Comparative Reflections on Persons and Selves

Comparative Reflections on Persons and Selves

Edited by

Pavel Stankov and Michael Dufresne

Cambridge
Scholars
Publishing

Comparative Reflections on Persons and Selves

Edited by Pavel Stankov and Michael Dufresne

This book first published 2023

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-0220-1
ISBN (13): 978-1-5275-0220-8

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

PAVEL STANKOV AND MICHAEL DUFRESNE

When Immanuel Kant wrote that two things "fill the mind with ever new and increasing admiration," namely, "the starry heavens above" and "the moral law within," he may have not anticipated that two hundred years later the outermost and the innermost would be the chief provinces of metaphysics left where the findings of speculative philosophy would not be quite superseded by empirical science. Today, the study of the self, whatever it is and if it exists at all, is almost as much a chase for something beyond the observable as it was in the Late Enlightenment. This is why a volume on the self such as this one, combining perspectives from various traditions, approaches, and time periods, is very much needed, relevant, and as timely as ever.

The borderlands between philosophy on the one hand and psychology and cognitive and neuroscience on the other, just as that between philosophy and physics, are a productive and flourishing field. The current volume presents but a small segment of it, all of which is the work of the young scholars who presented at the 2018 Uehiro Graduate Student Philosophy Conference, "Cross Currents: Persons and Selves" at the University of Hawai‘i at Mānoa. Five years and a pandemic later, their reflections are as relevant as they have always been in our lifetimes, and maybe in Kant's.

First, in his essay "Relation R Is Not What Matters," Elliot J. Thornley of University of Cambridge, UK, challenges one of the most celebrated theses developed by Derek Parfit in his *Reasons and Persons*, the thesis that Relation R, "psychological connectedness and/or continuity," is what matters and it matters more than the survival of personal identity. According to Thornley, Parfit is inconsistent because the outlined defense of the importance of Relation R commits us to caring both about our replicas and our ordinarily surviving selves in the Branch-Line Case of the famous Teletransporter thought experiment. We cannot, however, have egotistical concerns for more than one entity, and the ordinarily surviving person has no physical or psychological continuity with the replica to justify such concern.

With her paper "Grounding Self-Awareness: Contact and Material Subjectivity," University of Hawai'i's Emma Irwin-Herzog summarizes a discussion starting with Kant and Strawson and extending into today before she provides her response concerning the relationship between the self, the body, and the self's knowledge of objects. Irwin-Herzog argues that the material subject conceptualizes both itself and others through the unavoidable touching of something, in the most obvious case the ground, which connects us to all objects.

In "Persons, Selves, and Dementia," Mathew Sayball of the University of California Santa Barbara asks how we should treat people with dementia. Adopting a Humean/Buddhist conception of selfhood as illusory and a relational conception of personhood inspired by systems of thought as diverse as Ubuntu and Confucianism, Sayball characterizes the self as the "who" and the person as the "what" of human experience. Though personhood and selfhood typically co-exist, people with dementia become wholly or partially disconnected from their identities. For this reason, Sayball argues that we should hold people with dementia in personhood instead of holding them in identity, which means treating them as persons in general irrespective of who they are in particular.

Picking up on the last essay's cross-cultural influences, the next four essays give center stage to comparative philosophy. In "David Hume and Dōgen Zenji: Personhood, Self, and Identity as Evolutionary Impermanence," You Jeen Ha of Smith College, Massachusetts, brings together the ideas of David Hume and Dōgen Zenji to develop the notion of "evolutionary impermanence." According to Ha, both Hume's and Dōgen's accounts of the self permit us to shape and change our realities, providing us with a way to quell the anxieties associated with a permanent sense of self. If we are able to come to terms with the self's impermanence, then we can take control over our responses to the world and guide the self's evolutionary development.

"Watsuji Tetsurō's Rinrigaku and Environmental Ethics" focuses on Waisuji Tetsurō's relevance to environmental ethics. Specifically, Chih-Wei Peng of the University of Hawai'i at Mānoa disputes James McRae's application of Watsuji's philosophy to environmental ethics, declaring it to be strongly anthropocentric and fallacious in its appeal to nature. In opposition to individualist ethical theories, Watsuji claims that humans are both social and individual. He also argues that human "being"—in contrast to Heidegger's *Dasein*—is just as spatial as it is temporal. Based on these ideas, Watsuji develops his conception of "climate" (*fudo*) to show that humans are dependent their environments as much as they are on each other, which Peng describes as a weakly anthropocentric view.

In "Shamed Self to Mindfulness: A Buddhist Alternative to Confucian Shaming," Zulhaqem Zulkifli Zul of Nanyang Technological University, Singapore, offers a Buddhist critique of the Confucian emphasis on shaming in the Chinese tradition. For Zul, Confucian shaming depends on the concept of an individualized self and places the wrongdoer in a bifurcated state between their current degraded self and a potential ideal self. If this ideal self is not actualized, then the emotional states imparted and exploited by shaming could become engrained as a part of one's identity. To counteract this issue, Zul proposes Buddhist mindfulness as an alternative, which detaches individuals from all self-centric emotions and provides them with a more objective view of their actions and emotions.

Finally, in "From Non-Self to Impartial Benevolence: A Constructivist Interpretation of Śāntideva's Argument," Katherine Cheng from University of British Columbia, Canada, offers an interpretation of Śāntideva's thesis that compassion follows from understanding our selfless nature. That view has been challenged by Paul Williams who argues that it fails because it reifies the conventionally existing self or else is unintelligible because it demands the existence of pain without a suffering subject. Mark Siderits's response to Williams, according to Cheng, however, is also unsuccessful because it introduces consequentialist assumptions that need further defense. Cheng also argues against an interpretation based on textual evidence, according to which Śāntideva understands the relationship between self and others as like that between parts of a body working together to stop the body's suffering. According to Cheng's view, Śāntideva's claim is considerably narrower and more modest: once one is committed to impartial benevolence, the realization of the metaphysical truths that permanent selves do not exist is a necessary and instrumental guidance to living in accordance with equal benevolence to all.

The seven essays presented in this collection are diverse to say the least, spanning a multitude of traditions while discussing topics ranging from the relationships of selfhood and personhood to nature and mental health. Despite their many differences, however, we believe that they possess what Ludwig Wittgenstein refers to as a family resemblance, which we have demarcated with the name "persons and selves." While all the chapters have much to offer as stand-alone works, they also complement each other in fascinating ways, which we have done our best to draw out in the chapter order. Our ordering, however, need not limit your understanding of these works, or of this collection as a whole. If you wish to read them "out of order," we encourage you to do so. Or, if you only wish to read the essay(s) most relevant to your own interests, we will not (nor can we) stop you, though we encourage you to give each work a chance.

If these essays have taught us anything, it is that the nature of identity is open for debate. Although this collection may appear to us, the editors, to be unified in one way or another, this unity may prove nothing more than an illusion or a bundle of perceptions in the end. We shall leave this up to you, the reader, to decide.

CHAPTER 1

RELATION R IS NOT WHAT MATTERS

ELLIOTT J. THORNLEY

Introduction

In Part Three of his *Reasons and Persons*, Derek Parfit claims that personal identity is not what matters. What matters, in the way that personal identity is mistakenly thought to matter, is Relation R: "psychological connectedness and/or continuity."[1] He claims that we ought to be egoistically concerned about the fate of some future person if and only if that person is R-related to us.

Since Relation R is preserved by the Teletransporter, Parfit must (and does) claim that one ought to be egoistically concerned about the fate of one's replica. This is so even in *the Branch-Line Case* where one's original brain and body are not destroyed.

In this paper, I argue that this last claim is false. In the Branch-Line Case, one has no reason to be egoistically concerned about the fate of one's replica. This result, in conjunction with a plausible requirement on theories of what matters (that Parfit himself accepts), implies that the same is true of *the Main-Line Case*. It also implies that Relation R is not what matters.

I begin by clarifying some key terms. I explain how Parfit's claim that Relation R is what matters commits him to the claim that one ought to be egoistically concerned about one's replica in the Branch-Line Case. I then assess Parfit's defense of this claim. I conclude that his defense is unsuccessful and that we ought to believe the converse: one has no reason to be egoistically concerned about one's replica in the Branch-Line Case. I end by tracing the implications of this result: one has no reason to be egoistically concerned about one's replica in the Main-Line Case and Relation R is not what matters.

---

[1] Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), 215.

Relation R

In *Reasons and Persons* Derek Parfit defines Relation R as "psychological connectedness and/or continuity" [2] Two persons—X and Y—are psychologically connected in proportion to the number of psychological connections between them. Examples of psychological connections between X and Y include memories (X's remembering having had some experience that Y had), intentions (X's doing something or intending to do something that Y intended to do), desires (X's desiring something that Y desired) and beliefs (X's believing something that Y believed). Two persons are *strongly* psychologically connected if "the number of direct connections, over any day, is *at least half* the number that hold, over every day, in the lives of nearly every actual person."[3] Two persons are psychologically continuous if and only if they are united by overlapping chains of strong connectedness.

Parfit does not claim that Relation R is the true criterion of personal identity. Rather, he claims that Relation R is what *fundamentally matters*. It matters in the way that personal identity is mistakenly thought to matter. By this he means that Relation R is the thing that justifies *egoistic concern* about the fate of some future person. It is the relation we should want to bear to some future person if that person's life will be worth living and the relation we should *not* want to bear to some future person if that person's life will be worse than nothing. Parfit uses this notion of egoistic concern as a reliable indicator of what matters: our relation to some person contains what matters if and only if we ought to be egoistically concerned about the fate of that person.

The main-line case

Parfit asks us to imagine a machine that he calls the Teletransporter.[4] The Teletransporter has two components: a scanner on Earth and a replicator on Mars. In the Main-Line Case, the scanner destroys a person's brain and body (call this person *Scanny*) whilst recording the exact state of her cells. This information is transmitted at the speed of light to the replicator. The replicator then creates, out of new matter, a brain and body exactly like that of Scanny. Call the person that emerges *Replica*.

---

[2] Parfit, *Reasons and Persons*, 215. The "and/or" is provisional. Parfit later claims that both connectedness and continuity matter and that neither matters more than the other (Parfit, *Reasons and Persons*, 301). My argument is unaffected by this claim.
[3] Parfit, *Reasons and Persons*, 206. Emphasis in original.
[4] Parfit, *Reasons and Persons*, 199.

Parfit specifies that Replica will be strongly psychologically connected to and psychologically continuous with Scanny. This specification, in conjunction with Parfit's claim that what matters is Relation R, implies that what matters is preserved in the Main-Line Case. Scanny ought to be egoistically concerned about the fate of Replica. This implication is one that Parfit is happy to accept.[5]

The branch-line case

The Branch-Line Case is slightly different. The exact state of Scanny's cells is recorded and Replica is created on Mars just as before, but Scanny's brain and body are not destroyed. Instead, a person who is physically continuous with, strongly psychologically connected to and psychologically continuous with Scanny walks out of the scanner. Call this person *Scanned*.[6]

Is what matters preserved in the Branch-Line Case? Should Scanny be egoistically concerned about the fate of Replica?

Given that Relation R holds between Scanny and Replica, Parfit must (and does) answer in the affirmative.[7] However, he also recognizes that this claim is liable to strike one as seriously implausible. Given that Scanned is physically continuous with, strongly psychologically connected to and psychologically continuous with Scanny, it seems plausible to claim that Scanny should be egoistically concerned about the fate of *Scanned*. This intuition is likely to be made all the stronger by the recognition that the list of relations given above is not necessarily exhaustive. Aside from the creation of Replica millions of miles away, the relation between Scanny and Scanned is exactly like ordinary survival. Therefore, it seems as if Scanny should be concerned about the fate of Scanned in the same way that we should be concerned about our future selves. She should look forward to Scanned's pleasures and dread Scanned's pains.

However, if this is the case, then it seems that Scanny cannot be egoistically concerned about the fate of Replica. Scanned and Replica are

---

[5] "On my view, my relation to my Replica contains what fundamentally matters." Parfit, *Reasons and Persons*, 285.

[6] Although it seems plausible to claim that Scanny and Scanned are the same person, this claim could conceivably be denied. That is why I distinguish between the two.

[7] In fact, Parfit's claim is even stronger. He claims that the relation between *Scanned* and Replica contains almost everything that matters: "when I am on the Branch-Line, my relation to my Replica contains almost everything that matters" (Parfit, *Reasons and Persons*, 289). Since Relation R holds to a greater or equal degree between Scanny and Replica than it does between Scanned and Replica, this claim implies that the relation between Scanny and Replica contains what matters.

two distinct persons, and the very nature of egoistic concern seems to render egoistic concern about two distinct persons impossible. After all, a person can only feel one person's set of pains.[8] If Scanny expects to feel Scanned's pains, she cannot also expect to feel Replica's pains.

Parfit might try to resist this point by drawing our attention to another of his cases. He might ask us to consider *My Division*:

> My body is fatally injured, as are the brains of my two brothers. My brain is divided, and each half is successfully transplanted into the body of one of my brothers. Each of the resulting people believes that he is me, seems to remember living my life, has my character, and is in every other way psychologically continuous with me. And he has a body that is very like mine.[9]

Call one of the resulting persons *Lefty* and one *Righty*. Parfit claims that "[w]e ought to regard division as being about as good as ordinary survival."[10] He would likely claim that, in this case, one person (Parfit pre-injury) ought to be egoistically concerned about the fate of two distinct persons (Lefty and Righty). Therefore, he might claim, Scanny ought to be egoistically concerned about the fate of Replica as well as Scanned.

However, even if Parfit is right that he ought to be egoistically concerned about the fate of both Lefty and Righty, this does not imply that Scanny ought to be egoistically concerned about the fate of Replica. This is because there is an important difference between the two cases. In the case of My Division, there is physical continuity as well as psychological continuity. Both the brain of Lefty and the brain of Righty are connected to Parfit's pre-division brain by a physically continuous spatio-temporal path. So, if Parfit ought to be egoistically concerned about the fate of both Lefty and Righty, it might be physical continuity that justifies this concern. However, if that is the case, then Scanny ought not be egoistically concerned about the fate of Replica, because Replica is made from new matter. Since there is no physical continuity between Scanny and Replica, the relation between them might not contain what matters.

Given that the case of My Division is inconclusive, how can we determine whether Scanny ought to be egoistically concerned about Replica? I claim that a closer analogy makes the answer clear. Consider *Evil Villain*:

---

[8] These points apply equally to pleasures and all other experiences. I restrict my focus to pains for simplicity's sake.

[9] Parfit, *Reasons and Persons*, 254–55.

[10] Parfit, *Reasons and Persons*, 261.

> Unbeknownst to you, an evil villain obtains a record of the exact state of all your cells. He uses this information to create a replica of you in his secret hideaway. This replica is strongly psychologically connected to and psychologically continuous with you. He tortures this replica.

If this case were to obtain, could you expect to feel any pain? Have you any reason to be egoistically concerned about the fate of your replica? I claim that the answer to both questions is no. The villain might create, torture and dispose of your replica without you ever knowing about it. If you were to learn about your replica's fate, you might feel sympathy, but this is not egoistic concern. It is the same kind of concern you might feel for another human being.

This case is a closer analogue of the Branch-Line Case than My Division. In Evil Villain and the Branch-Line Case, there is psychological continuity without physical continuity. Therefore, if you have no reason to be egoistically concerned about the fate of your replica, Scanny has no reason to be egoistically concerned about the fate of Replica. The relation between Scanny and Replica does not contain what matters.

Parfit might try to push back against this last claim. In the section titled "The Branch-Line Case," he uses *My Physics Exam* to support the claim that the relation between *Scanned* and Replica contains almost everything that matters:[11]

> In this case I divide my mind for ten minutes. In both of my streams of consciousness, I know that I am now having thoughts and sensations in my other stream. But in each stream I am unaware of my thoughts and sensations in my other stream. My relation to myself in my other stream is again like my relation to another person.[12]

However, it is unclear how this case is supposed to establish that "when I am on the Branch-Line, my relation to my Replica contains almost everything that matters."[13] For this to be so, the relation between the two

[11] If the mattering relation is transitive, the truth of Parfit's claim would imply that the relation between Scanny and Replica contains what matters (given that the relation between Scanny and Scanned contains what matters). I make no argument either way concerning the transitivity of the mattering relation. Since I argue that the relation between Scanned and Replica does not contain what matters, the transitivity of the mattering relation makes no difference.

[12] Parfit, *Reasons and Persons*, 288.

[13] Parfit, *Reasons and Persons*, 289.

streams of consciousness in the Physics Exam would have to (1) be like the relation between two persons and (2) contain almost everything that matters. But if there is a person inhabiting each stream of consciousness then these persons will necessarily have different experiences: they will think different thoughts and feel different pains. One person cannot feel the pains of the other. Given that this is so, each person cannot be egoistically concerned about the fate of the other. The relation between the two persons cannot contain what matters. Therefore, My Physics Exam cannot help to defend the claim that the relation between Scanned and Replica contains almost everything that matters.

What matters is not preserved in the main-line case

Neither My Division nor My Physics Exam can help justify Parfit's claim that the relation between Scanny and Replica in the Branch-Line Case contains what matters. Rather, the Evil Villain case shows that it is much more plausible to claim that the relation between Scanny and Replica does *not* contain what matters. Scanny has no reason to be egoistically concerned about the fate of Replica.

This result, in conjunction with a plausible requirement on theories of what matters, implies that the relation between Scanny and Replica does not contain what matters in the Main-Line Case. The requirement is as follows:

> Whether my relation to some future person contains what matters must depend only on the intrinsic features of the relation between us. It cannot depend on what happens to other persons.

I believe this requirement is plausible *prima facie*, and Parfit himself accepts it.[14] But it implies that the relation between Scanny and Replica does not contain what matters even in the Main-Line Case where Scanny's brain and body are destroyed. After all, the intrinsic features of the relation between Scanny and Replica are the same in the Branch-Line Case and the Main-Line Case. In both cases, Replica is strongly psychologically connected to and psychologically continuous with Scanny. But if these facts do not justify egoistic concern in the Branch-Line Case, they cannot justify egoistic concern in the Main-Line Case. The non-existence of another

[14] "Instead of asking whether I shall be some future person, I ask whether my relation to this person contains what matters. Like [Bernard] Williams, I can claim that the answer must depend only on the intrinsic features of my relation to this future person" (Parfit, *Reasons and Persons*, 271).

person—Scanned—cannot make any difference. Therefore, what matters is not preserved in the Main-Line Case. If you were to use the Teletransporter, you would have no reason to be egoistically concerned about the fate of your replica.

Relation R is not what matters

My conclusion about the Branch-Line Case implies another, independent conclusion: Relation R is not what matters. The argument for this conclusion is very simple:

(P1) If Relation R is what matters, then the relation between Scanny and Replica in the Branch-Line Case contains what matters.
(P2) The relation between Scanny and Replica in the Branch-Line Case does not contain what matters.
(C) Relation R is not what matters.

The first premise is true in virtue of Parfit's stipulations about the Teletransporter. The Teletransporter is such that Relation R holds between Scanny and Replica in the Branch-Line Case. The second premise is my conclusion about the Branch-Line Case. The conclusion follows by *modus tollens*.

Conclusion

Parfit's claim that Relation R is what matters commits him to the claim that the relation between Scanny and Replica contains what matters in the Branch-Line Case. This claim is implausible *prima facie*, and Parfit's mentioning of cases like My Division and My Physics Exam fails to reduce this implausibility. Instead, Evil Villain gives us strong reason to believe that the relation between Scanny and Replica does *not* contain what matters. This result has two independent implications. The first is that, given the truth of a plausible requirement on theories of what matters, what matters is not preserved in the Main-Line Case. The second implication is that Relation R is not what matters.

Bibliography

Parfit, Derek. *Reasons and Persons*. Oxford: Clarendon Press, 1984.

CHAPTER 2

GROUNDING SELF-AWARENESS: CONTACT AND MATERIAL SUBJECTIVITY

EMMA IRWIN-HERZOG

Introduction

Is it possible to have knowledge of things? If so, how? Previous answers to these essential epistemological questions have made an indelible distinction between knower and known. That is, the bifurcation of that which *knows* and that which *is known* is taken to be a condition necessary for knowledge of things. The necessity of this split is supported by grammatical investigations of the nominative-accusative distinction and analyses of agents and recipients of action. For example, it is said that the finger cannot touch itself. In other words, the thing at hand cannot occupy both the subject and the object position in a proposition nor can it be both the agent and the recipient of action. With respect to the self, the implication of this division is that the self cannot be both subject (that which knows) and object (that which is known).

There are, however, traditions which speak of self-knowledge and self-awareness. What, then, can we say about knowledge or awareness of the self? Can the self be both that which knows (the subject) and that which is known (the object)?

I will attempt to problematize this crucial distinction. I will begin by sketching Kant's view of the self, and then I will raise Peter Strawson's objection that Kant does not adequately acknowledge the body as a criterion for the empirical application of subject-identity. Strawson's point raises further questions: does embodiment entail objectivity? How does our body relate to the self and to our knowledge of things? To address these questions, I will first consider the phenomenological account provided by Samuel Todes. I will explicate his thesis that the human body is the *material subject* of the world. Quassim Cassam similarly reconstructs Kant's philosophy of the self with a focus on addressing the supposed elusiveness of the self, i.e.,

the idea that the subject cannot be known as anything other than a formal unity. My overview of his position will focus on his thesis that self-awareness requires intuitive awareness of the self, *qua subject*, as an object among other objects. Both Todes and Cassam address self-awareness in the context of the Kantian epistemological project of identifying the conditions of the possibility of having knowledge of things. Both accounts also overcome the distinction between the self as subject and object.

My argument will supplement Cassam's claims and will make use of Todes' general phenomenological method. I will argue that the self known, qua subject, is not known as a *merely material* object nor as an object *merely among* other objects. My contention will be that the material subject is presented moreover as an object *in direct contact with* other objects. In plain terms, my claim will be that awareness of the material subject will always be awareness of a body which, at the very least, *touches something*—paradigmatically, the ground or the earth. The phenomenon of touching—of reaching out and placing one's finger upon, for example, a pane of glass—generally exemplifies the way in which the subjective body can be understood as both feeler and felt. This observation, in itself, undermines the strict bifurcation of knower and known, of subject and object. The phenomenon of *standing on the ground*, as a more universal and less contingent form of touching, constitutes an expanded condition of the possibility of self-awareness: intuitive awareness of the self, qua subject, as an object among and moreover *in contact with* a vast multitude of other objects similarly in contact with one another.

Kant on the self: the unity of consciousness

Kant admits three grades of the self: the bodily self, the empirical self, and the unity of consciousness. The bodily self for Kant is an empirically knowable embodied phenomenon. Kant's empirical self is the "I" understood as inner appearance, as whatever constitutes the contents of an empirical consciousness. This reflexive knowledge of empirical consciousness—of an inner succession of temporal and empirical objects—is essentially consistent with a bundle theory of the self. For Kant, the subject is not entirely captured by the Humean bundle or by knowledge of the empirical "I" understood as such. The unity of consciousness is the condition of empirical self-consciousness. It refers to the formal unity or general connectedness that is the transcendental unity of apperception. The transcendental unity of apperception is the *basis of the possibility of* the categories. As such, it must be presupposed in order to cognize an object at all. It therefore cannot be cognized as an object. In other words, the

transcendental unity of apperception is *the condition for the application of* the categories.[1] It therefore cannot be *conditional upon* them.

How do these three grades of the self relate to Kant's rejection of rational psychology? His rejection of rational psychology amounts to a rejection of immediate awareness or assurance of a substantial or identical self. It is clear that the "I" in the Cartesian proposition "I think" does not refer to the bodily self or to the empirical self as described above. Kant claims that the Cartesian philosopher of the soul confuses the unity of experiences (the transcendental unity of apperception) with the experience of unity (the substantial and identical soul). What is the difference between the two? The unity of consciousness is a formal unity. The rational psychologist, through invalid syllogistic argumentation, objectifies and reifies this merely formal unity. But according to Kant there are no criteria for the empirical application of the identify of a substantial self. This last point turns on what Peter Strawson calls "the principle of significance." According to Strawson, "in order to claim knowledge of the existence of an object falling under a certain concept, we must have, and must have occasion to make use of, empirical criteria for the application of that concept."[2] In other words, knowledge of any given object requires empirical application.

Let's take a closer look at Kant's paralogisms: again, the rational psychologist takes the totality of conditions of the thinking subject—the purely formal "I think"—to be an object, mind or soul. From this objectified subjective formality, the rational psychologist derives synthetic conclusions about the nature of that "object." The conclusions invalidly derived are as follows: substantiality, simplicity, numerical identity, immateriality, immortality, and incorruptibility.[3] The Cartesian self is thus *substantial yet immaterial*. Kant's contention with respect to the syllogisms used to derive these synthetic conclusions about the self is that the arguments commit equivocation on a term occurring in the major and minor premises. In the first paralogism,[4] the slippage in meaning occurs between the two senses of the term "subject:" the first premise establishes mere logical subjectivity and the second premise requires (or assumes) more than that Furthermore, for an inference to be made from the substantiality of the logical subject to the substantiality of the "I think" there must be a given object corresponding to the logical subject. The major premise, correctly understood, requires

[1] Immanuel Kant, *Critique of Pure Reason: Unified Edition*, trans. Werner S. Pluhar (Indianapolis: Hackett Publishing Company, 1996), A111–12, 161.
[2] Peter F. Strawson, *The Bounds of Sense* (New York: Routledge, 1991), 162.
[3] Kant, *Critique of Pure Reason*, A344/B402, 384.
[4] Kant, *Critique of Pure Reason*, A348–51, 387–89.

knowledge of objects as substances. But the category of substance requires empirical application. An inference cannot be made about the substantiality of the "I think" because the "I think" is purely formal and non-empirical. That is, the Cartesian fails to recognize, according to Kant, the conditions of the application of the category of substance. Thus the slippage between the two senses of the term "subject" can be rephrased as such: the first premise only establishes logical subjectivity *even though substantiality is predicated of it* (for there are no conditions given under which this category can be applied) and the second premise requires more than mere logical subjectivity. The rational psychologist gives no conditions under which the logical subject in the first premise can be known to be substantial, and those conditions are exactly what is needed. The application of the concept (the category) of substance *requires reference to experience of something permanent.* But there is nothing permanent about the "I think." There is no concept of substance involved in synthesizing the "I"—only the unity of apperception is involved in synthesizing it.

In the case of the third paralogism, the equivocated term is identity.[5] Identity can refer either to the sameness of two things, to the singleness of a subject of experience or to the retention of core-essence through alterations in time. Again, the major premise does not provide the conditions under which an object corresponding to the representation "that which is conscious of the numerical identity of itself" can be given. And this object must be given in order for the inference to be made. In the case of the first paralogism, the criteria for the applicability of the category of substance to the "I think" was missing because what was required was reference to something permanent and there is nothing permanent about the unity of apperception. What, in this case, is the empirical criterion for the applicability of the concept of identity? Again, permanence must be presupposed. But the "I" which accompanies all appearances—the transcendental unity of apperception—is the condition *for* the possibility of the categories and as such is not (and cannot be) conditional *on* them. It therefore cannot be cognized as a permanent or substantial object. Thus there are, according to Kant, no criteria for the empirical application of subject-identity.

Generally speaking, Kant claims that there can be no legitimate inference made from the unity of apperception to that of a temporally permanent substance. We mistake the transcendental unity of apperception (the necessary connectedness of experiences which is the condition for the possibility of the empirical consciousness) for awareness of a unitary (identical) subject. Kant exposes the illusion of there being *immediate*

[5] Kant, *Critique of Pure Reason*, A361–66, 396–99.

assurance of the substantive, *identical* self. He rejects the position that the self is an object of thought (a knowable subjective object) but not of intuition or experience. As such, he maintains the distinction between the subject-self and the object-self.

The inadequacy of Kant's answer: Strawson's criticism

Peter Strawson reminds us that the transcendental unity of apperception is the connectedness and unity of a set of experiences, "secured to it by concepts of the objective."[6] I take this reference to objectivity as a further reminder that the transcendental unity of apperception is not just a *subjective* necessity (not just a subjective unity of experience). Instead, the transcendental unity of apperception is an *objective* condition of all knowledge (this is what is meant by the *objective* unity of consciousness). Furthermore, the connected experiences are experiences of an objective world. According to Strawson, this objectivity, combined with the unity of one set of connected experiences, gives us *the basic ground* but not the *full conditions* of "the possibility of an empirical use for the concept" of the self as a persisting immaterial subject of experiences.[7] No empirically applicable criteria of identity are given by the transcendental unity of apperception understood as this merely formal unity of experiences.

Strawson points out that immediate self-ascription, or the ascription of any given mental state to oneself, or the use of the pronoun "I" in this context, does not require any criteria of personal identity as justification. In other words, I can say "I feel sad" without justifying my use of the personal pronoun "I" by making a distinction between myself and something else. It would be meaningless to ask, **"**I am hungry, but is that feeling of hunger manifesting in *me*?**"** First-person reference thus holds up even in the absence of necessary justification according to a criterion of personal identity. In other words, I can still *in practice* use the personal pronoun "I" to meaningfully refer to myself as a subject (subject: the unity of apperception, the necessary connectedness of experience). Given that immediate self-ascription doesn't require reference to criteria of personal identity, there is the tendency to continue to use "I" referentially, i.e., to think of it as referring to a subject only on the basis of inner sense, and to thus think of that subject as a Cartesian soul. We can see that the pronoun "I" has been abstracted from its ordinary setting, in which it is used as first-person reference to the empirical consciousness. Kant doesn't acknowledge

[6] Strawson, *The Bounds of Sense*, 163.
[7] Strawson, *The Bounds of Sense*, 163.

this explicitly. Instead, he connects the "delusive" use of I with the unity of consciousness. Our delusive non-referential thought of the subject, which abstracts from the condition of outer sensibility, is that of the connectedness of experiences—the transcendental unity of apperception. Strawson suggests that when, in the grips of the illusion, we abstract from empirical application, we nevertheless do not abstract from conditions required by the transcendental unity of apperception. As such, we mistake the transcendental unity of apperception (the necessary connectedness of experiences which makes possible self-reflexive awareness of the empirical consciousness) for awareness of a unitary (identical) subject.

Strawson's major criticism is that Kant does not make explicit enough the necessary role of empirically applicable criteria of subject-identity. Moreover, Kant does not press the point that in order to speak of *one* consciousness, we must make reference to *one* person. That is, we do indeed have criteria of identity for subjects of experiences—bodies. Reference to empirical or bodily identity is, in Strawson's view, necessary but underappreciated by Kant.[8]

Kant has shown us that subjectivity does not entail immaterial substantiality. That is, he rejects the Cartesian soul and affirms the formal unity of consciousness—the transcendental unity of apperception which is neither an immaterial substance nor a mere Humean bundle of mental states. Strawson's criticism suggests that reference to the body can satisfactorily serve as criteria for the empirical applicability of subject-identity. If reference to the body does indeed satisfy the principle of significance, what then, is the body in relation to the self? Is it merely the bodily self as it was understood by Kant—as an empirically knowable empirical phenomenon? Does embodiment entail objectivity? In other words, must the bodily self be considered an object? If so, in what sense? Is the bodily self an object like all other objects? Is there a sense in which the bodily self is both subject and object? What is the role of our body, qua subject, as an object, especially with respect to our knowledge of things?

Todes's phenomenological investigation: the material subject

Samuel Todes maps various views of the self in relation to the body. The classic view is of the human subject as what he calls the "migrant-subject." The migrant-subject travels from identification with the body to identification

[8] Strawson, *The Bounds of Sense*, 168.

with a subjectivity beyond the body.[9] For Aristotle, the human subject is something in the world with the capacity to conceptually identify with the unity of the world by forsaking its own substantial state. That is, the human subject cannot be both identified with the unity of the world and in the world.[10] Plato similarly advocates turning away from the body and toward the mind. For both Plato and Aristotle, "the measure of identity with the unity of the world achieved by the non-body element of the human subject is constructed…as a measure of release from being in the midst of things in the world, in favor of a state of pure thought beyond the world."[11] In other words, the human subject must move beyond that which is immersed in the world (the body) toward that which, through pure thought, attains knowledge of transcendental categories (being, unity, truth, etc.) and thus also succeeds in thinking itself. Self-awareness, in this model, is achieved through non-identification with the body.

Descartes rejects this classic view.[12] Todes introduces the Cartesian view of the necessary conceptuality of the human subject. Necessary conceptuality effectively erases the idea that the human subject is, in any sense, to be identified with the body. The Cartesian substantial soul, is, of course, the immaterial, substantial, thinking self. Todes suggests that Descartes vacillates between positing the subject as something equivalent to the ordering unity of the world and something only somewhat more than an object merely in the world. In sum, Descartes separates the subject from the body but does not raise the subject to identification with "the ordering unity of the world itself," as per the classic view. Todes calls the Cartesian treatment of the human subject "ambiguous."[13]

Hume, according to Todes, considered the self qua subject to be identified "only with the function of explicating the unity of the world" and never with the body.[14] Thus the Humean subject loses touch with the body.[15] The body, for Hume, is an object in the field of human experience, like all other objects.[16] The Humean view thus separates the subject from the body, relegates the body to the status of a mere object in human experience, and posits the subject as a spectator of the body as a dismembered object.

---

[9] Samuel Todes, "The Human Body as Material Subject of the World," PhD diss., (Harvard University, 1963), 98.
[10] Todes, "The Human Body as Material Subject of the World," 11.
[11] Todes, "The Human Body as Material Subject of the World," 12.
[12] Todes, "The Human Body as Material Subject of the World," 13.
[13] Todes, "The Human Body as Material Subject of the World," 24.
[14] Todes, "The Human Body as Material Subject of the World," 99.
[15] Todes, "The Human Body as Material Subject of the World," 44.
[16] Todes, "The Human Body as Material Subject of the World," 48.

How can we make sense of the subject as pure spectator? Does this view hold up? How might the body be known as an object unlike other objects in experience? To answer these questions, Todes engages in a phenomenological investigation of how the subject might perceive the body as a merely material object. He arrives at a number of conclusions regarding the operation of the active body in the midst of the world. The upshot of his conclusions is that *the body is not merely a material thing in the world*. It is rather and moreover "that material thing whose capacity to move itself generates and defines the whole world of experience in which any material thing, including itself, can be found."[17] Todes explains that both the classic view of the migrant-subject and the post-Cartesian view of the world-subject affirm the mutual exclusivity of the subject identified with the body as an object in the world and the subject identified with the "function of explicating the unity of the world."[18] He explains that both views make the mistake of taking the body to be *merely a material thing in the world*. We might ask this: if the human body is not merely a material thing in the world, what is it? Todes answers that it is a uniquely material thing in the world which is the defining subject of the world. His central contention is that "the human body is the material subject of the world."[19] Todes thus effectively synthesizes subject and object in the body.

Todes subsequently discusses Kant's view that "the human subject makes his world of experience."[20] According to Todes, Kant synthesizes dogmatism and skepticism with respect to the relation between the human subject and the ordering unity of the world. Todes reconstructs the Kantian position as such: representations (as opposed to objects) make objects possible as knowable objects, and representations are produced by the knower.[21] Todes rejects the idea that there are only two possibilities with respect to the relationship between synthetic representations and objects. Kant thinks that either objects make representations possible or representations make objects possible. Through his phenomenological investigations, Todes reveals a third possibility: "the conformity of knowledge to object seems to derive neither from the subject nor the object but from the commonality of the world condition under which they exist together."[22] That is, the subject is in the world not as a merely material object but as an object with determining power. Todes furthermore claims that the commonality of the

[17] Todes, "The Human Body as Material Subject of the World," 99.
[18] Todes, "The Human Body as Material Subject of the World," 99.
[19] Todes, "The Human Body as Material Subject of the World," 99.
[20] Todes, "The Human Body as Material Subject of the World," 184.
[21] Todes, "The Human Body as Material Subject of the World," 185.
[22] Todes, "The Human Body as Material Subject of the World," 187.