inner sense in the Refutation comes very close to an amazing doctrine defended by Moore in his 1903 essay, also called “The Refutation of Idealism,” a doctrine which he calls the “transparency of consciousness”:

48 [W]hen we refer to introspection and try to discover what the sensation of blue is, it is very easy to suppose that we have before us only a single term. The term ‘blue’ is easy enough to distinguish, but that other element which I have called ‘consciousness’... is extremely difficult to fix. That many people fail to distinguish it at all is sufficiently shown by the fact that there are materialists. And, in general, that which makes the sensation of blue a mental fact seems to escape us: it seems, if I may use a metaphor; to be transparent --we look through it and see nothing but the blue. We may be convinced that there is something but what it is no philosopher, I think, has yet clearly recognized (Moore, 1993: p. 37).

49 Here consciousness is *not an inner thing*, in spades: instead it is nothing but a noetic searchlight on outer things. Later, in the 1930s (but presumably without having read Moore), Sartre pushes this idea of transparency one step further and describes something he calls “the transcendence of the ego” (Sartre, 1987). Sartre’s idea is that the ego receives its first-order unreflective reflexive subjective unity solely and directly from the outer things it is transparently conscious of. So this is not merely *content-externalism*: it is also *phenomenal consciousness-externalism*. The conscious mind is

much “out there in the world” as it is “in here.” *Egocentricity* is representational *eccentricity*. Phenomenal consciousness is nothing but *consciousness-of* or intentionality. Essentially the same view is held by contemporary defenders of the “first-order representational theory of consciousness” (see, e.g., Carruthers, 1998) and, even more radically, by defenders of the “extended conscious mind thesis” (see, e.g., Rockwell, 2005). All of these later affinities shows how radical and philosophically prescient Kant’s doctrine really is. Add the Sartrean transcendence of the ego, the first-order representational theory of consciousness, and the extended conscious mind thesis to the Moorean transparency of consciousness, and you have, in effect, Kant’s doctrine in step 4 of the RI.

50 (5) “Now consciousness [of my existence] in time is necessarily bound up with consciousness of the [condition of the] possibility of this time-determination. Therefore it is also necessarily combined with the existence of the things outside me, as the condition of time-determination” (*CPR* B276).

51 *Commentary on step 5.* This step is fairly straightforward. Insofar as I am empirically self-aware, and uniquely individuate myself in time, as an individual empirical self, I must also be directly consciously aware of this act of time-determination. Hence I must also be directly consciously aware of the existence of a distal persistent thing outside me that individuates me, since this is the necessary condition of time-determination.

52 (6) “I.e., the consciousness of my existence is at the same time (*zugleich*) an immediate consciousness of the existence of other things outside me” (*CPR* B276).

53 *Commentary on step 6.* This adds a crucial factor to step 5. The “immediate consciousness of” something is a direct veridical consciousness of that thing. So Kant is saying that for any particular empirical apperception of myself as uniquely determined in inner sense, I am also *simultaneously* directly veridically perceptually aware, via outer intuition, of some existing or actual distal material object in space as the individuating substratum to which I ascribe or apply the changing conscious representational contents of my mind.

54 So, to sum up the entire RI: Necessarily, if I’m determinately aware of myself in empirical apperception, then I’m also thereby at that very same moment directly veridically perceptually aware of some actual distal material object in space.

55 I now move on to some critical objections.

56 It seems to me that both steps 1 and 3 are acceptable, assuming the correctness of both Kant’s philosophical psychology (of inner sense, outer sense, and apperception) and of the “weak” reading of the First Analogy that I proposed. Yet critics of the RI very often hold that the fundamental gap in the proof is to be found in step 3. Why, such critics ask, is it necessarily the case that the intuition of that which is persistent, be an intuition of something outside me? Why couldn’t it instead be an intuition of some persistent thing *inside* me--that is, of some “thinking thing”? This option immediately fails, however, when we remember just what sort of intuition an inner intuition is:

57 the determination of my existence can occur only in conformity with the form of inner sense, according to the particular way in which the manifold that I combine is given in inner intuition, and therefore I have **no**

**cognition** of myself **as I am** but merely as I **appear** to myself (*CPR* B157-158, boldfacing in the original).

58 In other words, in empirical self-consciousness I'm *not* directly aware of myself as a Cartesian ego-in-itself. That would require an "intellectual intuition" (*CPR* B72) of myself as an individual empirical self. But as a finite and "human, all-too-human" human cognizer, an individual empirical self, who is not merely rational but also an animal, my intuition is strictly sensory and not intellectual: in inner sense, I'm directly aware only of the phenomenal flotsam and jetsam of consciousness. That is, in inner sense, I am directly aware only of *my phenomenally conscious states and their phenomenal contents (whether objectively representational or not)*, not of some deeper noumenal substratum of those phenomenally conscious states.

59 But even if steps 1 to 3 hold up tolerably well, nevertheless, in my opinion, steps 4 to 6 as they stand are highly questionable.

60 Here's a worry about step 4. Even granting that my empirical self-consciousness of my stream of consciousness in inner sense requires an outer intuition of something persistent, nevertheless it does not seem to follow that inner intuition in general requires any outer intuition of actually existing distal material objects in space. For so long as *space alone, as an object*, can be represented by means of a "pure intuition" or "formal intuition," as Kant explicitly argues in the Transcendental Aesthetic and again later in the B edition's Transcendental Deduction of the Pure Concepts of the Understanding (*CPR* B160 n.), then *that* seems sufficient to meet the requirement that there be a single persistent thing over against me, to which I must intuitionally refer and ascribe my ever-changing conscious inner states. And the pure or formal intuition of space does not logically require the existence of any distal material objects *in* space. Kant says explicitly that "one can never represent that there is no space, although one can very well think that there are no objects to be encountered in it" (*CPR* A24/B38-39).

61 What is the pure or formal intuition of space? Five features are at least individually necessary for it. **First**, the pure or formal intuition of space is a *non-empirical presupposition of all empirical intuitions of objects in space*: "[this representation of space] is a necessary representation, a priori, which is the ground of all outer intuitions" (*CPR* A24/B39).<sup>2</sup> **Second**, the pure or formal intuition of space is *non-conceptual*: "[this representation of] space is not a discursive or, as is said, general concept of relations of things in general, but a pure intuition" (*CPR* A25/B39). **Third**, the pure or formal intuition of space represents space as a *unique object*: "one can represent only a single space" (*CPR* A25/B39). **Fourth**, the pure or formal intuition of space represents space as a *unified structured manifold*:

62 Space and time are represented a priori not merely as **forms** of sensible intuition, but as **intuitions** themselves (which contain a manifold), and thus with the determination of the unity of this manifold... Space, represented as **object** (as is really required in geometry), contains more than the mere form of intuition, namely the **putting-together** (Zusammenfassung) of the manifold (*CPR* B160, text and note combined, boldfacing in the original).

63 **Fifth** and finally, the pure or formal intuition of space represents space as an *infinite totality*: "space is represented as a given infinite magnitude" (*CPR* A25/B40). For our purposes, we need not unpack Kant's extremely interesting doctrine of pure spatial representation any further. My twofold point right now is only (i) that the pure or formal intuition of space is an a necessary priori non-conceptual representation having a

referent--i.e., space itself--which is represented as a unique unified structured manifold and an infinite totality, and (ii) that this unique unified structured manifold and infinite totality has not been ruled out as the persisting element Kant needs in order to meet the requirement of step 2. It's incoherent to suppose that time itself might function as a persisting entity in time, but *not* incoherent to suppose that space itself might function as a persisting entity in time. And if space itself can meet that requirement, then since Kant explicitly says that space can be represented as empty of all material objects (*CPR* A24/B38-39, A291/B347), it follows that Kant has not ruled out the possibility that I ascribe or apply my changing mental states directly to empty space itself.

64 Just in case my objection to step 4 is not convincing however, here's another objection, this time to step 6. Even if we grant what I think we should *not* grant—namely, that my self-conscious awareness of my stream of consciousness in inner sense entails that I have some direct outer experiences of actual distal material objects in space—it doesn't seem to follow from that, that on every occasion of self-awareness I must be *simultaneously* directly correctly perceptually aware of a distal material external object. What about dreams and hallucinations? In Note 3 of the Refutation, Kant himself admits that

65 from the fact that the existence of outer objects is required for the possibility of a determinate consciousness of our self it does not follow that every intuitive mental representation of outer things includes at the same time (*zugleich*) their existence, for that may very well be the mere effect of the imagination (in dreams as well as in delusions) (*CPR* B278, italics added).

66 So Kant is certainly aware of the dream problem, and he must then implicitly grant that step 6 as it stands, with the simultaneity condition, is false.

67 Where does this leave us? By virtue of his admission of the dream problem, Kant has implicitly admitted that not every self-conscious awareness of my own uniquely determined conscious existence in time entails a simultaneous direct correct perception of a distal external object. So since Kant is certainly no fool, it seems to me that his concluding step 6 is most charitably and plausibly interpreted as saying the same as these two alternative formulations of the conclusion of the RI:

68 The proof that is demanded must therefore establish that we have **experience** and not merely **imagination** of outer things, which cannot be accomplished unless one can prove that even our **inner experience**, undoubted by Descartes, is possible only under the presupposition of outer experience (*CPR* B275, boldfacing added).

69 The mere, but empirically determined, consciousness of my own existence proves the existence of objects in space outside me (*CPR* B275).

70 Taken together, these formulations say that my having a self-conscious awareness of my individuated stream of inner consciousness, as an empirical self, entails my also having some direct correct perceptions of distal material objects in space. Even so, there is no necessity that I have a direct correct outer perception of a distal material object at the very *same* time that I am in one of these self-conscious states, so long as I also have some direct correct outer perceptions of distal spatial objects at *other* times. The simultaneity condition can be dropped.

71 This charitable interpretation is backed up by a footnote appended to the RI, which says that even when we are dreaming or hallucinating, and merely *imagining* space, it is presupposed that we already have an outer sense through which we do sometimes get direct correct perceptual access to outer material things:

72 In order for us even to imagine something as external, i.e., to exhibit it to sense in intuition, we must already have an outer sense, and by this means immediately distinguish the mere receptivity of an outer intuition from the spontaneity that characterizes every imagining. For even merely to imagine an outer sense would annihilate the faculty of intuition, which is to be determined through the power of imagination (*CPR* B276-277 n.).

73 In other words, space cannot be even imagined without our already having a capacity, sometimes actualized or realized, for directly and correctly perceiving or empirically intuiting distal material objects in space. And this reading is in turn backed up by two other texts. **First**, in the *Anthropology* Kant notes that imagination “cannot bring forth a representation that was *never* given to the power of sense; we can always trace the material of its representations” (*A* 7: 168). And **second**, in one of the *Reflexionen* he is even more explicit:

74 Dreams can represent to us things as outer, which are not there; however, we would never be able to represent something as outer in dreams, if these forms were not given to us through outer things (*R* 5399; 18:172).

75 So it seems to be Kant’s view that even our capacity for “imagination of something as external” is parasitic upon some direct correct outer sense perceptions of distal material objects, at some time or another. If he is right, and if we interpret step 6 in such a way as not to commit Kant to the implausible thesis that every individuating act of empirical self-consciousness requires a simultaneous direct correct perceptual awareness of a distal material object, then he in fact avoids the dream problem.

76 But even so, *is* he right? Well it seems likely that it is generally empirically true, as a fact in cognitive psychology, that normal image-construction and manipulation is originally funded by direct correct sense-perception of distal material objects. But is it *necessarily true* for creatures like us? Surely we can conceive of a possible human being whose empirical imagination-content is entirely funded by some source other than direct correct sense-perception of distal material objects. Or, to put it another way, if a creature had been born with or developed a capacity for imagining external things that was entirely empirically funded in some non-standard way which was systematically insulated from direct correct perceptual contact with the distal outer world—suppose, for example, that someone, unfortunately (see sections IV and V below) was existentially “thrown” into this world as the child of fanatical transhumanist parents, and was fitted by them from birth with a microscopically thin computer-driven “virtual reality suit” covering her entire body, or that (as in *The Matrix*) she was born hooked up to the Matrix, so that again all her perceptions were in fact false digital images—would she thereby fail to be *one of us*? I think not. Such a human cognizer, cocooned inside her all-encompassing perceptual prosthetic, and/or non-self-consciously supplied with a massively complex and detailed but still phoney digital image of her actual surrounding world, would certainly be odd, and perhaps somehow slightly cognitively handicapped (or perhaps even seriously handicapped, in light of the actual empirical fact of “neural plasticity,” as manifest in the effective neural and behavioral adaptation of actual human cognizers to inverting lense goggles, Tactile-Visual Substitution Systems, etc.): but she would certainly nevertheless, I think, still fully share our basic human cognitive constitution. So, by virtue of this conceivably naturally or nomologically possible transhumanist thought-experiment, Kant’s thesis of the dependency of imagination on correct perception is false, and the RI is therefore unsound.

77 I’m now going to shift philosophical gears, and move from the negative evaluation of the RI given at the end of the last paragraph, toward a positive evaluation

that yields—in my opinion—a sound argument. Moreover, in this regard, I think that the RI implicitly contains something of real and indeed fundamental philosophical significance. Let's suppose that steps 4 and 6 are indeed fallacious as they stand. Nevertheless it seems to me the case that Kant *has* indeed proved *this* weaker thesis:

78                    [[Inner experience in general is possible only through outer experience in general (*CPR* B278-279).

79            My reasoning is this. Crucial to this thesis are two phrases: “inner experience in general” and “outer experience in general.” I want to read “inner experience in general” as meaning “to have a self-conscious awareness of myself in inner sense,” and I want to read “outer experience in general” as meaning “to have an actual outer sense.” That is, I want to read the thesis as saying:

80                    To have a self-conscious awareness of myself in inner sense is possible only through my also having an actual outer sense,

81            *not*:

82                    Each and every inner self-conscious experience of a given mental state of my own is possible only through some direct correct outer perception of an actual distal material object in space.

83            That my proposed reading is at least plausibly Kantian is also well-supported by a passage in the *Reflexionen*:

84                    The question, whether something is outside of me, is just the same as to ask, whether I represent to myself an actual space. For this is outside of me (*R* 5400; 18: 172).

85            Otherwise put, I want to distinguish quite sharply between three distinct meanings of the phrase “X is outside my own conscious mental state”:

86            (i) X is a mind-independent substance,

87            (ii) X is a material object in another part of space from that in which I am located

= X is a distal material object, and

88            (iii) X is necessarily spatial in character.

89            Granting that, then what I want to argue on Kant's behalf is that, in order to refute skeptical idealism, it's necessary only to prove that *I myself* satisfy (iii), not to prove that *something else* satisfies (i) or (ii). The issue on the table right now is whether a dreamer or hallucinator could have a capacity for imagining external things *without having an actual outer sense*. Again, I think not. That is, I would want to argue on Kant's behalf that a capacity for imagining external things, even in dreams or hallucinations, is not possible without an actual outer sensibility.

90            Here's how Kant defines outer sense:

91                    By means of outer sense (a property of our mind), we represent to ourselves objects outside us, and all as in space.... In order for certain sensations to be referred to something outside me (i.e., to something in another place in space from that in which I find myself), thus in order for me to represent them as outside one another, thus not merely as different but as in different places, the representation of space must already be their ground (*CPR* A22-23/B37-38, underlining added).

92            In order to interpret this definition, I will rely heavily heavily on a short essay published by Kant in 1768—significantly, only two years after the publication of *Dreams of a Spirit-Seer*—“Concerning the Ultimate Ground of the Differentiation of Directions in Space.” Despite its brevity, this essay is seminal for the Critical Philosophy as a whole because it effectively prepares the ground for the theory of space,

time, and sensibility that Kant first worked out in his Inaugural Dissertation in 1769, and held at the time of his famous letter to Marcus Herz in 1772. Indeed, as Kant remarks in one of the best-known *Reflexionen*: “the year ‘69 gave me great light” (*R* 5037; 18: 69). For these and other closely-related reasons, I’ve dubbed the period from 1768 to 1772 Kant’s *proto-Critical* period (Hanna, 2016). In any case, for my purposes here, the most important passage in “Directions in Space” is this one:

93           Because of its three dimensions, physical space can be thought of as having three planes, which all intersect each other at right angles. Concerning the things which exist outside ourselves: it is only in so far as they stand in relations to ourselves that we have cognition of them by means of the senses at all. It is, therefore, not surprising that the ultimate ground on the basis of which we form our concept of directions in space, derives from the relation of these intersecting planes to our bodies ... Even our judgments relating to the cardinal points of the compass are, in so far as they are determined to the sides of our body, subject to the concept we have of directions in general. Independently of this fundamental concept, all we know of relations in heaven or on earth is simply the positions of objects relative to each other. No matter how well I may know the order of the compass points, I can only determine directions by reference to them if I know whether this order runs from right to left, or from left to right .... The same thing holds of geographical and, indeed, of our most ordinary knowledge of the position of places. Such knowledge would be of no use to us unless we could also orientate the things thus ordered, along with the entire system of their reciprocal positions, by referring them to the sides of our body (*DS* 2: 378–9).

94           As I read this text, Kant is arguing as follows.

95           **First**, every space represented by creatures like us in sensory intuition contains *directions*, namely, special topological features that partially determine not only the qualitative or extrinsic non-relational properties of material objects and their intrinsic structural properties but also the extrinsic relative positioning of material objects, yet which cannot be determined merely by any intrinsic nonrelational properties that material objects might possess. Indeed, as Kant argues two pages later than the above text, the fact of “incongruent counterparts” or enantiomorphic objects—for example, 3-D physical objects such as my right and left hands, which (ideally speaking) are isomorphic mirror images of one another and yet cannot be made to coincide by any continuous translation of the figures in 3-D space—shows the falsity of the relational (or Leibnizian) theory of space, which says that spatial relations are logically strongly supervenient on the intrinsic non-relational properties of noumenal objects or monads, whether these are taken to be mental monads (as Leibniz did) or physical monads (as Kant himself did in the pre-Critical *Physical Monadology*): for if this were true, then the right and left hands would be congruent.

96           **Second**, directions in space are unintelligible unless there exists a fixed frame of reference for spatially orienting distal objects.

97           **Third**, any fixed frame of reference for distal spatial orientation is necessarily centered on an egocentric origin-point or “egocentric space” consisting of the 3-D rectilinear axes of the cognizing subject’s own *body*: up/down, right/left, in front/behind.

98           **Fourth**, therefore, necessarily for all creatures with cognitive constitutions like mine, if any one of those creatures represents space, then its own body must also exist in that space.

99           Now, can my own body, construed as an egocentric, indexically-fixed material reference-frame for any possible representation of an intrinsically directionally-structured space, function as the persisting substratum in the original version of the RI? If so, then it seems to me that the RI can be charitably reconstructed as really arguing that necessarily for every creature cognitively constituted like me, a self-conscious awareness of its own uniquely determined stream of consciousness in inner sense

requires the existence of its own body in space. Here's a more explicit version of that argument:

100 (1) I represent myself through empirical self-consciousness as an individuated stream of consciousness in inner sense, an individual empirical self.

101 (2) In order to represent myself in empirical self-consciousness as an individuated stream of consciousness in inner sense, an individual empirical self, I must represent myself, at the very least, in a direct relation to actual space itself, which in turn functions as a self-identical substratum for temporally determining my own stream of consciousness.

102 (The rationale for this crucial step is given by Kant's argument for steps 1 to 3 in the original version of the RI, plus my criticism of steps 4 and 6).

103 (3) In order to represent actual space itself in some empirical context, I must represent it as having directions.

104 (4) In order to represent directions in space, I must interpret those directions by means of a 3-D egocentric frame of reference centered on my own body.

105 (5) Therefore, necessarily, if I represent myself in empirical self-consciousness as an individuated stream of consciousness in inner sense, an individual empirical self, then my body must also exist in space.

106 By using *modus ponens* on (1) and (5), and then generalizing the result, we easily obtain the denial of Cartesian skeptical idealism and Berkelyan metaphysical idealism alike:

107 Necessarily, for every creature having a cognitive constitution like mine, something exists outside its own conscious mental states, i.e., its own body in space.

108 Reconstructed in this way, the RI conforms very smoothly to Kant's leading idea that all conscious changes in inner sense are necessarily immediately ascribed to an actual spatially existing self-identical persisting thing or substratum. For we can now see that the most natural way of reading this is as saying that necessarily the contents of my own consciousness literally belong to my own living organismic human animal body or *Leib*. More specifically, the big problem with steps 4 and 6 in the original version of the RI was the assumption that the external substratum in question was *distal*, not *proximal*, in relation to the self-conscious subject. But suppose instead that the external substratum Kant is talking about is strictly proximal: suppose that the external substratum is none other than my own living organismic human animal body in actual space. Then Kant is saying (i) that my conscious mind is necessarily and completely an embodied mind, or what I've been calling *an essentially embodied mind* (Hanna and Maiese, 2009; Hanna, 2011), and (ii) that in order to individuate myself psychologically, as an individual empirical self, and as a unique member of my own species, then I must ascribe each of my mental states directly to my own living organismic human animal body in actual space. Or, in other words, the ascription of my mental states to my own living organismic human body individuates my mental states, and constitutes me as an individual empirical self, by locating my mental states uniquely.

109 In the important 1786 essay, “What is Orientation in Thinking?,” Kant calls this capacity of a rational human animal for bodily self-location “self-orientation”:

110 To self-orient (sich orientiren) in the proper sense of the word, means to use a given direction—and when we divide the horizon into four of these—in order to find the others, and in particular that of sunrise. If I see the sun in the sky and know that it is now midday, I know how to find south, west, north, and east. For this purpose, however, I must necessarily be able to feel a difference within my own subject, namely that between my right and left hands. I call this a feeling, because these two sides display no perceptible difference as far as external intuition is concerned ... I can now extend this geographical concept of the process of orientation to signify any kind of orientation within a given space, i.e., orientation in a purely mathematical sense. In the darkness, I can orientate myself in a familiar room so long as I can touch any one object whose position I remember. But it is obvious that the only thing which assists me here is an ability to define the positions of the objects by means of a subjective distinction: for I cannot see the objects whose position I am supposed to find; and if, for a joke, someone had shifted all the objects round in such a way that their relative positions remained the same but what was previously on the right was now on the left, I would be quite unable to find my way about a room whose walls were in other respects identical. But in fact, I can soon self-orient simply by the feeling of difference between my two sides, my right and my left (*OT* 8: 134–135).

111 There are obvious parallels between this text and the passage from the “Directions in Space” essay I quoted earlier. But the most important thing is this: when we take the idea of self-orientation from “What is Orientation in Thinking?,” and add it to the charitably interpreted RI and to “Directions in Space,” then we can derive a profound Kantian doctrine to the effect that to be self-consciously aware of my own uniquely determined psychological life as an individual empirical self is automatically also to be intuitionally aware of the unique location—and also, I think, of the unique locus of movement, or motility (see, e.g., Hanna and Maiese, 2009: chs. 3-8)—of my own living organismic human animal body in space. All human empirical apperception is thus “orienting apperception”: necessarily, I become aware of myself as *myself* only by way of establishing my own living human body, which is the unique location and kinesthetic locus of all my mental states, as a 3-D egocentric frame of reference in a directionally-structured encompassing total singular space. Or to use the entirely apt and now canonical language introduced by Nagel: necessarily, when I introspectively find myself as a subject enjoying phenomenal consciousness, which is “what it is like to be, for an organism,” I also find myself essentially having a “single point of view” (Nagel, 1979: pp. 160, 167).

112 Therefore, given the interpretation of outer sense that I’ve just worked out, as what stands in an immediate or a mediate relationship to my self-locating living organismic human animal body, this in turn is equivalent to the thesis that a subject’s capacity for imagining external things is not possible without *her possessing a living organismic human animal body in space*. Reconstructed in this way, the RI conforms very smoothly to Kant’s original idea that all conscious changes in inner sense are necessarily immediately ascribed to an actual spatially existing persisting thing or substratum. For we can now see that the most natural way of reading this is as saying that necessarily the contents of my own consciousness literally belong to my own living organismic human animal body. The big problem with steps 4 and 6 in the original argument was the assumption that the external substratum in question was *distal*, not *proximal*, in relation to the self-conscious subject. But suppose that the external substratum Kant is talking about is strictly proximal: suppose that the external substratum is none other than my own living organismic human animal body. Then what Kant is actually saying in the RI is that in order to individuate myself psychologically and as a unique member of my own species, then I must ascribe each of my mental states directly to my own living organismic human organismic animal body in space. In other words, the ascription of my mental states to my own living human animal body

individuates my mental states by locating them uniquely. And this seems to me to be true. So the RI seems to me to be a sound demonstration, against the Cartesian skeptical idealist and also the Berkelyan metaphysical idealist, of the existence of my own living organismic human animal body in space. That is, Kant is saying in the RI that necessarily, if I'm self-consciously aware of myself as an individuated stream of consciousness in inner sense, then my own living organismic human animal body also exists as uniquely self-located.

113 According to what I've called *Kant's anti-mechanism* (Hanna, 2014, 2022: ch. 1), there's a fundamental ontological and metaphysical difference between (i) *natural mechanisms*, that is, deterministic—or indeterministic for that matter, although Kant himself wouldn't have been in a good historical-theoretical position to recognize that, since the very idea of indeterminism and probabilistic/statistical/stochastic laws was a 19<sup>th</sup> century discovery or invention (Hacking, 1990)—mechanistic processes in nature, and (ii) *natural purposes*, that is, spontaneous, teleological, self-organizing, living organismic processes in nature, including mental processes, all of which are also self-organizing, living, organismic processes:

114 [L]ife is the subjective condition of all our possible experience (*Prolog* 4: 335).

115 [T]he mind is for itself entirely life (the principle of life itself) (*CPJ* 5: 278).

116 But according to Kant in the second half of the third *Critique*, although natural science can and actually does know natural mechanisms, it *cannot* know natural purposes, as this text asserts:

117 It is quite certain that we can never adequately come to know the organized beings and their internal possibility in accordance with merely mechanical principles of nature, let alone explain them; and this is so certain that we can boldly say that it would be absurd for humans to make an attempt or to hope that there could ever arise a Newton who could make comprehensible even the generation of a blade of grass according to natural laws that no intention has ordered; rather we must absolutely deny this insight to human beings (*CPJ* 5: 400).

118 Hence, according to Kant in the second half of the third *Critique*, although natural science *cannot know the real difference between natural mechanisms and natural purposes*, it must also investigate nature *as if there were a real difference between them*, as a regulative Idea for the purposes of a coherent and progressive natural-scientific investigation of nature.

119 Now, although this “regulative” (hypothetical-practical) conception of natural purposes is not “constitutive” (assertoric-theoretical), nevertheless it also directly entails the synthetic a priori subjunctive conditional truth that necessarily, *if* natural purposes *were to exist*, then universal natural mechanism *would be* false. But if we allow for the existence of veridical non-conceptual cognition via empirical or pure intuition (Hanna, 2005, 2008, 2015), then we can also advance from Kant's necessary subjunctive conditional thesis to a corresponding assertoric thesis that I've called *Kantian anti-mechanism*. According to Kantian anti-mechanism, although natural science cannot *know* either natural purposes or the difference between natural mechanisms and natural purposes, nevertheless we human cognizers can and actually do also have *veridical non-conceptual cognition of natural purposes*, by means of the “feeling of life” in our aesthetic experience of the beautiful and the sublime in nature (*CPJ* 5: 204). In this way, natural purposes actually exist in manifestly real nature, because *we* actually exist in manifestly real nature and because *we veridically feel our own life* and *we are living*

*organismic human animals*: therefore, not everything in veridically apparent or manifestly real nature is a natural mechanism (see also Gambarotto and Nahas, 2022).

<sup>120</sup> Since we actually exist in veridically apparent or manifestly real nature, as self-locating individual empirical selves whose rational, conscious, self-conscious, intentional, emotional, volitional minds are essentially embodied in our own living organismic human bodies, and since we are natural purposes, and since it is *not only* the case, (i) according to the third section of the *Groundwork of the Metaphysics of Morals*, that we must conceive ourselves under a regulative Idea of our own free agency and act as if we were transcendently and practically free, and also *not only* the case (ii) according to the third Postulate of Pure Practical Reason in the *Critique of Practical Reason*, that we must have moral faith (*Glaube*) in our freedom, *but also* the case, (iii) according to the “Fact of Reason” in the second *Critique*, that we have a direct non-conceptual awareness of our own freedom, then it follows directly, according to what I’ve called *Kant’s biological theory of freedom* (Hanna, 2009) and also his *embodied agency theory* of free will and practical agency (Hanna, 2006: ch. 8), that (iv) transcendental, practical, and autonomous freedom really and truly exist in the manifestly real world, as biological facts about our own lives, at the source of our self-determining, creative agency. More specifically, just as conscious, self-conscious, intentional, emotional, volitional mind is essentially a mechanistically/deterministically and physicalistically *irreducible* form of life, so too our self-determining, creative, deeply free *agentive sourcehood* is essentially a mechanistically/deterministically and physicalistically irreducible form of life. Or in other and fewer words, when we conjoin the conclusion of the charitably-interpreted RI, Kant’s biological theory of freedom, and Kant’s embodied agency theory of free will and practical agency, then it follows that EEKST is true.

<sup>121</sup>

### **Swedenborgianism, The Impossibility of Strong AI, and the Crypto-Swedenborgian Noumenal Fantasy of Transhuman Selves**

<sup>122</sup> The classical Swedenborgian noumenal fantasy about the self is that our “human, all-too-human” self can somehow exist in disembodied form, as a *Geist*/ghost-self that hovers epiphenomenally above the manifestly real world, immortally, yet also retains *all* the representational content that we ever personally or sub-personally cognized. The recent and contemporary crypto-Swedenborgian noumenal fantasy of transhumanism selves is that this *Geist*/ghost-self is a functional system in digital format, that can exist in abstract, disembodied form like a computer program, and be repeatedly uploaded or downloaded to, and implemented or realized on, on different kinds of digital technology that also have other causal powers of various kinds, and thereby our transhuman self epiphenomenally hovers above the manifestly real natural and social world like a causally inert shadow, going from from one uploading, downloading, implementation, or realization to another, escaping accidental or natural human death for an indefinitely long time and perhaps even forever in the sempiternal sense, i.e., immortally.

<sup>123</sup> Transhumanism, in turn presupposes the truth of the *strong thesis of artificial intelligence*, aka *strong AI*, which in turn presupposes the truth of the

materialist/physicalist doctrine of *metaphysical computational-functionalism about the human mind*, which says that the human mind and all its contents can be explanatorily and ontologically reduced to a computable system of input-output mappings that can be implemented or realized on different sorts of machines, all of which operate as digital computing machines, in addition to whatever other causal powers these machines might have. But if EEKST is true, then both strong AI and metaphysical computational-functionalism about the human mind are necessarily false.

<sup>124</sup> More specifically, *strong AI* is the two-part thesis which says (i) that rational human intelligence can be explanatorily and ontologically reduced to Turing-computable algorithms and the operations of digital computers (aka *the thesis of formal mechanism, as it's applied to rational human intelligence*), and (ii) that it's technologically possible to build a digital computer that's an exact counterpart of rational human intelligence, such that this machine not only exactly reproduces (aka simulates) all the actual performances of rational human intelligence, but also outperforms it (aka *the counterpart thesis*) (see, e.g., Block, 1980: part 3; Kim, 2011: ch. 6). The strong AI thesis is not only immensely controversial, but also strongly apt to be seriously muddled, for at least three reasons.

<sup>125</sup> **First**, the strong AI thesis is very often confused with the *weak AI* thesis (for details, see p. 29 below), but (i) the weak AI thesis is *itself* ambiguous as between a *non-trivial* version and a *trivial* version, and (ii) even if both of the versions of the weak AI thesis were true, nevertheless strong AI could still be false.

<sup>126</sup> **Second**, the strong AI thesis, as such, overlooks the fact that rational human intelligence is also *conscious*: hence if strong AI were true, then human consciousness would also have to be explanatorily and ontologically reducible to Turing-computable algorithms and the operations of digital computers, which is equivalent to metaphysical computational-functionalism about *human consciousness*, which in turn puts an extra-heavy burden of proof on defenders of strong AI. Indeed, Ned Block, David Chalmers, Frank Jackson, and others have developed (admittedly, controversial) arguments which, if sound, show that phenomenal consciousness does not logically strongly supervene on functional physical facts (see, e.g., Chalmers, 1996: chs. 1-4). But even if, as I've argued, the conception of phenomenal consciousness that's assumed by all *those* anti-functional arguments is false (Hanna and Maiese, 2009; Hanna, 2011), nevertheless if EEKST is true, then necessarily, consciousness in creatures like us is essentially embodied, so metaphysical computational-functionalism about human consciousness is false and indeed synthetic a priori impossible.

<sup>127</sup> **Third**, the very idea of "being artificial" is ambiguous as between (i) being mechanical, as opposed to *being organic*, and (ii) being able to be built or constructed or synthesized, as opposed to *not being able to be built or constructed or synthesized*, for whatever reason, but (i) and (ii) are mutually logically independent of one another: something could be mechanical but not buildable, constructible, or synthesizable (for example, digital computations involving more digits or computations than there are particles or future moments of time in the cosmos), and conversely something could be buildable, constructible, or synthesizable but not mechanical (for example, certain exactly reproducible uncomputable, non-equilibrium thermodynamic biochemical processes, including organismic processes).

128 Over and above the controversies, the strong AI thesis is demonstrably false, for at least four good reasons, and the weak AI thesis is either false (the non-trivial version) or boringly trivially true (the trivial version), as follows.

129 **1.** If EEKST is true, then necessarily, intelligent rational human minds are *alive*, but systems that conform to the strong AI thesis are *inherently mechanical and non-living*, so the strong AI thesis is necessarily false.

130 **2.** If EEKST is true, then necessarily, intelligent rational human minds are *essentially embodied*, but systems that conform to the strong AI thesis are *possibly disembodied*, so the strong AI thesis is necessarily false.

131 **3.** If EEKST is true, then necessarily, rational human knowledge requires a *non-accidental* connection between judgment or belief and truth, and also a *non-accidental* connection between true belief and justification, but systems that conform to the strong AI thesis only ever provide *accidental* content-connections, hence not only (i) is the strong AI thesis *necessarily false*, but also (ii) if the strong AI thesis *were* true, then our intelligent rational human minds would be nothing more than digital computers, and therefore we couldn't ever *know* the truth of the strong AI thesis, so the strong AI thesis is also *epistemically self-undermining*.

132 **4.** If EEKST is true and *also* Kurt Gödel's incompleteness theorems (Gödel, 1967), which say (i) that all *Principia Mathematica*-style systems of mathematical logic based on the Peano axioms for arithmetic will contain undecidable, unprovable sentences, and (ii) that no such system of mathematical logic can prove its own consistency, hence the truth of mathematical axioms has to be demonstrated *outside* those systems—for example, by acts of rational human *mathematical intuition* (see, e.g., Hanna, 2015: chs. 6-8)—are true, then EEKST together with Gödel's incompleteness theorems jointly entail that there will be uncomputable, undecidable, unprovable mathematical axioms *that only intelligent rational human minds, i.e., essentially embodied Kantian selves, can know*, so systems that conform to the strong AI thesis inherently fall short of the actual performances of rational human intelligence, and therefore the strong AI thesis is false.

133 Moreover, if the weak AI thesis says that *not all* but only *some* actual performances of rational human intelligence are *exactly reproducible* (aka can be simulated) on digital computers (i.e., the non-trivial version), then since the strong AI thesis is not only false but impossible, then the non-trivial version of the weak AI thesis is false and impossible too. But if the weak AI thesis says merely that *some behavioral or formal features* of *some* actual performances of rational human intelligence are *either operationally or isomorphically representable* on digital computers (the trivial version), then this is indeed true, but at best *boringly trivially true*, since the very same thesis is true of even the simplest counting or calculating procedures, using, for example, one's fingers, hockey pucks, or an abacus.

134 In short, the very idea of transhuman selves is nothing but a noumenal *Hirngespinnst*, a noumenal fantasy—a digital ghost—of precisely the sort so effectively criticized and wittily derided by Kant in his 1766 critique of Swedenborgianism, *Dreams of a Ghost-Seer*.

## Conclusion

136 As we've seen, EEKST is arguably true, and if so, then the very idea of transhuman selves is conceptually empty or incoherent and synthetic a priori impossible. Now, the only remaining philosophical question is, *why* are contemporary transhumanists so attracted to, and indeed intellectually and emotionally *addicted* to, the crypto-Swedenborgian fantasy that their own “human, all-too-human” selves are digital *Geist*/ghost-selves? By way of answering that question, I strongly agree with Arran Gare's critical diagnosis of the postmodernist/post-humanist “quest for disembodiment”:

137 The illusory quest for disembodiment ... appears to be common among "macroparasites" who live off the work, products and lives of others. Historically, this illusion of disembodiment appears to have legitimated exploitation of others, but in doing so has led the ruling classes of civilizations to destroy the real conditions of their own existence. The ... postmodernist forms of this illusion ... are transmogrifications of the illusion of disembodiment on which medieval civilization was based (Gare, 2016: p. 27).

138 So, assuming that Gare's critical diagnosis is correct, then not only are the transhumanist crypto-Swedenborgian noumenal fantasists of transhuman selves—digital *Geist*/ghost-selves—self-deniers of their own living organismic human animal embodiment (i.e., they're committed to an inherently inauthentic way of life by virtue of existential self-stultification), they're also explicitly or implicitly intellectually and emotionally addicted to a rationally unjustified and immoral coercive authoritarian sociopolitical theory (i.e., they're committed to a social-institutional system that violates the universal obligation to treat all rational human animals with sufficient respect for their human human dignity, and is thereby inherently oppressive). The very idea of transhuman selves is therefore not only a crypto-Swedenborgian noumenal fantasy, it's also an existentially and morally *reprehensible* fantasy. Now, it's an obvious and tragic truth that people can become addicted to things that are not only bad for them personally, but also objectively bad. And so it is with transhumanism.

139

## References

140 For convenience, I cite Kant's works in parentheses. The citations include both an abbreviation of the English title and the corresponding volume and page numbers in the standard “Akademie” edition of Kant's works: *Kants gesammelte Schriften*, edited by the Königlich Preussischen (now Deutschen) Akademie der Wissenschaften (Berlin: G. Reimer [now de Gruyter], 1902—). I generally follow the standard English translations of Kant's works, but have occasionally modified them where appropriate. For references to the first *Critique*, I follow the common practice of giving page numbers from the A (1781) and B (1787) German editions only. Because the Akademie edition contains only the B edition of the first *Critique*, I have also consulted the following German composite edition: *Kritik der reinen Vernunft*, ed. W. Weischedel, Immanuel Kant Werkausgabe III (Frankfurt: Suhrkamp, 1968). For references to Kant's *Reflexionen*, i.e., entries in *Kants handschriftlicher Nachlass*—which I abbreviate as “R”—I give the entry number in addition to the Akademie volume and page numbers. The translations from the *Reflexionen* are my own. Here is a list of the relevant abbreviations and English translations, along with their dates of first publication as well as their Akademie volume numbers and page ranges (except for the first *Critique*):

141 *A Anthropology from a Pragmatic Point of View*. Trans. M. Gregor. In I. Kant, *Immanuel Kant: Anthropology, History, and Education*. Cambridge: Cambridge Univ. Press, 2007. Pp. 227–429. [1798, Ak 7: 117–333]

*CPJ Critique of the Power of Judgment*. Trans. P. Guyer and E. Matthews. Cambridge: Cambridge Univ. Press, 2000. [1790, Ak 5: 165–485]

*CPR Critique of Pure Reason*. Trans. P. Guyer and A. Wood. Cambridge: Cambridge Univ. Press, 1997. [1781, 1787]

*DDS* “Concerning the Ultimate Ground of the Differentiation of Directions in Space.” Trans. D. Walford and R. Meerbote. In I. Kant, *Immanuel Kant: Theoretical Philosophy: 1755–1770*. Cambridge: Cambridge Univ. Press, 1992. Pp. 365–372. [1768, Ak 2: 375–383]

*DSS* “Dreams of a Spirit-Seer Elucidated by Dreams of Metaphysics.” Trans. D. Walford and R. Meerbote. In Kant, *Immanuel Kant: Theoretical Philosophy: 1755–1770*. Pp. 301–359. [1766, Ak 2: 315–373]

*OT* “What Does It Mean to Orient Oneself in Thinking?” Trans. A. Wood. In I. Kant, *Immanuel Kant: Religion and Rational Theology*. Cambridge: Cambridge Univ. Press, 1996. Pp. 7–18. [1786, Ak 8: 133–46]

*Prol Prolegomena to Any Future Metaphysics*. Trans. G. Hatfield. Cambridge: Cambridge Univ. Press, 2004. [1783, Ak 4: 253–383]

---

#### Примечания:

1. For a full presentation and detailed defense of the analytic-synthetic distinction, including the synthetic a priori, see (Hanna, 2001, and 2015: esp. ch. 4).

2. There is, however, an interpretive subtlety here: the pure or formal intuition of space is *a* presupposition of all empirical intuitions of objects in space, but it is not *the* presupposition: only the bare intuitional representation of space, the form of our outer intuition, is uniquely presupposed. The pure or formal intuition of space entails the form of outer intuition, but the form of our outer intuition does not entail the pure or formal intuition of space. For the important distinction between *formal intuitions* and *forms of intuition*, see (*CPR* 160).

---

#### Библиография:

1. Block, N. (ed.) (1980) *Readings in the Philosophy of Psychology*. 2 vols., Cambridge, MA: Harvard Univ. Press. Vol. 1.

2. Britannica (2022) “Transhumanism.” .

3. Carruthers, P. (1998) Natural Theories of Consciousness. *European Journal of Philosophy* 6: 203–220.

4. Chalmers, D. (1996) *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford Univ. Press.

5. Gambarotto, A. and Nahas, A. (2022) Teleology and the Organism: On Kant’s Controversial Legacy for Contemporary Biology. *Studies in History and Philosophy of Science* 93: 47–56. .

6. Gare, A. (2016) “Beyond Modernism and Postmodernism: The Narrative of Re-embodiments,” in R. Thomas-Pellicer, V. De Lucia and S. Sullivan (eds.), *Contributions*

- to Law, Philosophy and Ecology: Exploring Re-Embodiments. London: Routledge. Ch. 1. Pp.27–46. Also available online .
7. Gödel, K. (1967) “On Formally Undecidable Propositions of Principia Mathematica and Related Systems,” in J. van Heijenoort (ed.), *From Frege to Gödel*. Cambridge MA: Harvard Univ. Press. Pp. 596–617.
  8. O’Connell, M. (2018) Five Books to Understand Transhumanism. *The Guardian*. 10 May. .
  9. Hacking, I. (1990) *The Taming of Chance*. Cambridge: Cambridge University Press.
  10. Hanna, R. (2001) *Kant and the Foundations of Analytic Philosophy*. Oxford: Clarendon/Oxford Univ. Press.
  11. Hanna, R. (2000) Kant and Nonconceptual Content. *European Journal of Philosophy* 13: 247–290.
  12. (Hanna, 2006). Hanna, R. *Kant, Science, and Human Nature*. Oxford: Clarendon/Oxford Univ. Press.
  13. Hanna, R. (2008) Kantian Non-Conceptualism. *Philosophical Studies* 137: 41–64.
  14. Hanna, R. (2009) Freedom, Teleology, and Rational Causation. *Kant Yearbook* 1: 99–142.
  15. Hanna, R. (2011) Minding the Body. *Philosophical Topics* 39: 15–40.
  16. Hanna, R. (2014) Kant’s Anti-Mechanism and Kantian Anti-Mechanism. *Studies in History and Philosophy of Biological and Biomedical Science* 45. .
  17. Hanna, R. (2015) *Cognition, Content, and the A Priori: A Study in the Philosophy of Mind and Knowledge*. *The Rational Human Condition*, Vol. 5. Oxford: Oxford Univ. Press.
  18. Hanna, R. (2016) “Directions in Space, Non-Conceptual Form, and the Foundations of Transcendental Idealism,” in D. Schulting (ed.), *Kantian Nonconceptualism*. London: Palgrave Macmillan. Pp. 99–115.
  19. Hanna, R. (2022) *The Philosophy of the Future: Uniscience and the Modern World*. Unpublished MS. .
  20. Hanna, R. and Maiese, M. (2009) *Embodied Minds in Action*. Oxford: Oxford Univ. Press.
  21. James, W. (1950) *Principles of Psychology*. 2 vols., New York: Dover.
  22. Kim, J. (2011) *Philosophy of Mind*. 3rd edn., Boulder, CO: Westview.

23. Moore, G.E. (1993) “The Refutation of Idealism,” in G.E. Moore, G. E. Moore: Selected Writings. London: Routledge. Pp. 23–44.
24. Nagel, T. (1979) “What is it Like to be a Bat?” in T. Nagel, *Mortal Questions* Cambridge: Cambridge Univ. Press. Pp. 165–80.
25. Rockwell, T. (2005) *Neither Brain Nor Ghost: A Nondualist Alternative to the Mind–Brain Identity Theory*. Cambridge, Mass.: MIT Press.
26. Sartre. J.-P. (1987) *The Transcendence of the Ego: An Existentialist Theory of Consciousness*. Trans. F. Williams and R. Kirkpatrick. New York: Farrar, Strauss, and Geroux.
27. Senderowicz, Y. (2021) The “I” that Thinks: Kant on Self-Reference as a Thinker. *Iyyun: The Jerusalem Philosophy Quarterly* 69: 221–247.
28. Swedenborg Foundation (2022) “Who Was Swedenborg?” .
29. Wikipedia (2022) “Transhumanism.” .

# Essentially Embodied Kantian Selves and The Fantasy of Transhuman Selves

**Robert Hanna**

*USA*

## **Abstract**

By “essentially embodied Kantian selves,” I mean necessarily and completely embodied rational conscious, self-conscious, sensible (i.e., sense-perceiving, imagining, and emoting), volitional or willing, discursive (i.e., conceptualizing, judging, and inferring) animals, or persons, innately possessing dignity, and fully capable not only of free agency, but also of a priori knowledge of analytic and synthetic a priori truths alike, with egocentric centering in manifestly real orientable space and time. The basic theory of essentially embodied Kantian selves was spelled out by Kant over the course of slightly less than two decades, between 1768 and 1787, but above all, it flows from an empirical realist and metaphysical reading of the “Refutation of Idealism” that Kant inserted into the Postulates of Empirical Thought section in the 1787 edition of the first Critique. In my opinion, all rational but also “human, all-too-human” creatures like us are, synthetic a priori necessarily, essentially embodied Kantian selves. Let’s call that the essentially embodied Kantian selves thesis, or for short, EEKST. If EEKST is true, then it’s synthetic a priori impossible for the selves of creatures like us to exist independently of our own living organismic animal bodies or beyond the deaths of those bodies, whether temporarily or permanently, by any means whatsoever. Indeed, the very ideas of disembodied selves, their survival after death, and of human immortality, while minimally logically consistent, are in fact conceptually empty and incoherent, even over and above the synthetic a priori impossibility of such things, since the term “myself” indexically picks out an essentially embodied Kantian self, all of whose core features require grounding in a particular living organismic animal body. According to the recent and contemporary movement of transhumanism, the selves of creatures like us can not only exist independently of our bodies, as functional systems of representational content that are inherently able to be implemented or realized in digital-mechanical technology and uploadable to servers, but also to survive accidental or natural human death in server-limbo, then be downloaded into technologically enhanced partially mechanical humanoid bodies or even into wholly artificially-created completely mechanical non-humanoid bodies, survive in these new implementations or realizations for an indefinitely long time, repeat that process, and possibly even become immortal. Transhumanism is in fact metaphysically equivalent to Swedenborgianism, which Kant so effectively criticizes and wittily derides in his 1766 book, *Dreams of a Spirit-Seer Elucidated by Dreams of Metaphysics*. Moreover, and more importantly, if EEKST is true, then, just like Swedenborgianism, transhumanism is not only conceptually empty and incoherent, but also synthetic a priori impossible. And what’s more, it’s also existentially and morally reprehensible. In short, then, the belief in transhuman selves is nothing but a reprehensible noumenal fantasy or *Hirngespinnst*.

**Keywords:** Immanuel Kant, Selves, Embodiment, Swedenborgianism, Transhumanism

**Publication date:** 29.12.2022

**Citation link:**

Hanna R. Essentially Embodied Kantian Selves and The Fantasy of Transhuman Selves  
// Studies in Transcendental Philosophy – 2022. – V. 3. – Issue 3 [Electronic resource].  
URL: <https://transcendental.su/S271326680021060-6-1> (circulation date: 20.05.2024).  
DOI: 10.18254/S271326680021060-6

---

Код пользователя: 0; Дата выгрузки: 20.05.2024; URL - <http://ras.jes.su/transcendental/s271326680021060-6-1> Все права защищены.