QUOTE:AUGUSTINE, DECARTES, HUME, BUDDHISM.

A PRIORI SUBJECTS:

Kant and the Existence of the Soul

Stephen Priest

Stephen Priest is Senior Research Fellow in Philosophy at the University of Oxford.

Preface

There is a tension between Kant's anti-metaphysics and his Pietism. In his moral philosophy, Kant argues that the immortality of the soul is a postulate of pure practical reason. Yet, in the *Critique of Pure Reason* Kant offers arguments for the conclusion that the solutions to some problems of the self do not entail the existence of an immaterial soul. My restricted aim in this book is to decide whether any member of these sets of arguments is sound, and the tension relieved. The tension matters, because philosophy subsequent to Kant tends to endorse his anti-metaphysics but repudiate his Christian faith.

Because this is a philosophy book, not a work in the history of ideas, I have not attempted to reconstruct the whole of Kant's thinking about the self, nor to locate it in a historical context. In philosophy, the relevant units are the problems, their putative solutions, and the arguments for those putative solutions. My approach to Kant is therefore more like Jonathan Bennett's than those of Karl Ameriks or Paul Guyer.

By 'understanding' Kant I mean the allocation of propositions to his sentences in a way that maximises their consistency as a set. The justification of this method is: To the extent to which some sentences form an inconsistent set they express nothing. I point out places where Kant's arguments have been misunderstood by major commentators and substitute more viable interpretations. Because we operate in an anti-metaphysical age, in which Cartesianism is held up as a paradigm of philosophical falsehood, commentators tend to be sympathetic to Kant's attack on rationalist psychology. I argue that the problems, once understood in their profundity and with clarity, can only be solved if we are in fact souls. This book is therefore unusual in being largely a defence of the conclusions of 'the rationalist psychologist' against Kant's attack.

I thank A. J. Ayer, Graham Bird, Edward Craig, Benedikt Göcke, Peter Hacker, Michael Inwood, Adrian Moore, Terence Penelhum, Bernard Williams and Timothy Williamson for discussion of Kant's philosophy or problems of the self or both. I am grateful to the Philosophische Fakultät of the Westfälische Wilhelms-Universität Münster, Germany for their hospitality. In particular I thank, Prof. Dr. Andreas Hütteman, Prof. Dr. Harald Holz and Prof. Dr. Peter Rohs for useful discussion of the invited lectures I delivered there about Kant. I also thank the Faculteit der Filosophie of the Radboud Universiteit, Nijmegen, Netherlands, for their hospitality. In particular I thank Professor Grahame Lock, and Tjeerd and Sjoerd van Hoorn for their contributions to a two-day seminar generously devoted to my ideas.

Stephen Priest Trinity 2019
Oxford

Contents

I

Kant's Attack on Rational Psychology

In the *Critique of Pure Reason*, Kant calls the project of proving Cartesian dualism by *a priori* means 'rational psychology'. (1) The structure of Kant's argument against the possibility of rational psychology is:

(1) 'The proposition 'I think' (taken problematically) contains the form of each and every judgement of the understanding and accompanies the categories as their vehicle.'

Da nun der Satz: Ich denke, (problematisch genommen,) die Form eines jeden Verstandesurteils überhaupt enthält, und alle Kategorien als ihr Vehikel begleitet

(2) 'It is evident that the inferences from it admit only of a transcendental employment of the understanding.'

So ist klar, daß die Schlüsse aus demselben einen bloß transzendentalen Gebrauch des Verstandes enthalten können

(3) 'This employment excludes any admixture of experience.' Welcher alle Beimischung der Erfahrung ausschlägt

(4) 'We cannot, after what has been shown above, entertain any favourable anticipations in regard to its methods of procedure.'

Und von dessen Fortgang wir nach dem, was wir oben gezeigt haben, uns schon zum voraus keinen vorteilhaften Begriff machen können.
(CPR 368, B406)(1)

His claim in (1) that the expression 'I think' contains the form of every judgement, is a formal requirement on any judgement that it admit of being couched in first person singular grammatical form, and thereby could in principle be thought. This constraint is familiar from the doctrine of the transcendental unity of apperception: If a judgement it to count as such, and be an episode in a self-conscious rational mind, it

must logically admit of re-phrasing using the sentential prefix 'I think [...]'. If it were logically impossible for a putative judgement to be couched in this form, then it could not be an occurrence in a self-conscious rational mind, so could not be a thought, and so could not be a judgement either. It is for Kant a necessary truth, because analytic, that a person thinks their own thoughts, and all thoughts are somebody's, so the proposition 'I think my thoughts' is analytic. It follows that 'I think my thoughts' is not a material or informative sentence, and the 'I think […]' prefix is merely a formal requirement on my thinking and does not give any information about my nature.

It is because the 'I think' accompanies judgements as a formal condition for their existence that it is also a formal condition on the use of the categories. If the categories' only use is in judgements, and if the possible employment of the 'I think [...]' prefix is a condition of the use of judgements, then it follows that the possible use of the 'I think' prefix is a condition for the use of the categories also. (1) then, expresses synoptically Kant's theory of the correct use of 'I think', and exhausts the essentials of what is philosophically legitimate to claim about it.

The second premise, (2) follows from (1) if (1) is correctly understood. If the possibility of the 'I think [...]' prefix accompanying the use of judgements is a condition for the use of judgements, then that is a transcendental fact about it. If that fact about the 'I think' is the only fact about it that is philosophically well-founded, then the transcendental fact about it is the only fact about it that is philosophically well-founded. Further, if the 'I think' admits only of a transcendental employment, then any sentence which may be logically derived from 'I think' must admit also of only a transcendental employment. This is why for Kant it is 'evident' that the inferences obtained from 'I think' will have only a transcendental import. He means logically evident.

(2) follows from (1) because no transcendental fact is an empirical fact. Kant would not wish to deny that the sentence 'I think' has an empirical use, nor that the expression might have an empirical use when prefixed to a judgement. But from the fact that an expression possess both a transcendental and an empirical use it does not follow that the transcendental use possesses any empirical features, or that the empirical use possesses any transcendental feature. Indeed, Kant insists that transcendental and empirical descriptions are mutually exclusive (and collectively exhaustive) so that if *p* expresses a transcendental fact then *qua* transcendental fact it is not an empirical fact, and if *p* expresses an empirical fact then *qua* empirical fact it is not a transcendental fact. Clearly, this is quite consistent with one and the same set of words being used to express both transcendental and empirical facts.

So, by (2) Kant is claiming that the transcendental unity of apperception's 'I think' is a transcendental fact; a condition on self-conscious thought, and *qua* transcendental fact it admits of only a transcendental use. No empirical facts follow from 'I think' construed transcendentally.

(3) follows from (1) and (2) because if 'I think' is a transcendental fact then *qua* transcendental fact its use is not empirical. If it were empirical it would include an

'admixture of experience', that is, it would express information that could be inferred from experience.

In (4) we have to read 'favourable' to mean 'metaphysically favourable' or 'favourable to the projects of rational psychology'. Kant thinks nothing non-transcendental follows from 'I think' so, *a fortiori* not only does nothing empirical follow from 'I think' but nothing metaphysical either. So, in (4), 'after what has been shown' refers back to the first three premises interpreted according to the theory of the transcendental unity of apperception, and 'its methods of procedure' refers to the transcendental method of treating ' I think'.

Kant's argument is valid, and if the premises are true then it is also sound, but accepting those premises largely depends on accepting the central doctrines of the transcendental deduction, so accepting the conclusion depends on accepting those doctrines also.

In particular, it might be doubted that 'I think' *only* has empirical or formal import. For example, 'I think' might not be straightforwardly empirical if the docrine of the 'systematic elusiveness of the I' is true or if, when aware of my own thoughts, I am not aware of them through experience of them or both. If the self systematically eludes experience then it eludes experience, but nothing that eludes experience is empirical, so knowledge of the I is not empirical. If I am not aware of my own thoughts through experience, but knowledge of anything empirical is by experience, then my knowledge of my own thoughts is not empirical. In either case, 'I think' is not empirical or, if some empirical content may be granted to me and my thoughts, not only empirical or not wholly empirical. On the other hand, depending on what we allow and exclude under 'formal', it might be hard to construe 'I exist' as only a formal truth. It entails, for example, at least: There is something rather than nothing. There is thinking. There are events. Time exists. There is thinking. I am (even I should turn out to be nothing over and above the thinking). There is as much unity of consciousness as is necessary for thinking to go on. These entailments are not the hallmark of a wholly formal proposition.

Arguably, the logical status of 'I think' is contingent *a priori*. 'I think' is *a priori* because its truth is decidable by thought, independently of sense experience. 'I think' is contingent because I might not have thought, or that thinker who I am might not have existed. It is therefore wrong to think 'I think' is a necessary truth. To conclude this is almost invariably to confuse the necessity of the inference from 'I think' to 'I think' with the modal status of the conclusion. Being *a priori* is inconsistent with being empirical so if 'I think' is contingent *a priori* then 'I think' is not empirical. Being contingent is inconsistent with being only formal so if 'I think' is contingent *a priori* then 'I think' is not only formal.

If Kant is wrong to conclude that the sentence 'I think' only expresses an empirical proposition or a formal proposition what does it express? If all the facts are empirical, formal or metaphysical and 'I think' is not empirical or formal then 'I think' expresses a metaphysical proposition. Of course, Kant's transcendental idealism, which entails a metaphysical anti-realism, is designed to rule out this

possibility. However, if the arguments adduced above are sound, not all the presuppositions of ‘I think’ are empirical or formal. If the referent of ‘I’ in ‘I think’ denotes a subject it denotes an *a priori* subject: one which may be known independently of sense experience. It is known as a presupposition of experience, not as an item falling within experience.

Kant has a second, more general, argument for the conclusion that nothing metaphysical follows from 'I think':

(1) 'I do not know an object merely in that I think, but only in so far as I determine a given intuition with respect to the unity of consciousness in which all thought consists.'

Nicht dadurch, daß ich bloß denke, erkenne ich irgend ein Objekt, sondern nur dadurch, daß ich eine gegebene Anschauung in Absicht auf die Einheit des Bewusstseins, darin alles Denken besteht, bestimme, kann ich einen Gegenstand erkennen.

(2) '*Modi* of self consciousness in thought are not by themselves concepts of objects (categories) but are mere functions which do not give thought an object to be known.'

Alle modi des Selbstbewusstseins im Denken, an sich, sind daher noch keine Verstandesbegriffe von Objekten (Kategorien) sondern bloße logische Funktionen, die dem Denken gar keinen Gegenstand [zu erkennen geben]

(3) 'Accordingly, (they) do not even give myself as object.'

Mithin mich selbst auch nicht als Gegenstand zu erkennen geben.

(4) 'The object is not the consciousness of the determining self, but only that of the determinable self, that is, of my inner intuition.' (CPR 368, B 406-7)

Nicht das Bewusstsein des Bestimmenden, sondern nur des bestimmbaren Selbst, d.i. meiner inneren Anschauung [...] ist das Objekt.

(5) 'Consequently, I do not know myself through being conscious of myself as thinking, but only when I am conscious of the intuition of myself as determined with respect to the function of thought.'

Also erkenne ich mich nicht selbst dadurch, dass ich mich meiner als denkend bewußt bin, sondern wenn ich mir die Anschauung meiner selbst, als in Ansehung der Funktion des Denkens bestimmt, bewußt bin.

(1) is the claim that nothing can be known about what exists, by mere thinking: so all knowledge of what is, is *a posteriori* and not *a priori*. If knowledge of what exists could be obtained by pure thought, then there would be *a priori* knowledge of what exists. In fact all knowledge of what exists is empirical. This is the force of 'only in so far as I determine a given intuition'. Only in so far as a person finds intelligible some item of experience is it possible for that person to obtain knowledge of what exists, or of the nature of what exists.

Thus far (1) expresses a version of empiricism and denies a version of rationalism, but also included in this is the very Kantian requirement that any putative experience occur to a unified consciousness. This too is a condition of knowledge of objects because if it were not fulfilled we could not talk of persons (or 'my') acquiring knowledge of objects.

I take it 'determine' here is implicitly a reference to synthesis because if an experience occurs within a unified consciousness then that experience is subjected to synthesis, that is, it is organised categorically. Only in that context where, for example, the categories of substance and causation find their only legitimate application does it make sense to talk about objects at all.

In (2) Kant says he means by a modus of self consciousness a mere function. This is a reference to the 'I think […]' prefix. By 'function' he means 'grammatical function'; or perhaps 'logical function'. We may think of the 'I think […]' prefix as a quasi-formal operator on sentences which exhibits their first person singular psychological properties. Then (2) can be interpreted as the claim that the expression 'I think' is not or does not contain a category. For example 'substance' is a category, but the sentence 'I think' does not contain any categories so the category of substance cannot be logically unpacked from it.

When he says the 'I think […]' prefix does not 'give thought an object to be known' he means that it cannot be concluded that any entity with specifiable properties may be inferred to exist from the fact that 'I think' has a use. It might also be part of his meaning that the thinking of 'I think', its mental entertainment, does not present the thinker to themselves as object. Clearly that is impossible on Kant's account, because objects are only given in experience and, in thinking 'I think', I do not appear to myself to be an object (at least in the sense of a substance) in experience. 'Know myself' here could either mean 'come to be indirectly acquainted with myself' or 'come to acquire knowledge of acts about myself'. Whichever way we read Kant here, he is saying knowledge of oneself as an object is not possible by the thinking of 'I think'.

A mode of self consciousness, as a function, is explicitly contrasted with being a category because categories are concepts of objects, and when applied to intuition do indeed yield what may be properly called 'knowledge of objects'. The 'I think […]' prefix has no such function.

(3) follows logically from (2), because if it is true that the 'I' think […]' function cannot be used to obtain knowledge of any object, then *a fortiori* it cannot be used to obtain knowledge of myself as an object either.

(4)'s 'the object' is the object of thought: what thought is thought about. Kant is saying that if there is an object of thought in a sentence prefixed by the 'I think […]' function then that object is myself as I appear to myself in intuition: say, in inner sense.

He distinguishes the determining self from the determinable self. This is not an ontological distinction between two numerically distinct entities. It is the difference between the self *qua* subject of thought and action, and the self *qua* object of thought and experience. As a determining self I am active in synthesizing the contents of my intuition through the application of the categories. As a determinable self I am the possible object of that synthesis and those intuitions.

Now, Kant is saying that when I am self-conscious, necessarily, I am conscious of myself only as a determinable self and not as a determining self. In other words, I may only be conscious of myself as a possible object of my own intuition. It follows that there cannot be self-consciousness that is consciousness of a putatively metaphysical self, which, by implication, would be irreducibly subjective, active, non-physical and yet substantial.

(5) follows from (4) because it is largely a rewriting of it. If I do not know myself as an object though mere thinking then it clearly follows that 'I do not know myself through being conscious of myself' so long as this means 'only by being conscious of myself. An additional point is made in (5), however, when Kant adds 'as thinking' to 'being conscious of myself'. He is saying, suppose I am conscious of myself as thinking because, for example, I might judge that ''I think' is true. Then even that thought: the thought that 'I think' is true, does not yield any empirical or metaphysical knowledge of myself, (other than trivially, that 'I think' is true). Specifically, nothing follows about the nature of the 'I'. *a fortiori*, nothing Cartesian follows about the nature of the I.

The rest of (5) is a reiteration of the thesis that I may only know myself as an object of my own intuitions, and those intuitions are subject to the formal requirement of the 'I think' prefix having its use in the transcendental unity of apperception.

This argument is also valid, but like the last, its soundness depends upon the truth of transcendental idealism.

The First Paralogism

The First Paralogism is stated in the first edition and left unaltered for the second. Called ‘Of Substantiality’, it is:

(1) ‘That, the representation of which is the *absolute subject* of our judgments and cannot therefore be employed as determination of another thing, is *substance*.’

Dasjenige, dessen Vorstellung das absolute Subjekt unserer Urteile ist und daher nicht als Bestimmung eines andern Dinges gebraucht werden kann, ist Substanz.

(2) 'I, as a thinking being, am the *absolute subject* of all my possible judgments, and this representation of myself cannot be employed as predicate of any other thing.'

Ich, als ein denkend Wesen, bin das absolute Subjekt aller meiner möglichen Urteile, und diese Vorstellung von Mir selbst kann nicht zum Prädikat irgend eines andern Dinges gebraucht werden.

(3) 'Therefore I, as thinking being (soul), am *substance*.' (A 348)

Also bin ich, als denkend Wesen (Seele), Substanz.

(1) is Kant's definition of 'substance'. It is awkward in two ways. His talk of 'representation' or 'presentation' (*Vorstellung*) introduces an epistemological consideration essentially irrelevant to the semantics of 'substance' or the ontology of substance. Secondly, Kant partly assimilates 'subject' in the sense of 'grammatical subject' and subject in the sense of the (paradigmatically) extra-linguistic denotation of a grammatical subject. Leaving aside both distractions, where 'F' ranges over properties, (1) may be construed as entailing the quasi-Aristotelian:

x is a substance iff $\forall F, \forall x \, \neg(x = F) \, \& \, \Diamond(\exists x) \, (\exists F) \, (Fx)$

(2) is a claim about the grammar of the first person singular pronoun: 'I' (that is, 'this representation of myself') may be used as the grammatical subject of any judgement I make but may not function as any kind of predicate. The first conjunct of (2) entails the rather tautological thought that any of my judgements is thought by me.

(3) identifies the denotation of the first person singular pronoun with a substance. (Even so, as Jonathan Bennett sees, 'soul' (*Seele*) carries no metaphysical or theological force at just this stage of the argument.) (3) is straightforwardly ontological and so does not plausibly follow from the linguistics of (2). Kant does not need as a premise:

(2)* 'I' is only usuable as a grammatical subject, never as a grammatical predicate.

but Kant does need:

(2)** I am not a property of anything but I do bear properties.

If we allow Kant (2)** instead of the grammatical claim (2) and allow the quasi-Aristotelian entailment from (1) then (3) validly follows:

x is a substance (def.) iff $\forall F, \forall x -(x = F) \ \& \ \Diamond(\exists x\ )\ (\exists F\ )\ (Fx) \ \&\ (\mathrm{I} = x)$

In the absence of an argument for (2)** it is less clear that the argument is sound. One could be forthcoming from the coherence of idealism, in particular the coherence of solipsism. One plausible explanation for the coherence of solipsism is that the self is a substance: It may be *coherently supposed* that only I exist because, in reality, I depend upon nothing.

Kant's Critique of the First Paralogism in the first edition.

Kant does not quarrel with the validity of the First Paralogism just as stated (at A 348) (despite his definition of 'paralogism' as 'fallacious argument') or even insist that its conclusion is false. His quarrel is with the attempt to derive metaphysical or theological conclusions from that Paralogism's conclusion as premise. He aims for the conclusion that it does not follow from the fact that I am a substance that I cannot begin to be or cease to be ('arise or perish' A 349). If his argument is sound, Cartesian dualism and the Platonic tenet that the soul can neither begin nor cease to be (because the same sorts of things that are immortal cannot originate) may not be soundly derived from the premise that the self is a substance. He is willing to grant the word 'substance' so long as nothing metaphysical or theological turns on it:

'The proposition, '*The soul is substance*', may, however, quite well be allowed to stand, if only it be recognised that this concept [of the soul as substance] does not carry us a single step further, and so cannot yield us any of the usual deductions of the pseudo-rational doctrine of the soul, as, for instance, the everlasting duration of the human soul in all changes and even in death -- if, that is to say, we recognise that this concept signifies a substance only in idea, not in reality.' (A 351)

Kant's argument is:

(1) '[In the analytical part of the Transcendental Logic we have shown that] pure categories, and among them that of substance, have in themselves no objective meaning, save in so far as they rest upon an intuition, and are applied to the manifold of this intuition, as functions of synthetic unity.' (A 348-9)

Wir haben in dem analytischen Teile der transzendentalen Logik gezeigt: daß reine Kategorien (und unter diesen auch die der Substanz) an sich selbst gar keine objektive Bedeutung haben, wo ihnen nicht eine Anschauung untergelegt ist, auf deren Mannigfaltiges sie, als Funktionen der synthetischen Einheit, angewandt werden können.

(2) ‘I can say of any and every thing that it is substance, in the sense that I distinguish it from mere predicates and determinations of things.’

Von jedem Dinge überhaupt kann ich sagen, es sei Substanz, so fern ich es von bloßen Prädikaten und Bestimmungen der Dinge unterscheide.

(3) ‘Now in all our thought the 'I' is the subject, in which thoughts inhere only as determinations; and this 'I' cannot be employed as the determination of another thing.’

Nun ist in allem unserem Denken das Ich das Subjekt, dem Gedanken nur als Bestimmungen inhärieren, und dieses Ich kann nicht als die Bestimmung eines anderen Dinges gebraucht werden.

(4) ‘Everyone must, therefore, necessarily regard himself as substance, and thought as [consisting] only [in] accidents of his being, determinations of his state.’

Also muss jedermann Sich selbst notwendiger Weise als die Substanz, das Denken aber nur als Akzidenzien seines Daseins und Bestimmungen seines Zustandes ansehen.

(5) ‘That I, as a thinking being, *persist* for myself, and do not in any natural manner either *arise* or *perish*, can by no means be deduced from it.’ (A 349)

Dass ich, als eine denkend Wesen, vor mich selbst fortdaure, natürlicher Weise weder entstehe noch vergehe, das kann ich daraus keinesweges schließen.

(1) entails that the categories have only an empirical use, as does Kant’s auxiliary claim: ‘In the absence of this manifold, they are merely functions of a judgment, without content.’ (A 349) We may read ‘meaning’ (*Bedeutung*) in ‘objective meaning’ to mean ‘referent’. If the categories are not used empirically then they do not refer to anything. If the categories only have an empirical use then they have no metaphysical or theological use. So, if ‘substance’ is a category then ‘substance’ has no metaphysical or theological use.

(2) is designed to show that there is nothing special, in particular, nothing metaphysical or theological about the use of ‘substance’. Paradigmatically, a substance is a physical object. A thing is a substance:

$(\forall x)$ (thing x) $\rightarrow$ (substance x)

A thing has properties but is not a property:

$\forall x$ (thing x) $\rightarrow$ $-(x = F)$ & $\Diamond(\exists x)$ $(\exists F)$ (Fx)

which, with the quasi-Aristotelian concept of substance:

x is a substance iff $\forall F, \forall x -(x = F) \,\&\, \Diamond(\exists x)\,(\exists F)\,(Fx)$

yields

$(\forall x)$ (thing x) $\rightarrow$ (substance x)

[For Kant's anti-metaphysical conclusion to go through Kant needs instead:

$(\forall x)$ (substance x) $\rightarrow$ (thing x)

because that leaves no room for substances that are not things.]

In (3) Kant concedes the second premise of the First Paralogism: 'I' cannot function as a predicate and I am not a property of anything. He cites the special case of thoughts as properties of the self but any property, however tenuous, would do for the argument. '[…] in all our thought' ascribes a certain pervasive psychological tendency to us. Whatever thoughts we think, 'I' functions only as a grammatical subject.

In (4) there is a slide from the 'all' of 'all our thought' in (3) to 'must' in 'Everyone must […] regard himself as substance […]'. This slide is unwarranted without extra premises. Something like: 'The only coherent and legitimate function of the first person singular pronoun is as subject and never as predicate' would help but is not enough. There still seem to yawn two insuperable gaps, one between having to use 'I' only in the subject place in thoughts and taking onself to be a substance, the other between taking oneself to be a substance and having to do that. The first gap is between grammar and ontology. The second gap is between universal quantification:

$(\forall x)$ (x is a thought of mine, I use 'I' only as subject expression)

and necessity:

$\Box(\forall x)$ (x is a thought of mine, I use 'I' only as subject expression)

Indeed, Hume and the Buddhists do not think of themselves as substances, even in this rather weak sense of 'substance'. *A fortiori*, Hume and the Buddhists are under no (psychological, logical or semantic) compulsion to think of themselves as substances. Nothing in Kant's argument allows him to bridge these two gaps. In particular, Kant's 'therefore' in (4) has no force.

(5) is Kant's answer to his own question at (A 349) 'But what use am I to make of this concept of a substance?'. Kant is ruling out two putative inferences from 'I am

a substance': 'I […] arise' and 'I […] perish'. He thinks a third inference significantly does go through: 'I […] persist'. What do these amount to? Suppose I 'arise' if and only if I begin (to be) and

I begin (to be) iff - $(\exists x)$ t1 & $(\exists x)$ t2 & (I = x)

Suppose I 'perish' iff and only if I cease (to be) and

I cease (to be) iff $(\exists x)$ t1 & - $(\exists x)$ t2 & (I = x)

Suppose

I persist iff $(\exists x)$ t1 & $(\exists x)$ t2 & (I = x)

To rule out intermittent existence, the definition of 'I persist' should be supplemented with (for example):

& $(\forall tx)$ (tx > t1 & tx < t2) $\rightarrow$ $(\exists x)$ tx & (I = x)

so then we have:

I persist iff $(\exists x)$ t1 & $(\exists x)$ t2 & (I = x) & $(\forall tx)$ (tx > t1 & tx < t2) $\rightarrow$ $(\exists x)$ tx & (I = x)

Kant is right that the two claims 'I do not arise' and 'I do not perish' may not be validly derived from the conclusion of the First Paralogism because

x is a substance (def.) iff $\forall F, \forall x$ –($x = F$) & $\Diamond(\exists x)$ $(\exists F)$ (Fx) & (I =x)

tells us nothing about the truth value of:

- $(\exists x)$ t1 & $(\exists x)$ t2 & (I = x)

or

$(\exists x)$ t1 & - $(\exists x)$ t2 & (I = x)

However, *pace* Kant, it does not entail

$(\exists x)$ t1 & $(\exists x)$ t2 & (I = x) & $(\forall tx)$ (tx > t1 & tx < t2) $\rightarrow$ $(\exists x)$ tx & (I = x)

without an extra assumption which would have to be as strong as: *If something exists it lasts*. Kant has no conclusive argument for such an assumption. For example, whatever force the arguments in the Transcendental Aesthetic, the Schematism and

the Analogies chapters for time as a necessary condition for empirical objects might have they do not rule out as incoherent a-temporal non-empirical objects. ‘x exists but has no duration’ entails no contradiction. (In particular, the fact that ‘exists’ is a present tense verb entails no ontological prohibition of timeless items. We just read it timelessly.) Because ‘I am a substance’ does not entail ‘I am temporal’ the conclusion of the First Paralogism leaves logical room for ‘I am timeless’, a component of Platonism and Cartesianism.

Kant’s apparent victory over Cartesianism is due to his weakening of the concept of substance. A substance is that which may bear properties but is not itself a property. Kant is quite happy with this weakening. He says

‘Yet there is no other use to which I can put the concept of the substantiality of my thinking subject, and apart from such use I could very well dispense with it.’

This is not right. On the Cartesian conception, the self is a substance not only in the sense of that which bears properties but in the sense of *that which depends upon nothing else for its existence*. In order to refute entailments to absence of natural origin and immortality Kant would have to either deny that the self is a substance in this sense or deny that these entailments go through from the soul’s being a substance in this sense. He does neither.

The Cartesian or Platonist might defend the absence of natural origin as follows:

(1) Whatever has a natural origin depends upon something natural.
(2) I, as substance, depend upon nothing natural.
(3) Therefore, I have no natural origin.

They might defend immortality as follows:

(1) Whatever admits of natural destruction depends upon something natural.
(2) I, as substance, depend upon nothing natural.
(3) Therefore, I do not admit of natural destruction.

Mindful of the empirical constraint on the meaningful employment of the categories, Kant entertains an alternative strategy for proving non-natural genesis and nemesis:

‘So far from being able to deduce these properties merely from the pure category of substance, we must, on the contrary, take our start from the permanence of an object given in experience as permanent.’

GERMAN

‘For only to such an object can the concept of *substance* be applied in a manner that is empirically serviceable.’

However, the First Paralogism does not contain any premises about experience so is of no use in the alternative strategy.

‘In the above proposition, however, we have not taken as our basis any experience; the inference is merely from the concept of the relation which all thought has to the 'I' as the common subject in which it inheres.’

Kant’s Humean Intuitions

Kant is confident that introspection is powerless to show that the self is a substance. He thinks that ‘the permanence of an object given in experience as permanent’ would be sufficient for the self’s being a substance. I interpret this criterion to be met if I am given to myself veridically as something persisting which can have no natural beginning and no natural end. Kant says:

‘Nor should we, in resting it upon experience, be able, by any sure observation, to demonstrate such permanence.’

Wir würden auch, wenn wir es gleich darauf anlegten, durch keine sichere Beobachtung eine solche Beharrlichkeit dartun können.

so, in his view, even if there were such a permanent self it could not be disclosed as such to introspection. The ground he provides for this is thoroughly Humean:

‘The 'I' is indeed in all thoughts, but there is not in this representation the least trace of intuition, distinguishing the 'I' from other objects of intuition.’

Denn das Ich ist zwar in allen gedanken; es ist aber mit dieser Vorstellung nicht die mindeste Anschauung verbunden, die es von anderen Gegenständen der Anschauung unterschiede.(A 350)

In Kantian vocabulary, there is no intuition of an ‘I’ distinct from whatever else I intuit. In Humean vocabulary, when I introspect I always happen upon some idea or other, none of which is sufficiently fixed and enduring to give rise to the idea of self. HUME

This Humean picture is widely endorsed. Its grip on subsequent philosophy is pervasive. Unfortunately, it is wholly and catastrophically mistaken.

Hume misses everything that matters in introspection: subjectivity, inner phenomenological space, that the time is always now, the felt tone of me-ness, the no-thing-ness of the interiority of immateriality, the uniqueness of being oneself. These features of one’s own existence (which are mysteriously not available through one’s perceptions of another) amount to the inner site or zone in which Humean ideas

and impressions arise and perish. Kant and Hume only see the coming and going of the content of the self. They miss the self entirely. Being it is an obstacle to their experiencing it. Modelling inner sense on outer sense they misconstrue the internal world on the model of the external world. This claim is therefore wholly false:

'We do not have, and cannot have, any knowledge whatsoever of any such subject.'

Die Erkenntnis des realen Subjekts der Inhärenz [...], von welchem wir nicht die mindeste Kenntnis haben, noch haben können.(A 350)

Kant reiterates that any of my thoughts could in principle be prefaced by 'I' and says we should not mistake this grammatical point for the experience of a permanent self:

'Thus we can indeed perceive that this representation is invariably present in all thought, but not that it is an abiding and continuing intuition, wherein the thoughts, as being transitory, give place to one another.'

Man kann also zwar wahrnehmen, dass diese Vorstellung bei allem Denken immer wiederum vorkömmt, nicht aber, dass es eine stehende und bleibende Anschauung sei, worin die Gedanken (als wandelbar) wechselten.

Kantian grammar and a Heraclitean ontology are not enough to solve the problems of the self. Something has remained utterly unchanging in one's own nature. As we shall see, this is a necessary condition for solving the problem of personal identity. But independently of that, it should be possible through a *Gestalt* switch to be Parmenidean rather than Heraclitean about the self. AUGUSTINE

Nevertheless, there is an ambivalence in Kant's rejection of Cartesianism. On the one hand he is indignant about the fictitous author of the Frist Paralogism trying to deceive us into endorsing Cartesianism based on linguistic premises:

'It follows, therefore, that the first syllogism of transcendental psychology, when it puts forward the constant logical subject of thought as being knowledge of the real subject in which the thought inheres, is palming off upon us what is a mere pretence of new insight.'

Hieraus folgt: dass der erste Vernunftschluß der transzendentalen Psychologie uns nur eine vermeintliche neue Einsicht aufhefte, indem er das beständige logische Subjekt des Denkens vor die Erkenntnis des realen Subjects der Inhärenz ausgibt

On the other hand, Kant glimpses that grammar and Heraclitus are not enough:

'Consciousness is, indeed, that which alone makes all representations to be thoughts, and in it, therefore, as the transcendental subject, all our perceptions must be found;

but beyond this logical meaning of the 'I', we have no knowledge of the subject in itself, which as substratum underlies this 'I', as it does all thoughts.'

Weil das Bewusstsein das einzige ist, was alle Vorstellungen zu Gedanken macht, und worin mithin alle unsere Wahrnehmungen, als dem transzendentalen Subjekte, müssen angetroffen werden, und wir, außer dieser logischen Bedeutung des Ich, keine Kenntnis von dem Subjekte an sich selbst haben, was diesem, so wie allen Gedanken, als Substratum zum Grunde liegt.

Three levels of explanation need separating here: ontological, logico-linguistic, and epistemological. Consciousness and the transcendental subject belong to ontology. The logical meaning of 'I' belongs to language and logic. That we might have no knowledge of the subject, no knowledge of what we are, is an epistemological claim. Now, if I last *something* lasts. Even if the possibility of the 'I think […]' accompanying any of my (re)presentations is logico-linguistic (or belongs to 'logical meaning') we may ask how this possibility is grounded ontologically. Suppose 'I think […]' accompanies my thought at t1, but not at t2 (It is a disposition that does *not* have to be perennially exercised). Suppose it accompanies my thought at t3. A necessary condition for this is that I last from t1 through t3 (or, at least, exist at both t1 and t3). This means that Kant's 'logical meaning' is grounded in what exists. Now, what thereby exists? Promisingly, Kant offers us the ontological theses 'Consciousness is […] that which alone makes all representations into thoughts' and 'in it […] all our perceptions must be found'. In a moment of metaphysical temptation and unease he identifies consciousness with the transcendental subject. He talks about consciousness 'as the transcendental subject'. Instantly he recants. The transcendental subject is only the transcendental unity of consciousness bestowed by the 'I think […]' possibility. This, predictably, is to be understood as the 'logical meaning of the 'I'. Kant's metaphysical unease was justified. He was right to be tempted and should have given in to temptation. All our perceptions must be found in consciousness. The Heraclitean self is grounded in the Parmidean self which Kant, in this passage, half concedes. The Humean and the Buddhist arises and subsides within the Augustinian or Cartesian.

Kant is even willing to concede both the dependence of the Heraclitean on the permanent, and the linguistic on the ontological, so long as 'we have no knowledge of the subject in itself'. He is right to concede the dependency but wrong to be pessimistic about the epistemology. By way of entailment, if not always overt characterization, we learn a great deal in the *Critique of Pure Reason* about the self as thing-in-itself: I am it. It is not physical. We know that because if it were physical it could be subsumed under the forms of intuition and the categories and would not be unknowable. It is unknowable, therefore it is not physical. It is free. We know this explicitly from the resolution of the Third Antinomy. It is individuated. Kant (unlike Schopenhauer) talks about 'things-in-themselves' in the plural. How deeply Cartesian this whole picture suddenly looks. After all, Descartes himself thought the

soul did not directly experience itself as itself, only its operations (*pensées*). Kant is a Cartesian *malgre lui*. Kant is logically committed to the metaphysics he would eschew.

Kant's Critique of the The First Paralogism in the Second Edition

Kant does not restate the First Paralogism at CPR 369, B407, only the putative refutation of it which may be extracted as this argument:

(1) 'In all judgements I am the determining subject of that relation which constitutes the judgement.'

In allen Urteilen bin ich nun immer das bestimmende Subjekt desjenigen Verhältnisses, welches das Urteil ausmacht.

(2) 'That the 'I', the 'I' that thinks, can be regarded always as subject and as something which does not belong to the thought as a mere predicate must be granted'

Daß aber Ich, der ich denke, im Denken immer als Subjekt, und als etwas, was nicht bloß wie Prädikate dem Denken anhänge, betrachtet werden kann […]

(3) 'It is an apoplectic and indeed identical proposition; but it does not mean that I, as object, am for myself a self-subsistent being or substance'

[…] gelten müsse, ist ein apodiktischer und selbst identischer Satz, aber er bedeutet nicht, dass ich, als Objekt, ein, fpr mich, selbst bestehendes Wesen, oder Substanz sei.

Although Kant does not employ the expression 'propositional attitude', he has the concept of a judgement as a propositional attitude. He thinks that if I judge that *p* then I am thereby psychologically related to *p*, and it is at least true of me that I believe that *p*. Premise (1) says, 'I' refers to the judge, or whoever makes the judgement.

(2) allows to the rational psychologist the grammatical point that 'I' has no adjectival use but is always used in the subject position in sentences used to make first person singular ascriptions. 'It' in (3) refers back to (2), so is used to assert that the grammatical fact about 'I' is a necessary truth. When Kant says it is an 'identical proposition' he implies it is not only necessary but analytic. This is because it is part of the definition of 'I' to possess the subject role in sentences.

The rest of (3) denies the validity of a certain inference, from:

'I' has only a subject role

to:

I, the formulater of judgements in which 'I' features, am a substance.

Kant is denying that grammatical subjects necessarily denote substances, (rather as it might be misleading to assume that adjectival expressions always designate properties) and and denying *a fortiori* that ‘I’ used as a grammatical subject entails any denotation of a substance. Kant refuses to draw an ontological conclusion from only linguistic premises.

Kant is incorrect in this thought because it does follow from the fact that the first person singular pronoun has the grammatical function it does that the self *qua* user of that pronoun, is a substance in several senses. The subject of a sentence does denote a substance in the sense of that which bears properties but is not itself a property. It also denotes a substance in the minimal sense of a 'self-subsistent being', that is, a being that endures long enough for the ascription or misascription of properties in the use the predicate. ‘I’ denotes a substance in one of the Aristotelian senses of ‘substance’ to mean 'that which exists independently' or 'that which is not logically dependent on something other than itself for its existence'. (CPR p. 371-2) To see this, it is necessary to spell out the presuppositions of being an ‘I’ user in the full sense: being conscious, being someone (in Nagel’s sense), being conscious of being the being one is, not being what one is not, possessing a psychological interiority. It is implausible to suggest that these metaphysical properties could be possessed by anything physical and their only plausible bearer is the Augustinian or Cartesian subject or soul inadvertantly entailed by Kant’s critique of Rational Psychology.

II

Simple Souls: Kant and the Unity of Consciousness

In the so called Second Paralogism of the *Critique of Pure Reason* Kant is concerned to repudiate the idea that the self possesses one of the essential properties of the soul: simplicity.

I shall understand by 'simplicity':

x is simple if and only if (*def.*) *x* is not even in principle divisible.

So 'simple' here means 'logically simple'. Kant's repudiation takes the form of a critique of the following argument deployed by 'the rational psychologist' as a putative proof of the simplicity of the self:

(1) 'That, the action of which can never be regarded as the concurrence of several things acting is simple.' (CPR 335, A 351)(1)

Dasjenige Ding, dessen Handlung niemals als die Konkurenz vieler handelnden Dinge angesehen werden kann, ist einfach.

(2) 'Now the soul, or the thinking 'I' is such a being!

Nun ist die Seele, oder das denkende Ich, ein solches:

(3) 'Therefore etc.'

Also etc.

I shall comment on this argument and then comment on Kant's comments on it.

The first premise (1) is a definition of 'simplicity'. As I shall read it, it entails the idea of logical simplicity offered above. (1) uses a causal criterion which restricts what is simple to what has causal efficacy. It asserts that some putative entity, *x*, is simple if and only if its effects may never be rightly regarded as issuing from more than one cause. I take it this precludes one of two *prima facie* logical possibilities. Firstly, if A is putatively the cause of B, then it might turn out that B is in fact the result of several causes amongst which A features. This is one way of taking 'several things acting'. Secondly, it might turn out that A is indeed the cause of B, but by virtue of the causal efficacy of certain or all the parts of A. Then 'A' would be a kind of generic term for what could equally correctly be thought of as a set of causes of B. On the first interpretation there are causes of B other than A, on the second A is a plurality of causes of B. (1) is designed to preclude the second possibility.

We have to decide now what the force of 'cannot be regarded as' is. If A is the cause of B, but not in virtue of the operations of any parts of A, it does not follow that the only reason why A has to be thought of in this way is that A is like that. There might exist some theoretical limitation, or imaginative constraint on persons making A intelligible which precludes their thinking of A in any other way.

Suppose we strengthen Kant's 'cannot' so that it is a logical 'cannot'. We then have the claim that A cannot, even in principle, be regarded as made up of parts. Why might this be? It would not seem to follow from 'A is not made up of parts' that 'A cannot (logically) be thought of as made up of parts'. That would seem to make a certain sort of mistake about A's nature not just impossible but logically impossible, and it would require a great deal of additional argument to show that it was logically impossible for someone to be mistaken about something. Suppose we supplement

the account still further so that we have: 'It is logically impossible for a person to correctly think of A as made up of parts'. Then, if that is true, 'A is simple' might follow with greater plausibility.

To decide that, we need to know two things: whether 'A is simple or A is composite' is exhaustively disjunctive (whether if one conjunct is true then the other is false), or whether there might exist a third possibility. Secondly we need to know whether it is always true that if it is logically impossible to correctly think of *x* as *F* then this must always be because *x* is not *F*, ie. we need to know whether it is logically possible that there should be reasons other than *x* 's being not F why it should be logically impossible to think of *x* as *F* correctly.

On the first of these, if 'A is simple and A is composite' is a contradiction then 'A is simple or A is composite' is exhaustively disjunctive. 'Simple' means 'not composite' and 'composite' means 'not simple' so the conjunction is a contradiction, so the distinction is exclusive. The disjunction is also exclusive, because if it is true of A that A is simple then A is not composite and *vice versa*. 'A is simple' is either true or false, and the claim 'A is composite' is either true or false, so 'A is simple or A is composite' is exhaustively disjunctive.

On the second, it is not impossible that there should exist reasons other than A's not being *F* why it is impossible to correctly think that A is *F*. The thought that such reasons should exist does not contain a contradiction. What is clearly the case is that if A is not *F* then it is logically impossible to correctly think of A as *F*. But it is the first and not the second of these thoughts that is needed for the first premise of the Second Paralogism. If the only possible reason why A's action 'can never be (correctly) regarded as the concurrence of several things' is that it is not a concurrence of several things then it will follow that A is simple and the rational psychologist's definition goes through. But if there are reasons other than A's not being the concurrence of several things why it is (logically) impossible to (correctly) think of A as the occurrence of several things than the definition does not go through, because clearly, A might then be the concurrence of several things and so not be simple.

In addition to these considerations, it is not at all evident that someone who thought a simple object composite would thereby have produced a contradiction, and that seems to be needed if it is to be logically impossible to think of A as *F* just because A is not *F*. But the thought 'A is *F* ' is not contradictory if A is not *F*: just false. Obviously 'A is *F* and A is not *F* ' is contradictory but that thought is of no use to Kant's attack on the rational psychologist.

It might have been more satisfactory if the first premise of the First Paralogism had not been formulated using the expression 'can never be regarded as' with simply with 'is not'. Then it would have amounted to a straightforward enough definition of 'simple'.

The second premise, (2), is the claim that the soul falls under the description 'its action can never be regarded as the concurrence of several things acting'. We are to imagine 'the soul' or 'the soul's action' as substituted in the first premise for 'That the

action of which', but 'soul' here is ambiguous. It could just mean 'self', or something like 'that which thinks'. This latter is suggested by the clause which qualifies 'soul' with 'or the thinking I'. There are difficulties in taking either reading.

If 'soul' just means 'self' here then no reasons have been adduced, no premises supplied within the argument, for the claim that the soul falls under the description in (1). It is an unargued assumption. If, on the other hand, 'soul' means 'immaterial substance' then the soul does fall under the description in (1). But this is of little use to Kant's attack on the rational psychologist because it goes towards proving that souls exist. If we accept that in some sense there is a subject of experience, something that thinks, we are not thereby *obviously* committed to the view that the subject is a soul. And if we accept the rationalist definition of 'soul' we are not thereby committed to the existence of souls. Kant thinks this equivocation is fatal to the persuasive power of the second premise: It is either putatively informative, but unsubstantiated, or else vacuously true but uninformative.

The conclusion (3) 'therefore etc' should read 'Therefore the soul or the thinking 'I' is simple'. Clearly the argument of the Second Paralogism is valid: The conclusion follows from the two premises on either of the two interpretations of the second premise. But the interesting question is: Is it sound?

The truth of 'The soul is simple' cannot be decided if we take the informative reading of the second premise, only if we take the uninformative one, which gives us 'the soul is simple' as true, but just as true by definition. So soundness seems to be obtained in the case of the Second Paralogism at the price of vacuousness.

The best strategy for the rational psychologist is to try to adduce reasons for the truth of the second premise interpreted non-analytically. Otherwise, because the first premise is a definition, the threat is that the conclusion will be vacuous because derived from vacuous premises. Clearly such reasons could be adduced, perhaps based on the unintelligiblity of an ontologically divided subject, but I leave consideration of this possibility until later.

Kant on the Second Paralogism

At CPR 335, A352, before criticising the Second Paralogism, Kant reconstructs an argument that the rational psychologist might adduce as a set of reasons for the truth of the second premise, interpreted informatively. I call this the Argument for the Simplicity of the Self:

(1) 'An effect which arises from the concurrence of many acting substances is indeed possible, namely, when this effect is external only.'

Nun ist zwar eine Wirkung, die aus der Konkurrenz vieler handelnden Substanzen entspringt, möglich, wenn diese Wirkung bloß aüßerlich ist.

(2) 'Suppose it be the composite that thinks, then every part of it would be part of the

thought.'

Denn, setzet, das Zusammengesetzte dächte: so würde ein jeder Teil desselben einen Teil des Gedanken [enthalten]

(3) 'Only all of them taken together would contain the whole thought.'

Alle aber zusammengenommen allererst den ganzen Gedanken enthalten.

(4) 'But this (2) and (3) cannot be consistently be maintained.'

Nun ist dieses aber widersprechend.

(5) 'Representations (for instance the single words of a verse) distributed among different beings, never make up a whole thought (a verse).'

Denn, weil die Vorstellungen, die unter verschiedenen Wesen verteilt sind, (z.B. die einzelne Wörter eines Verses) niemals einen Gedanken (einen Vers) ausmachen.

(7C') 'It is therefore impossible that a thought should inhere in what is essentially composite.'

So kann der Gedanke nicht einem Zusammengesetzten, als einem solchen, inhärieren.

(8C") 'It is therefore only in a single substance, which not being an aggregate of many, is absolutely simple.' (CPR 335, A352)

Er ist also nur in einer Substanz möglich, die nicht ein Aggregat von vielen, mithin schlechterdings einfach ist.

We know that the cause of an 'effect which arises from the concurrence of many acting substances' (as described in (1)) is composite. This is because Kant defines 'composite' at CPR 335, A351 not only as 'an aggregate of several substances' but also uses the causal criterion: 'the action of a composite [[...]] is an aggregate of several actions or accidents, distributed amongst the plurality of substances'. (1) is designed to show that this is not just an empty definition but that there do exist some entities that are not simple but genuinely composite. Kant thinks, for example, that the motion of a physical object is the motion of the parts of that object. There is an equivocation here between 'is' and 'is caused by' which is significant for what follows. The point of the example is to contrast those conditions under which it is correct and legitimate to think in terms of composites with those under which it is not. The contrast is marked by Kant by his use of 'external', which presumably, but

not explicitly, is contrasted with 'internal'. So 'external' causes may be composite but an 'internal' cause must be simple. This corresponds to a subjective-objective distinction, because if *x* is 'internal' then *x* pertains only to the psychology of the subject. There seems little need to quarrel with in the first premise, if we accept that any physical object is in principle infinitely divisible. If we reject that (say, because of a limitation of Newtonian atomism) we can still allow that a physical object has parts and so is in principle divisible but not infinitely so. If so then physical objects are obviously not simple even if they have some simple parts. .

(2), together with (3) and (4) have the form of a *reductio ad absurdum*. Kant wants to prove that 'With thoughts, as internal accidents belonging to a thinking being it is different (from (1))', ie, that thoughts are not composite. There is a glaring ambiguity in what is to be proved here because the proof that thoughts are (each) simple is not the same as the proof that that which thinks those thoughts is itself simple. This ambiguity runs throughout the argument, but the first part of (2) is clearly concerned with what thinks. In keeping with the *reductio* format, Kant assumes in (2) the contradictory of that which he seeks to prove, and the whole of (2) expresses the thought that if what thinks is composite then what is thought would itself be composite, and in a sense, what is thought would be identical, in every part, with what thinks. The difficulty here is just that which arises in the case of the efficacy of a physical object in motion. Compare:

(a) A physical object's motion is caused by the motion of its parts.

and

(b) A physical object's motion is the motion of its parts.

with

(a') Thoughts are caused by the parts of that which thinks.

and

(b') Thoughts are the parts of that which thinks.

But the relation expressed by *a=b* is not the same relation as that expressed by '*a* causes *b*', indeed, *prima facie*, the truth of '*a* causes *b*' would seem to presuppose that '*a=b*' is false, and the truth of '*a=b*' would seem to presuppose '*a* causes *b*' is false. If so, then (a) and (b) are mutually inconsistent, and (a') and (b') are mutually inconsistent.

But it might be that which relations are causal relations, and which relations relations of identity is a matter of conceptual stipulation. In the two examples above, the decision might be partly governed by the assumed criteria for the individuation of

physical objects and subjects of thought respectively. If so, then some of the tension between (a) and (b) and (a') and (b') is relieved, because each member of each pair is then an alternative description, and there are then no strong ontological grounds for saying their disjunction has to be exclusive.

The notion of simplicity and composition discussed in the first premise of the Second Paralogism, and in the first premise of the argument for the simplicity of the self are established by a causal criterion which presuposes a distinction between an object and its effects such as that expressed by (a) and (a'), but the second premise of the argument for the simplicity of the self assumes no ontological distinction between a (mental) object and its effects.

There are at least two ways of taking (2), and these need to be separated out before we can decide whether (2) is true. If we endorse the supposition expressed by the first clause of (2) then we assume that the subject of a thought, the thinker, is not simple but *qua* thinker, is made up of parts. Then, the second clause of (2) has it as a logical consequence of the thinker's composition that every part of the thinker is a 'part of the thought'. This is what (2) means if 'it' in the second clause refers back to 'the composite that thinks' in the first clause. The grammar of (2) does suggest this.

Now, there is a sense in which this claim is highly implausible. If there is a putative thinker 'S' and a putative thought '*m*' and *m* is thought by S, then it does not really make sense to say that every part of S is a part of *m*. This is because S's being a part of *m* is inconsistent with S's thinking *m*, but that is what is assumed in (2)'s first clause. Further, suppose S thinks *m* but also previously thought thoughts *k*, *l*, and subsequently thought thoughts n, o, p, then presumably every part of S is putatively a part of each of *k*,*l*,*n*,*o*, and *p* as well as *m*. But if A is wholly a part of B it is impossible for A to be wholly a part of C unless B is a part of C. But in this case, clearly the thoughts *k*,*l*,*n*,*o*, are not themselves parts of *m* because they exist at different times, and their all being parts of a putative thought series is not sufficient for any one of them being part of any other. So, if we read (2) literally, (2) contains this inconsistency and so is false.

To free (2) of the inconsistency we have to weaken the second occurrence of 'part of' so that it reads 'would be a part of what thinks the thought'. Then we obtain the claim that if a composite self thinks then each part of the composite thinker thinks a part of a (composite) thought. So, although the thinker as a whole thinks the whole thought, no single part of the thinker thinks the whole thought, and the whole of the thinker does not think a single part of the thought, because there are some parts of the thinker which do not think some parts of the thought. We may assume this if there is a one-one maping between thinker-parts and thought-parts. This is a much more plausible way of reading (2) even though it requires weakening 'part-of'.

Then we can take (3)'s 'all of them' to refer back to 'each part of the thinker'. Then (3) follows from (2), because if it is true that each part of the thinker thinks a part of the thought, then each part of the thinker considered together as a set, or series, thinks the whole thought (assuming still the part-part relation between thinker and thought is one-one).

Notice that (3) is phrased using 'contain', so what it literally expresses is that the parts of the thinker ('them') 'contain the whole thought'. This suggests a strong reason for not taking 'contain' literally here, and confirms the correctness of not taking 'part of' literally in its second occurrence in (2). This is because if A contains B then B is a part of A. But, (2), on the first interpretation, makes the thinker part of the thought.

Now, (3), inconsistently with that, makes the thought part of the thinker. Clearly, if A is a part of B it cannot also be true that B is a part of A because of the logic of 'part of'. But then it is clearly false that it can both be the case that the thinker is part of the thought and that the thought is a part of the thinker, where, evidently, 'thinker' means 'same thinker' and 'thought' means 'same 'thought'.

The most sympathetic construal, that on which Kant makes most sense, is: Each part of the thinker thinks a part of the thought. Then it is analytic that the totality of the parts of the thinker that think the thought think the totality of the thought. Clearly this is not quite the same as saying that the whole of the thinker thinks the whole of the thought. That might not leave room for a thinker thinking more than one thought if the one-one thinker part/thought part relation entails 'one and only one' thinker part.

(4) is the proposition that (2) and (3) are mutually inconsistent. We have already noted one inconsistency between (2) and (3), but Kant has mind mind a different one here. Suppose some part of S called 'S1' thinks the part of some thought '*m*' called '*ml*' at at a time t^1, S2 thinks m2 at t^2, etc, so we can read (3) here, (the claim that only all of the parts of the thinker taken together think the whole thought) to mean that only, say, S1… S4 inclusive may be truly said to think the whole 'm' composed of, say, ml… m4 inclusive. This thinking of ml… m4 takes place over $t^1 \ldots t_4$. (4) entails that this model cannot be a coherent account of the thought - thinker relation. The reasons for this are partly to do with the individuation of thinkers and partly to do with the individuation of thoughts. They are contained not in (4) but in (5) so we should examine that next.

In reading (5) we may allow ml…m4 inclusive to be a putative (re)presentation, in Kant's example, a verse. Then S1…S4, inclusive, are 'different beings', meaning at least that each one of S1…S4 is numerically distinct from every other. Putatively S1…S4 are constitutive of a thinker over a time $t^1 \ldots t4$.

Then (5) is partly the claim that m1…m4 do not jointly make up a thought, and S1…S4 do not jointly make a up a thinker. The implication is that no criterion has been provided for m1…m4's counting as parts of a single thought, nor for S1…S4's counting as parts of a single thinker.

There are ways in which Kant's criticisms of the model may be made explicit. For example, it is merely stipulative within the model that S1…S4 are constitutive of one thinker over $t^1 \ldots t4$, but there is no *a priori* obstacle to either of the following alternative interpretations.

Firstly, S1, S2, S3, and S4 are each numerically distinct thinkers and not parts of one and the same thinker. Each would then last as long as each of $t^1 \ldots t4$, and would think only one of the sequence ml, m2, m3, m4 during each of those times.

Secondly, any member of the set S1…S4 could, in principle, be arbitrary combined with any other member of S1…S4 to constitute a single thinker and, clearly, any such set might form a subset of S1…S4, and not be the set S1…S4.

Next, the putative thought parts m1…m4 are constitutive of a whole thought on the model, but again, this is merely stipulative. Nothing prevents *a priori* m1…m4 being numerically distinct thoughts, whether S1…S4 constitute a unitary thinker or are numerically distinct thinkers. So the model does not rest on criteria for the individuation of thoughts and thinkers.

One further way in which Kant's criticism may be expanded is illustrated by this diagram;

TIME	THINKER PARTS	'thinks	THOUGHT PARTS	relation'
t^1	S1	S1'	ml	ml'
t^2	S2	S2'	m2	m2'
t3	S3	S3'	m3	m3'
t4	S4	S4'	m4	m4'
t5	S5	S5'	m5	m5'
t6	S6	S6'	m6	m6'

Here there exist qualitatively similar putative thinker parts simultaneously (for example, S1 and S1') and there exist qualitatively similar putative thought parts simultaneously (for example *m*1 and *m*l'), and, further, a qualitatively similar pair of thinker parts exists simultaneously with a qualitatively similar pair of thought-parts (for example, S1, S1', *m*l and *m*l' and t^1).

Now, there is nothing so far on the model to preclude S1 and S1' from being parts of one and the same thinker at t^1 because it is not logically impossible, if a thinker is composite, that a thinker be (partly) composed of numerically distinct yet qualitatively identical parts at a single time. Similarly, there is nothing which prevents *a priori* ml and ml' being numerically distinct yet qualitatively identical parts of one and the same thought. Whether these two logical possibilities are actual on the model is wholly undetermined.

This shows in a different way that the model includes no criterion for the individuation of thoughts and thinkers, this time by showing there is no way of specifying which parts of thinkers belong to which thinkers, or which parts of thoughts belong to which thoughts. There is nothing *a priori* to preclude the existence of a thinker who is constituted by, say, all and only S2, S3', S4' and S5', or a thought which is made up of, say, *m*3, *m*3' and *m*4 (or, I should say, S1 and S6', and *m*l and *m*6).

Equally problematic is the 'thinks' relation. A similar interpretation is that S1 thinks *m*l, S1' thinks *m*l' etc. But nothing precludes S1 thinking *m*' or S1' thinking *m*l. The model incorporates no conceptual apparatus for linking up specific thoughts and thinkers.

Suppose, finally, one wished to argue that S1…S6 constitute a thinker, and S1'…S1' constitute a numerically distinct thinker. It is clearly not sufficient for this that S1…S6 is a set of thinker parts each of which is numerically distinct from each of the thinker parts S1'…S6', even partly constituted by two numerically distinct but qualitatively identical sets of simultaneously existing thinker-parts. But it is necessary because if a thinker is numerically distinct from some second thinker then every part of the first thinker must be numerically distinct from every part of the second thinker. If we wish to give up this assumption, say to allow fusion and fission, then nothing in the model protects it. (I consider giving up the assumption below).

(5), in the sense that Kant intends it, is a necessary truth. This is because in the only other sense in which it could be plausibly taken it is clearly false. It is false if it means that different beings, numerically distinct thinkers, could never each think a part of a whole verse (representation), where the criterion for individuating verses makes no mention of a thinker, say, if verses are individuated by their semantic, syntactic or poetic properties. Then obviously several thinkers could think the separate parts (say lines) of what is a single verse. We could imagine a poetry recital where this was the convention.

So it is best to read 'a whole thought' in (5) as meaning 'a whole thought thought by a single thinker'. Then it is a necessary condition for the putative parts of a thought being parts of a whole thought that there is just one thinker of those parts of that thought.

In just that sense it is not a sufficient condition of some putative parts of a verse being the parts of a whole verse that any purely semantic, syntactic or poetic criterion be met, even though that may still be allowed as a necessary condition.

So reading (5) as necessary, the new requirement that the verse have one thinker and only one must also be met. (In (5) we may read 'representations' as '(putative) thought parts'. Then several representations may constitute one whole thought).

If we allow Kant the 'one thinker' criterion for the individuation of thoughts then conclusion (7C') follows, from the premises as interpreted. Clearly then, if it follows that 7C' it follows that 8C" because if it is true that *x* is not composite then it follows that *x* is simple, if *x* is the sort of individual which may be coherently characterised using those predicates.

Kant on the Argument for the Simplicity of the Self

Kant rightly identifies (5) as the crucial premise and offers two reasons for doubting it. One of these I shall call the *a priori* objection, the other the empirical objection. (They are located at CPR 335-6, A352-355).

The *a priori* Objection

Kant thinks it illegitimate, or at least not to the rational psychologist's purpose, to treat (5) as analytic, and clearly its putative necessity does derive from its putative analytically:

'The proposition "A thought can only be the effect of the absolute unity of the thinking being', cannot be treated as analytic.'(CPR 336, A353)

Der Satz: Ein Gedanke kann nur die Wirkung der absoluten Einheit des denkenden Wesens sein, kann nicht als analytisch behandelt werden.

It can be treated as a conceptual truth if we adopt the 'thinker' criterion for the individuation of thoughts but if the rational psychologist makes this stipulation his conclusion begs the question. Kant opts for any criterion for the individuation of thoughts that does not make 'same thinker' a condition for putative thought parts being parts of a whole thought. There is some merit in this, both because it accords with our pre-philosophical intuitions about what, say, counts as a whole verse, and because to insist on the unity of the thinker when thinking as a sufficient and necessary condition of the unity a putative thought would yield two undesirable consequences: Semantically and syntactically incoherent thoughts would count as complete if thought by a single thinker. Semantically and syntactically coherent thoughts would count as incomplete if thought by more than one thinker.

But the main force of Kant's criticism is this: No ontological conclusions may be logically derived from any number of conceptual truths. This is why he says 'No one [..] can prove this proposition [5] from concepts.' (CPR 336, A352) Diesen Satz aber kann niemand aus Begriffen beweisen.

The *a priori* objection is effective because Kant shows that two sorts of possibility exist: A single thought (verse) may be thought by one unified self, or by several. It is obviously not a necessary condition for there being verses or thoughts in general, individuated semantically and syntactically, that each be thought by a separate thinker.

The Empirical Objection

The empirical objection divides into two claims: that 'the concept of absolute unity is quite outside its [experience's] province', “der Begriff der absoluten Einheit [ist] weit über ihre [d.i. Erfahrung] Spähre” and that 'experience yields us no knowledge of necessity.' Diese gibt keine Notwendigkeit zu erkennen”(CPR 336, A353)

There are three reasons why a putative soul substance should not be available to empirical observation.

Firstly, it is sometimes maintained that *qua* subject of experience a self, that which experiences, may not be an object of its own experience.

Secondly, the soul is allegedly simple and so not even in principle divisible. This is the point Kant captures by 'absolute unity' here. Clearly the sentence '*x* is not

even in principle divisible' could never be conclusively verified empirically, though it could be conclusively falsified empirically. In the case of the soul in particular, both verification and falsification conditions would seem to be lacking, unless fission 'experienced from the inside' is thinkable.

Finally, the soul is putatively a non-material object and it is a plausible assumption that only objects with at least some physical properties may be detected empirically. These, taken together, amount to good reasons for pessimism about any empirical proof of the simplicity of the self.

The second part of the empirical objection is less satisfactory. We may allow Kant that experience yields no knowledge of necessity as he intends it, but it is not clear in what sense the rational psychologist is dealing in necessities. Two sense of 'necessity' are relevant to deciding this.

It is part of the definition of 'soul' that the soul is simple, so sentences like 'the soul is simple', 'the soul is non-divisible' and so on have the status of necessary truths. That is one kind of necessity the rational psychologist tries to establish, but it is obviously of no ontological significance is these sentences are analytic.

Secondly, there is one of Kant's senses of 'necessity': That certain sentences are true is a necessary condition for experience. In particular, that the sentences of transcendental idealism are true is a necessary condition for the experience of free, rational, self-conscious beings. Here 'necessity' does not have the sense of 'logical necessity' but 'irrefutable by experience'. Clearly, if *p* is necessary for experience then experience is sufficient for *p*. Then, nothing about experience could refute *p*.

Whichever way Kant reads the rational psychologist here, it is an essential part of his own critical philosophy that neither of these sorts of necessity may be established empirically. So, if we allow Kant that, and the rational psychologist is engaged in producing either or both sorts of necessity, then he cannot successfully derive them from purely empirical premises.

The Subjective 'I'

The *a priori* objection is designed to show the impossibility of a certain kind of *a priori* ontology: an ontology of the self. The empirical objection shows that there cannot be any non-empirical ontology of the self which would rest on empirical premises (Clearly Kant thinks in a sense there can be an empirical study of the self: psychology, or 'anthropology' as he calls it).

But there is room between these two impossibilities and this possibility for a genuine philosophical statement about the unity of the self, and this is an utterly formal or conceptual statement. To use an analogy, there exists an indivisible subjective self in roughly the sense that there exist indivisible points in mathematics. Formal truths may be formulated about each but if we ask ontological questions about them then the only true answers will be empirical ones, sentences about empirical persons and empirical points on paper. Kant has an argument to convince us of the formal nature of the unity of the self:

The Argument for the Formal Unity of the Self

(1) 'If I wish to represent to myself a thinking being, I must put myself in his place, and thus substitute as it were, my own subject for the object I am seeking to consider (which does not occur in any other kind of investigation).'

Es ist offenbar: dass, wenn man sich ein denkend Wesen vorstellen will, man sich selbst an seine Stelle setzen, und also dem Objekte, welches man erwägen wollte, sein eigenes Subjekt unterschieben müsse [welches in keiner anderen Art der Nachforschung der Fall ist]

(2) 'We demand the absolute unity of the subject of a thought.'

Und dass wir nur darum absolute Einheit des Subjekts zu einem Gedanken erfodern

(3) 'Otherwise we could not say 'I think".'

Weil sonst nicht gesagt werden könnte: ich denke.

(4) 'It is this 'I' that we presuppose in all thinking.'

Und dieses Ich setzen wir doch bei allem Denken voraus.

(5) 'Although the whole of the thought could be divided and distributed among many subjects the subjective 'I' can never be thus divided.' (CPR 336, A353-4)

Denn obgleich das Ganze des Gedanken geteilt und unter viele Subjekte verteilt werden könnte, so kann doch das subjektive Ich nicht geteilt werden.

(1) is the claim that the thought of a thinking being *qua* thinking being requires the ability to imagine being that being. (We should read 'conscious' for 'thinking' here). The assumption is that we learn what it is like to be conscious, and what it is like to be a being, each from our own case. If we each wish to have the thought of other conscious beings, for example other people, then a necessary condition for this is that we each possess the capacity for empathy with those beings.

When Kant speaks of substituting his own subject for the object of another thinking being he mean 'substitute in imagination'. We each have direct acquaintance only with our own subjectivity so to think of another person as possessing subjective properties requires that one, partly at least, model one's thought of the other on oneself. Kant thinks this does not occur in any other kind of investigation because he assumes that only conscious ('thinking') beings possess subjectivity, and only in their case is there is something it consists in to be one.

Premise (2) marks a departure from the question: Does a unified thought presuppose a unified thinker? and introduces this new question: Is the individual subject of experience divisible? (2) raises this issue at least by reporting our pre-philosophical intuition that the subject is simple, and perhaps also by expressing the view that we could not find ourselves intelligible unless we regarded ourselves as non-composite.

Whichever way we read (2), there is merit in the view. It is difficult to makes sense of the idea of subjective fission: that I should become two numerically distinct sources of thought and experience, each of which is identical with me. We can perhaps make sense of two subjective sources of consciousness occupying one body, and even make sense of 'they are both person *x*' so long as we do not substitute 'me' for 'person *x* '. The difficulty is in thinking of oneself as two subjective conscious selves. Being it is an obstacle to thinking its bifurcation. Why this should be so is part of the unsolved metaphysical problem of why there should obtain the asymmetry between the being who you are and all the remainder who you are not.

Kant says our belief in the absolute unity of the subject, and the fact that thinking of other minds requires extrapolation from one's own case, are 'obvious'. (CPR 336, 353) I take it, therefore, he does not wish to imply anything contentious by these claims, rather he takes them to be commonsensical. I shall take up subjective fission again later.

(3) points to a consequence of denying (2): If we did not believe in the unity of the subject we could not use 'I'. The implication is that we do use 'I' so we do believe in the unity of the subject. But what sort of unity of the subject is presupposed by the use of 'I', and what sort of presupposition obtains between the two?

Kant thinks the unity involved is a purely formal unity. He thinks it is not required that there exist a metaphysical subject, a soul substance, in order for 'I' to have a use. 'I' is the word that each human being uses to refer only to him or herself. This empirical fact about the function of 'I' requires minimally that there be self-reference, that user and referent be numerically identical, and that the ‘I’ user be an individual, and that the 'I' user know that he or she is an individual. Indeed, Kant thinks the only sort of unity of the self presupposed by the use of 'I' is that expressed by the transcendental unity of apperception: That I possess a disposition to self-consciousness is a necessary condition of my using 'I', and the statement 'I think my thoughts' is analytic. It is the conjunction of these two sentences which is expressed elliptically by (4), because the 'I' there is the 'I' of the transcendental unity of apperception. When Kant says the 'I' is presupposed in all thinking he means that any thought by a rational self-conscious being must in principle be capable of being couched in the grammatical first person singular. This entails that it makes no grammatical sense to try to postulate thought without a thinker. What does not follow from this according to Kant is that that thinker is metaphysically indivisible, only that the formal requirements of pure apperception, of the use of 'I', are met.

If premises (2), (3), and (4) are read, as they must be to have any plausibility, as an elliptical reiteration of the transcendental unity of apperception doctrine then given the second part of (5), the argument for the formal unity of the self is valid.

(5) entails that it is logically and practically possible for each of the constituents of a thought to be thought by a numerically distinct subject, but the 'subjective 'I' which each of those subjects has not similarly divisible. This subjective 'I' is however a purely formal property of the subject, not a metaphysical one.

Not only does Kant think the only non-empirical self is the 'I' of pure apperception, but he also holds the thesis that the metaphysical doctrine of the self advanced by the rational psychologist is logically parasitic upon the apperception doctrine. This is made clear by;

'The formal proposition of apperception 'I think' remains the sole ground to which rational psychology can appeal when it thus ventures upon an extension of its knowledge.' (CPR 336, A354)

Also bleibt eben so hier, wie in dem vorigen Paralogism, der formale Satz der Apperzeption: Ich denke, der ganze Grund, auf welchen die rationale Psychologie die Erweiterung ihrer Erkenntnisse wagt.

When he says this proposition 'belongs to and precedes every experience' he means it is a transcendental condition of rational self-conscious thought. The rational psychologist illegitimately takes the 'I' to refer to a metaphysical entity. This is the force of 'we have no right to transform it into a condition of a knowledge of objects'.

This critique of rational psychology, which makes use of Kant's theory of the formal requirements on knowledge, may be usefully appreciated in the context also of his views on the empirical self. Kant does not deny that 'I' has an empirical use, even an empirical use to refer. But he thinks there is nothing in our experience which allows us to use 'I' referentially but not empirically.

But clearly nothing follows from that about the simplicity or divisibility of the empirical self. The only sense in which it is warranted to talk about the simplicity of the self is this one:

'The proposition 'I am simple' must be regarded as an immediate expression of apperception.' (CPR 337, A354)

Der Satz: Ich bin einfach, muss als ein unmittelbarer Ausdruck der Apperzeption angesehen werden.

The Second Paralogism and Kant’s Critique of Descartes

Descartes is the prime target of the Second Paralogism. As the rational psychologist *par excellence*, in Kantian terms, Descartes reifies the formal unity of apperception

into the metaphysical simplicity of the soul. Descartes' alleged mistake is precisely to attach ontological significance to a purely formal proposition. There are two stages to the supposed error. The first is to think 'I' refers to a metaphysical self (roughly, the mistake of the First Paralogism). The second is to think facts about the use of 'I' warrant the ascription to the putative metaphysical self one of its essential properties: simplicity. Here is Kant on the first mistake:

'What is referred to as the Cartesian inference, *cogito ergo sum*, is really a tautology.' (CPR 337, A355)

So wie der vermeintliche cartesianische Schluß, cogito, ergo sum, in der Tat tautologisch ist.

Kant is only right about this if we allow that 'I am' (*sum*) is already expressed by 'I think' (*cogito*). But I should be inclined to allocate *cogito ergo sum* a less straightforward logical status. It is perhaps true that 'I think therefore I am' can be formulated, that the sentence itself may exist, just on condition it is true. So then we have 'cogito ergo sum' may be asserted, written, etc. if and only if 'cogito ergo sum' is true. This is not the same as saying the sentence is a tautology if a tautology is a logical truth, true just in virtue of the meanings of its constituent terms. But whatever the correct logic of the *cogito*, it is clear that Kant is correct in thinking of Descartes as a rational psychologist because he thought that not only are there thoughts but there necessarily thereby exists a metaphysical thinker of them: a substantial soul:

'No thought can exist apart from a thing that thinks, and in general no activity or accident can be without a substance in which to exist [[...]] There are activities which we call intellectual [[...]] The substance in which they reside we call a thinking thing or a mind.' (HR 64)

Even when Descartes sometimes talks of thoughts as properties of the soul, it is clear that he does not think the soul is a mere ontological aggregate or series of such thoughts;

'Besides the attribute which specifies a substance we must recognise the substance itself beneath the attribute; for instance soul; being a thinking thing is, in addition to thought, a substance which thinks.' (AK 64)

Kant's critique of Descartes on the self has two stages. The first is a defence of his interpretation of the logic of *cogito ergo sum* as a tautology 'since the *cogito* (*sum cogitans*) asserts my existence immediately'. (In other words 'sum' is already asserted by 'cogito' because it is part of the meaning of it.) The second stage is to invoke the pure apperception reading of 'I think' (*cogito*):

'I am simple' means nothing more than that this representation 'I' does not contain in itself the least manifoldness and that it is absolute (although merely formal) unity.'

Ich bin einfach, bedeutet aber nichts mehr, als dass diese Vorstellung: Ich, nicht die mindeste Mannigfaltigkeit in sich fasse, und dass sie absolute (obzwar bloß logische) Einheit sei.
(CPR 337, A355)

Kant thinks that unless a series of representations could be unified into a non-chaotic whole by the possibility of self-consciousness then 'I' could not have the use it has in our language. But the word 'I' does not itself mean 'manifold' or 'unity of the manifold'. (Otherwise, presumably, 'manifold' could be substituted for 'I' in linguistic contexts with no resulting change of meaning.) When Kant thinks of the 'I' as possessing a purely formal unity we can think of him as affirming 'I's use as an indexical, as a self-referential device in the language that does not carry any ontological or metaphysical import. The 'unity' or 'indivisibility' is a logical property of the word 'I' itself for Kant, and not a metaphysical property of its putative referent. This is why he talks of 'the indivisible unity of a representation I, where the representation is 'I' used meaningfully. Descartes' alleged error it to impart the logical simplicity of 'I' to I:

'The simplicity of the representation of a subject is not *eo ipso* knowledge of the simplicity of the subject itself.' (CPR 337, A355)

Die Einfachheit aber der Vorstellung von einem Subjekt ist darum nicht eine Erkenntnis von der Einfachheit des Subjekts selbst.

So, the logical simplicity of 'I' does not warrant any conclusion about the simplicity of the self. Indeed, the logical simplicity of 'I' mainly means that the first person pronoun has a grammatical function and it is not the name of an object the properties of which might be discovered from any alleged unpacking of predicates covertly contained in 'I'. There are no such predicates.

Interestingly, Kant substitutes the notion of a person for that of a metaphysical self when he allows 'I' an empirical referent. So Descartes' argument that he is a simple thinking substance rests on a grammatical mistake:

'Thus the renowned psychological proof is founded merely on the indivisible unity of a representation, which gives only the verb in relation to a person.' (CPR 337, A355)

Also ist der berühmte psychologische Beweis lediglich auf der unteilbaren Einheit einer Vorstellung, die nur das Verbum in Ansehung einer Person dirigiert, gegründet.

Here the indivisible representation is 'I' and the verb is 'think' and the unity is the sure apperception account of how the role of 'I' is possible. Kant's point is that the grammatical function of 'I' is left intact whatever the ontology of the subject, and nothing follows about that ontology from grammar alone. 'I' could be used by any self-conscious thinker and so does not presuppose a simple soul:

'The entirely empty expression 'I' (is) an expression which I can apply to every thinking subject.' (CPR 337, A355)

[...] wenn es lediglich durch den an Inhalt gänzlich leeren Ausdruck Ich, welchen ich auf jedes denkende Subjekt anwenden kann, bezeichnet wird.

Although he thinks Descartes is misled by ordinary language, Kant does not conclude from this that ordinary language should be revised or parts of it given up. On the contrary, we may retain, indeed we need to retain, 'I' and even 'substance' to make sense or ourselves as rational self-conscious beings. All we should guard against is misconstruing these terms metaphysically. 'Substance' is a category, and the categories have only an empirical use, so 'substance' has only an empirical use so 'substance' has no metaphysical use. 'Substance' is used to subsume a series of (re)presentations under a concept and make them into a single unified subject or object. There is no 'underlying' intuition of a substance that gives 'substance' this use:

'This concept ('substance') [[...]] tells us nothing whatsoever in regard to myself as an object of experience, since the concept of substance is itself used only as a function of synthesis, without any underlying intuition, and therefore without an object. It concerns only the condition our knowledge; it does not apply to any assignable object.' (CPR 338, A356).

[...] aber dieser Begriff,[Substanz] lehret uns nicht das mindeste in Ansehung meiner selbst als einen Gegenstandes der Erfahrung, weil der Begriff der Substanz selbst nur als Funktion der Synthesis, ohne unterlegte Anschauung, mithin ohne Objekt gebraucht wird, und nur von der Bedingung unserer Erkenntnis, aber nicht von irgend einem anzugebenden Gegenstande gilt.

'Substance' does not apply to any object (or subject) because the subsumption of intuitions under 'substance' in synthesis is constitutive of objects (and subjects). The idea of a subjective subject or an objective object already makes use of 'substance' in a way that presupposes the synthesis of intuitions. It is not as though there is some further intuition 'underlying' the others to which 'substance' could refer. In this way Kant hoes to show that if his account of categorial synthesis is correct, then the formulation of Descartes rational psychology is logically dependent on it. 'Substance' can be misused only on condition its genuine use exists. Descartes is involved in a kind of ontological duplication. The concept of substance is allegedly exhausted in

yielding us our possible talk of ourselves as empirical subjects and physical objects. Kant thinks there is no extra work for it to do in referring to a soul which underlies the empirical self and makes it possible, any more than there is extra work for it to do in referring to a putative Lockean substratum which underlies the appearances of physical objects and makes them possible.

So long as we recognise the empirical uses of 'I' and 'substance', and the transcendental facts which make those uses possible, then not only does Kant think using 'I' and 'substance' may be philosophically harmless, he even thinks 'I am a simple substance' may have a use: 'I may therefore legitimately say 'I am a simple substance' so long as this rests only on those empirical facts and transcendental presuppositions.

Kant, perhaps a little cynically, ascribes to the rational psychologist two motivations for attempting a proof of the simplicity of the self: They want to be mental and They want to be immortal. Kant is not clearly right to note that neither conclusion would in fact go through. From the bare fact that *x* is in principle indivisible it does not follow that *x* is mental (or spiritual) but it does follow that x is not physical. x might be neither mental nor physical (for example, if x is a matheamtica point). Nor does it follow from the bare fact that x is in principle indivisible that *x* could not cease to exist. That which is simple could be wholly annihilated. Nevertheless, nothing simple could be naturally destroyed. If natural destruction entails separation into components but nothing simple has any components then nothing simple admits of natural destruction. Even the rational psychologist will agree that the soul admits of supernatural annihilation.

Kant as an Identity Theorist

Kant says that if we have to give up the Cartesian thesis that 'only souls think' then

'We should have to fall back on the common expression that men think'
(CPR 340, A359-60)

Es würde wie gewöhnlich heißen, dass Menschen denken.

It is not as though these are philosophical options of equal status: The critique of Descartes construed his mistaken metaphysics of the self as dependent upon our ordinary empirical language of the self. Now Kant continues by offering a philosophy of mind which shows what rational self-conscious beings minimally have to be like in order for the language of the self to have those empirical roles. This account is: the concept of a person is presupposed by that language but the concept of a soul is not. We need now to explicate Kant's concept of a person. It is essentially that which possess both subjective and objective properties, and both mental and physical properties. The allegedly erroneous idea of a soul substance is logically parasitic upon this concept of a person as that to which both sorts of predicate

essentially apply. (I do not enter into the interesting question of the relationship between this theory and those of Merleau-Ponty, Strawson and Ryle.)

Kant draws a distinction between 'mental' and 'physical' which is not metaphysical. This is broadly: *x* is mental if and only if *x* is an object only of inner sense, and *x* is physical if and only if *x* is an object only of outer sense. Here is the criterion for 'mental':

'Our thinking subject is not corporeal; in other words [...] in as much as it is represented by us as object of inner sense it cannot in so far as it thinks be an object of outer sense, that is, an appearance in space'

Dass unser denkendes Subjekt nicht körperlich sei, das heißt: dass, da es als Gegenstand des inneren Sinnes von us vorgestellt wird, es, in so fern als es denkt, kein Gegenstand äußerer Sinne, d.i. keine Erscheinung im Raume sein könne.

Two remarks are needed here: He only says this is a criterion for *x*'s being non-physical (not corporeal), but the use he puts the distinction to in his argument makes it clear than 'not physical' here means 'mental'. Secondly, it expresses the Kantian thesis that if *x* is an object of inner sense then *x* cannot even in principle be an object of inner sense. That thesis is required by the formulation of the criterion for *x*'s counting as physical also:

'Matter is mere outer appearance'

Nun ist [Materie] bloß äußere Erscheinung.
(CPR 339, A359)

where

'Extension, impenetrability, cohesion and motion (are) everything which outer sense(s) can give us'

Ob nun aber gleich die Ausdehnung, die Undurchdringlichkeit, Zusammenhang und Bewegung, kurz, alles, was uns äußere Sinne liefern können.
(CPR 339, A358)

Clearly, this Kantian mind/matter distinction is not between two different sorts of metaphysical substance, and Kant insists that the following conclusion does not follow from it, even though it has been thought to 'from earliest times';

'souls (are) different entities from their bodies'

Seelen als von den Körpern ganz unterschiedene Wesen zu betrachten.

(CPR 339, A358)

To see that mind-body substance dualism does not follow from the mind-matter distinction we have to invoke the Kantian thesis that 'substance' is a category with only an empirical use. Then it follows that 'substance' has the function of synthesizing the intuitions of inner and outer sense so does not denote a metaphysical bearer of mental or physical properties. The distinction between inner and outer intuitions is logically prior to that between mental and physical substance in the sense that we could not possess the concept of the latter if we did not possess the concept of the former. So the only understandable substances are empirical substances, and these are appearances subsumed under that category. We may if we wish still talk about mental and physical substance but this is just a distinction between two sorts of intuitions which may be subjected to synthesis: those of inner sense and outer sense respectively. It is fair to say that for Kant the distinctions between mind and matter, subject and object are logical constructions out of phenomena, not in a straightforwardly phenomenalist sense, in which any sentence about a mind or a physical object could in principle be translated into a sentence (or set of sentences) about intuitions, but in the sense that 'substance' allows intuitions to be thought of in a particular way, as empirical subjects and as empirical objects, or as objects of inner and outer sense. So I prefer to call Kant's theory 'monism'.

We need this empirical use of 'substance' to understand the next part of Kant's monism:

'The substance which in relation to our outer sense possesses extension is itself the possessor of thoughts.' (CPR 340, A360)

[...] und dass also der Substanz, der in Ansehung unseres äußeren Sinnes Ausdehnung zukommt, an sich selbst Gedanken beiwohnen.

and

'The very same being which, as outer appearance, is extended, is (in itself) internally a subject.' (CPR 340, A360)

Eben dasselbe was, als äußere Erscheinung, ausgedehnt ist, innerlich (an sich selbst) ein Subjekt sei.

So, his thesis that 'men' think' requires the following concept of a person: A person is both subject and object, the subject of his or her own experiences but the object of their own or another's experiences. A person has both mental and physical properties: He or she appears physical *qua* object of outer sense, but he or she appears mental *qua* object of inner sense. So the difference between mental and physical ceases to be a distinction between two different metaphysical substances and becomes instead a

distinction between two different perspectives on the person, one subjective and one objective. This theory is wholly congruent with his view that 'substance' has an application only to appearances.

I call Kant's philosophy of mind 'monism' for two main reasons. Firstly, phenomena, the content of experience, may be described using the predicates 'mental', 'physical', 'subjective', and 'objective' depeding on how they may be accessed, but it allegedly does not make sense to ask the metaphysical question whether phenomena are intrinsically any of these. They are in fact neutral between various possible descriptions of them. Secondly, the empirical idea of the whole person must be possessed by anyone drawing a mental/physical, or mind/body distinction, because the concept of the mental has meaning only because a person can 'inwardly intuit himself' in inner sense, and the concept of a physical object has meaning only because empirical persons may exercise outer sense. This is what I mean by saying Kant's mental/ physical distinction rests on a difference between two ‘perspectives’ a person has: one only on him or herself, one on him or herself and also on other person. Kant's word for these perspectives is 'relations':

'In this way, what in one relation is entitled corporeal would in another relation be at the same time a thinking being.' (CPR 340, A359)

Auf solche Weise würde eben dasselbe, was in einer Beziehung körperlich heißt, in einer andern zugleich ein denkend Wesen sein.

This monist theory is inconsistent with mind-body dualism, as Kant points out, because dualism is the view that there exist two different sorts of (non-empirical) substance, something that it does not make sense to assert for Kant. Finally, Kant's philosophy of mind is a monism in yet another sense. [He does not say that unless they had mental properties they could not have physical properties. I call these 'weak materialism' and 'weak idealism' respectively.] His theory it properly monist because no fundamental ontologically dualist conclusions may be adduced from the findings of inner sense of outer sense; only conclusions about appearances. That is Kant's solution to the mind-body problem.

The Second Paralogism in the Second Edition

The putative refutation of the *a priori* argument from the simplicity of the self is:

(1) 'The 'I' of apperception, and therefore the 'I' in every act of thought, is one, and cannot be resolved into a plurality of subjects, and consequently signifies a logically simple subject'

Dass das Ich der Apperzeption, folglich in jedem Denken, ein Singular sei, der nicht in eine Vielheit der Subjekte aufgelöst werden kann, mithin ein logisch einfaches Subjekt bezeichne, liegt schon im Begriffe des Denkens.

(2) '[It] is therefore an analytic proposition'

[...] ist folglich ein analytischer Satz.

(3) That proposition ['I' is a simple substance'] would be synthetic'

Dass das denkende Ich eine einfache Substanz sei, welches ein synthetischer Satz sein würde.CPR 369, B407-8)

(4) 'This does not mean that the thinking 'I' is a simple substance'

Aber das bedeutet nicht, dass das denkende Ich eine einfache Substanz sei.

The first premise is a reminder that the transcendental unity of apperception doctrine is a purely formal one: that the word 'I' is logically simple. Conceptual analysis could not reveal it to have any ontological import.

When Kant says that 'I' in every act of thought is one, he means that each thought is thought by a unitary subject: Every thought is someone or other's. This is a necessary truth for Kant, because he is individuating thoughts through the thinker of them. Although numerically distinct thoughts may be thought by the same thinker, it is logically possible that the whole of numerically one and the same thought be thought by numerically distinct thinkers. This does not preclude two possibilities which Kant allows: Each of the parts of one and the same thought may be thought by more than one thinker, and qualitatively identical but numerically distinct thoughts may be thought by more than one thinker.

So, some episode counts as a thought on this account of 'thought' only if it is thought by just one thinker. We should read 'signifies' not as 'refers to' but as 'means' because if 'I' could be known to refer to a simple substance on the ground of its meaning alone, then there could be an *a priori* rational psychology, at least in so far as the simplicity of the self would follow from linguistic facts alone: The simple self would would be the referent of 'I' on that reading. Clearly Kant wishes to exclude that possibility, so we should read (1) as including the claim that 'I' is a simple subject, taking 'subject' grammatically, not metaphysically. Then there is no danger of being misled by 'subject' into drawing ontological conclusions about the subject of experience (a putative metaphysical entity) from grammatical facts about the subject of first person sentences: which is a word.

(2) says that (1) is analytic. This is not just the usual assertion that that 'I think […]' may prefix any of my thoughts is analytic, though it relies on that. It is the claim that the 'I' of apperception is logically simple is analytic.This may be allowed if we

take (1) as a definition of the first person singular pronoun. Any such definition will take the form of an explication of the grammatical function of 'I', and if it is correct then it will be analytic just because all verbal definitions are analytic.

Kant's explication of (1) is not complete. For example, Kant does not make explicit the fact that 'I' is the word each person uses to refer knowingly only to himself. It is part of the grammar of 'I' that 'I' is logically simple, but occupies the grammatical subject role in sentences, so (1) is only a part definition of 'I'. It is not a necessary condition for the analyticity of some sentence that that the proposition it expresses be a complete definition, even though that would be sufficient. Suppose a sentence is analytic if and only if either its meaning is sufficient for its truth or its truth may be discovered merely by inspection of the meanings of its constituent terms. It is decidable simply by the inspection of 'I is a logically simple subject' that it is true, so by that crieterion that sentence is analytic even though it is not a complete definition of 'I'. So (2) is true and follows from (1).

(3) is the assertion that a certain ontological claim is synthetic *viz* 'I am a simple substance. This proposition clearly is synthetic because it gives (putative) information about the self that is not already expressed by the use of 'I'. If true, it could count as informative for someone that the I is a simple substance. Kant introduces this premise to contrast it explicitly with the analyticity of '"I" is a logically simple subject'. The two propositions belong to distinct logical categories because if *p* is analytic then *p* is not synthetic, and if *p* is synthetic *p* is not analytic. It follows that the grammatical truth about the first person singular pronoun is not the same proposition as the ontological claim about the self, just because the first is analytic and the second is synthetic.

However, it does not follow that the ontological (or putatively metaphysical) claim does not follow from the linguistic claim. Although no synthetic proposition follows from any proposition which is wholly analytic, *pace* Kant, his (a?) and (b?) are not purely formal or analytic. If *p* entails some synthetic proposition *q* then it follows that *p* is not wholly analytic because *p* is partly informative by containing *p* as an embedded sub-sentence. '"I" is a logically simple subject' is not wholly analytic, and does entail the synthetic proposition 'The I is a simple substance'. This is a synthetic proposition derived from the first.

(4) follows from its premises, despite the fact that "I' is a simple substance' occurs both in the conclusion and in the third premise. The argument is valid because it does not follow that 'I is a simple substance' is true, from the fact that 'I is a simple substance' is synthetic. We cannot logically derive any conclusion about the truth value of a proposition from any set of facts about the logical category that proposition belongs to unless the truth value of that proposition is already expressed by those facts. For example, if a proposition is analytic we may safely allocate to it the truth value 'true', because the fact of its belonging to that logical category entails that it is true. But in (3) Kant has allocated 'I is a simple substance' to the logical category 'synthetic', and it is consistent with some proposition's being synthetic that it be true or that it be false.

So Kant is incorrect to deny the rational psychologist the truth of the thesis of the simplicity of the self as it is derived from (3). Clearly, that thesis does follow either from either (1) or (2) or from the conjunction of (1) and (2). This is because (1) and (2) do not express purely formal truths. The simplicity of the self is a substantive or synthetic thesis, and although this substantive or synthetic thesis does not follow from any purely formal facts, Kant's premises do not express purely formal facts. So in (4) we may read 'this does not mean that' to mean 'these premises do not logically imply that' so Kant is here denying the validity of an inference which is valid.

I conclude that Kant's argument against the putative derivation of the conclusion of the Second Paralogism is unsound.

III

Kant and the Problem of Personal Identity

In virtue of what exactly is a person at one time numerically identical with a person at a different time? In the Third Paralogism Kant is concerned to repudiate one answer to that question: *By having the same immaterial soul* and he substitutes an alternative account. (1) I shall argue that Kant's attack on the soul is unsuccessful and that there are reasons for thinking the solution to the problem of personal identity is strongly dualist. Kant states the dualist argument which is his target in this way:

(1) 'That which is conscious of the numerical identity of itself at different times is in so far a person.' (A 361)

(1) 'Was sich der numerischen Identität seiner Selbst in verschiedenen Zeiten bewußt ist, ist so fern eine Person.' (A 361)

(1) $\forall x\ (Cxn \supset Px)$
$\forall x\ (Kx\ (x = x) \supset Px)$

(2) 'Now the soul is […] [conscious of the numerical identity of itself at different times]' (A 361)

(2) 'Nun is die Seele […] [der numerischen Identität seiner Selbst in verschiedenen Zeiten bewußt ist]' (A 361)

(2) Csn

Ks (St1 = st2)

(3) 'Therefore it is a person.' (A 361)

(3) Also ist sie ein Person.' (A 361)

(3) Ps

Clearly the argument of the First Paralogism is logically valid. The problem is whether it is sound. It is valid because $\forall x (Fx \supset Gx) \& Fa \vdash Ga$ is a valid pattern of argument:

1. $\forall x (Fx \supset Gx)$	Assumption
2. Fa	Assumption
3. $Fa \supset Ga$	UE 1
4. Ga	MP 2, 3

It might not be sound because either or both of its premises might be false. It is in fact Kant's strategy to argue that the second premise is false. Before examining Kant's criticisms of the argument we should decide what the three premises mean, by clarifying their component concepts.

The First Premise

(1) is one definition of 'person': [If *x* is conscious and] if *x* [thereby] knows that it is just the same entity it is at some earlier or later time then that is [both a necessary and] a sufficient condition for *x* 's being a person. It is worth separating out the conceptual components of this recurrent view in the history of philosophy. On the definition, *x* is a person if and only if

(a) *x* is conscious
Bx
K*x*

(b) *x* persists over time

WEAK PERSISTENCE (WP): $\exists t \exists t' (t \neq t' \wedge t<t' \wedge E(x,t) \wedge E(x, t'))$

STRONG PERSISTENCE (SP):

$\exists t \exists t' (t \neq t' \wedge t<t' \wedge E(x,t) \wedge E(x, t') \wedge \forall t'' (t \leq t'' \geq t' \supset E(x, t'')))$

(c) *x* is conscious of persisting over time
C(x, WP OR SP)
K (*x* WP or SP)

(c) implies *x* has memory, given that *x* is not conscious of existing at a future time, and that its consciousness of its existing is a present consciousness, so we may add

(d) *x* has memory
M *x*
K *x p* at t1 (where p is true at t-1)

Does it also follow that *x* has a concept of itself on this account? It is certainly true that if *x* is conscious of its numerical identity over time then *x* is conscious of the persistence of the being that is in fact *x* over time, but in order to achieve this it would seem not to be logically necessary that *x* be possessed of the concept ' *x* ', or 'person'. Minimally, if *x* is conscious at t^2 of its numerical identity with Y at t^1, then *x* must know that it is itself that being which at t^2 is numerically identical with Y at t^1. What does knowing *that* consist in? Minimally, *x* must know that it is identical with the being that is [conscious of] being identified with Y at t^1. Arguably, if *x* knows, under one description, that it is numerically the same being as one subsumed under a different description then *x* possesses a concept of self. We do not have a couch *x*'s achievements in terms of 'descriptions' to obtain this result. If a concept is a recognitional ability, then *x* has the concept itself because *x* has the ability to identify (and perhaps reidentify) that being which it in fact is (at an earlier time) with that being which it in fact is (at the present time). Before we can add

(e) *x* has a concept of itself
I*x* & K*x* (*x* = *x*)

to the analysed elements of 'person', one objection remains. Could *x* at t^2 successfully and correctly identify *x* at t^2 with Y at t^1 without *x* thereby knowing that it is Y at t^1 was one and the same X as that X that identifies Y at t^1 with X at t^2? In this situation, *x* has a use (if he is a language user) for the expressions 'Y at t^1' and *x* at t^2' and can formulate the (true) sentence '*x* at t^1 is numerically identical with *x* at t^2'. There is one sentence in particular however which he cannot formulate: 'I am *x*'. If this is a logical possibility, then it follows that even if *x* is identical with *x* at t^2 (and hence at t^1) *x* nevertheless has no concept of self in this respect: He does not know who he is, where 'who he is' means ' *x* at t^1 and Y at t^2'.

Prima facie Kant might seem to have met this objection in advance by using 'conscious [[...]] of itself' in formulating the first premise of the Third Paralogism. But, in fact, 'itself' is crucially ambiguous on just this point. It could mean 'that being which in fact it is', or, more strongly, 'that being which it in fact is, where that being

knows it is that being'. From the fact that *x* is conscious of *x* at t^1, and Yt^2, or conscious that Y at t^1 is identical with *x* at t^2 it does not follow that *x* at t^2 knows that it is identical with either Y at t^1 or *x* at t^2. What requirement would have to be met in order for this condition to be fulfilled? Not only must *x* be capable of consciousness of *x* but *x* must be further capable of consciousness that it is *x*. But what does that thought consist in? If it means, '*x* must be capable of the consciousness that it is itself *x*' then we employ 'itself' in a question-begging way. To unpack the residual component of 'itself' we need to distinguish between two descriptions of *x*:

' *x* is conscious of *x* at t^1'
$C(x, x(t^1))$
$K(x, x(t^1))$

and

'X is conscious that the *x* that is conscious of *x* at t^1 identical with the *x* that *x* is conscious of at t^1'.
$C(x(t^1), [\iota(y) (C (y, x(t^1))) = (\iota(z) (C(x(t^1), z)))])$
$K(x(t^1), [\iota(y) (K (y, x(t^1))) = (\iota(z) (K(x(t^1), z)))])$

It is this self-reference of reflectivity that is captured elliptically by the use of 'I'.

It might seem open to the skeptic to object in this way: Although *x* might be conscious that the *x* that is conscious of *x* at t^1 is identical with the *x* that *x* is conscious of at t^1, there might still be one residual but relevant fact *x* does not know: 'I am *x* ', and so we cannot say *x* has a fully fledged concept of self. There is a threatened regress here in that no matter however many times we preface

'*x* is conscious of Y's identity T^1 with *x* 'at t^2'

with

'*x* is conscious that *x* is the *x* that […]'

x will not know that *x* is himself. There seems room for a reflexive fact to be expressed by 'I am x' that is not captured by these two thoughts:

' *x* is self-identical'
x =x

which is a modal fact about *x* , and

' *x* is the *x* who is conscious that[...]'
x = ι(y)….

A Priori Subjects: Kant and the Existence of the Soul

or some other reflexive description of *x* made by *x* . In any such reflexive description *x* refers to *x* and knows that describer and referent are numerically identical. This exhausts the essential self-reflexive function of 'I'. If we incorporate this feature of 'I' into 'itself' in the first premise of the Third Paralogism, then we are entitled to add (e) '*x* has a concept of itself' to the conceptual components of the concept of a person it defines. Then we obtain:

(a) *x* is conscious

(b) *x* persists over time

(c) *x* is conscious of persisting over time

(d) *x* has memory

(e) *x* has a concept of (it)self

So

(c) *x* is a person

What does (a) mean? 'Conscious' is an imprecise word, but one deep ambiguity here is between

'*x* is capable of having experiences'

and

'*x* is capable of knowledge'.

The second is needed for the first sentence of the First Paralogism because no present experience could logically be said to be an experience of oneself at another time. If we object that memories could, then although memory might have an experiential dimension this is not essential to what it is, and an epistemic notion has already been smuggled in. This is because if *x* at t^2 (correctly) remembers being Y at t^1 then *x* thus acquires the knowledge, or at least the true belief, that he is identical with *x* at t^1. If this epistemic component of memory is subtracted then it hardly makes sense to talk of *x*'s t^2 experience of itself at t^1, and if that is unpacked as something like 'image of' then, shorn of an epistemic component, it falls short of memory.

(b) is not forced on us with logical compulsion by Kant's formulation of the first premise of the First Paralogism, but is presumably an assumption Kant makes here. Suppose:

x at t^1 is conscious of being numerically identical with *x* at t^2

C(*x* (t^1), (*x* (t^1) = *y*(t^2)))

and this thought [or thought content] of *x* is correct. It follows that

x at t^1 is identical with *x* at t^2
(*x* (t^1)=*y*(t^2))

Does it follow that

x persists between t^1 and t^2?
SP, last part

I do not think this does follow with logical certainly. There are no 'intermittent selves', so far as we know, but it seems nevertheless logically and metaphysically possible that Y at t^1 is *x* at t^2 even if *x* ceased to exist for some period between t^1 and t^2. This is incompatible with a widely held theory of personal identity; that the spatio-temporal continuity of some person's body between t^1 and t^2 is not just sufficient but necessary for the person existing at t^1 being numerically identical with the person existing at t^2. But suppose the person you are holding a conversation with at t^1 disappears, ceases to exist for a minute or two, then at t^2, to all appearances reappears, recommences existing, and continues the conversation as though absolutely nothing had happened. Would you feel forced to say that the person after t^2 is not numerically identical with the person who existed up to and including t^1? I see nothing to force such a conclusion. It would be much more natural to say that your interlocutor had, rather oddly, ceased to exist for a minute or so. (We would perhaps be less inclined to opt for identity if at t^2 two, five or five hundred million persons qualitatively identical to the earlier person were to appear.)

But could such a putative intermittent person remember their earlier states? If all continuity between *x* at t^1 and *x* at t^2 is broken then so too is any causal connection which could be the empirical basis for the application of a causal criterion for *x* at t^2 remembering being Y at t^1, and on most plausible views of memory if A remembers B then being B or experiencing B is a necessary condition for remembering being B. It would beg the question to insist that *x* at t^2 did genuinely remember being *x* at t^1 for those two reasons, even though if *x* at t^2 is *x* at t^2 then they are fulfilled. If *x* at t^2 is able to tell us a great deal about *x* up to and including t^1. Then I think we would be false to say *x* did not remember its states up to and including t^1. The puzzle then would not be 'Are these memories?' but 'How are these memories possible?'

It may be doubted whether *x*'s being able to remember its earlier states in this fashion is a necessary condition of *x*'s being a person. Suppose a human being to suffer suddenly from total amnesia. We could not correctly say that such a person is

not numerically identical with the person who he was before he lost his memory. (If we did say this we could intend it perhaps colloquially to mean different sort of person, not different one.) It follows that (c) is not entailed by the first premise of the paralogism if 'persists over time' means 'persists with unbroken continuity'.

(d) does follow from the first premise of the paralogism, but one final ambiguity needs to be dispelled.

'That which is conscious of the numerical identity of itself at different times […]'

could mean;

'That which is, at different times, conscious of its numerical identity (ie; its modal property of being self-identical) […]'

$C(x , \forall t\ (x\ =(t)\ x\ ))$
$K(x, \forall t\ (x\ =(t)\ x\ ))$

Dispelling this interpretation we have;

'That which is conscious at one time that it is identical with a being that exists at a different time from that time […].'
$C(x\ (t^1), (x\ (t^1) = y(t^2), t^1 \neq t^2))$
$K(x\ (t^1), (x\ (t^1) = y(t^2), t^1 \neq t^2))$

As argued above, such a being has memory. (e) has already been dealt with.

I take it that (a) - (e) amount to the sense in which 'person' is intended by the First Paralogism so it is not pertinent here to question that identification, and we may turn instead to the meaning of the second premise.

The Second Premise

(2) substitutes 'soul' for 'that which' in (1). So (2) can be read either (semantically) as a part definition of 'soul', or (ontologically) as the postulation of the soul as that which meets the description given in (1). If (2) is a part definition then it needs to be supplemented by traditional accounts of what the soul essentially is, but if it is merely stipulative there seems little value in criticising it as inadequate. If the second premise is read as the postulation of the existence of a soul (as an immaterial substance) then there is a philosophical question here: Must a person, as described in (1) be really or essentially a soul, to meet that description? What is the rational psychologist's concept of the soul? Kant's target is Platonic, Augustinian and Cartesian not Aristotelian or Thomist. I summarise it as a substance with these properties:

(1) Immaterial

(2) Spiritual

(3) Temporal

(4) Conscious

(5) Immortal

(6) A person's self

to which we could add a property not treated by Kant in the Paralogisms chapter:

(7) Created

What do these properties amount to?

If *x* is immaterial then *x* is not a physical object, nor a physical property of a physical object. By this I mean *x* possesses none of Locke's primary qualities.

If *x* is spiritual then in some sense pertains to God.

By saying *x* is temporal it is implied that temporal predicates truly apply to at least the operations of the soul, and that the soul itself has duration; because it lasts from the moment of God's creation of it, to the possible moment of God's annihilation of it.

If the soul is conscious then it is capable of thought, in the broadest sense that includes having experiences. I leave it open as a further, Cartesian, possibility that the soul perpetually thinks, does not stop thinking, so long as it exists.

The soul is 'immortal' means that it can only be destroyed by God, and apart from this possibility, is indestructible because it is simple.

Finally, by (6), is meant not just that if a certain soul exists then a certain person exists, and if a certain person exists then a certain soul exists, but also that the soul-person relation is numerical identity. This implies that a person could not survive the destruction of their soul, because they are it essentially, but they could survive the destruction of their body, because they are it only contingently, or they 'have' it.

Is there any reason to suppose that only an entity with properties (1) - (6) could also be the subject of the predicates which feature in the first premise of the Third Paralogism? Specifically, may we derive (1) - (6) embedded in 'There exists an entity such that that entity is [...]' from (a) - (e)? To decide this we need to know whether any of (1) - (6) follows from each of (1) - (e). I shall indicate the philosophical problems which lie in the way of deciding some of these logical relations, and point out some possible solutions.

Whether if *x* is conscious *x* must be immaterial depends upon a solution to the mind-body problem. If a person is a physical object with no non-physical parts and

yet is conscious, then it is not incoherent to conjecture that is conscious' and ' *x* is not immaterial' may both be true. However, if Cartesian dualism is true then the concept of a material object being conscious is incoherent. The inference required is from

' *x* is conscious'
C*x*
K*x*

to

'whatever is conscious is not physical'

$\forall x \ (\mathrm{C}x \supset \neg \mathrm{P}x)$
$\forall x \ (\mathrm{K}x \supset \neg \mathrm{P}x)$

but this would seem not to follow if, for example, the brain is conscious.

A more modest inference is from

'*x* is conscious'

to

'*x* possesses at least one non-material property'

$\exists F \ (F \notin \{G\text{: } G \text{ is a physical property}\} \wedge Fx)$

This follows if consciousness is not material, but is not strong enough for the rational psychologist. A further premise:

'Whatever possesses non-material properties is necessarily itself non-material'

$\forall x \ (\exists F \ (F \notin \{G\text{: } G \text{ is a physical property}\} \wedge Fx) \supset \Rightarrow \neg Px)$

is needed for

' *x* is non-material'

$\neg Px$

to follow from

'*x* is conscious'.

Without this (1), 2(1)s cannot be derived from (a).

If it does not follow that *x* is immaterial from ' *x* is conscious', then it does not follow that *x* is spiritual either even if being non-material is a necessary condition for being spiritual. However, if Christ is both material and spiritual then being non-material is not a necessary condition for being spiritual tout court. It is in any case doubtful whether it could be demonstrated that non-material objects are spiritual from premises which did not themselves include religious predicates. So, even if it is the case that being immaterial is a necessary condition for being conscious, it does not follow that being spiritual is a necessary condition for being immaterial. So (2) does not follow from (a).

Although, as just argued, if *x* is conscious then it is logically possible that *x* have no origin and no end, it is clearly *prima facie* consistent with '*x* is conscious' that *x* come to be totally annihilated. There are different sorts of destruction. For example, *x* may in one sense be truly said to have ceased to exist if:

(ED) All the parts of *x* continue to exist but if they are separated from one another, so as not to form whole that could be correctly termed ' *x* '

In another sense of 'destroyed' however:

(MD) *x* is truly said to have ceased to exist when no part of *x* exists: when prior to *x*'s destruction there was something called '*x*' but subsequent to *x*'s destruction there was nothing that is truly called '*x*'.

x's being conscious is consistent with *x*'s annihilation in either of these senses of 'annihilate'. *x* could, for example, be subject to annihilation in the first sense first and then in the second sense. It might be argued that the first sort of annihilation presupposes that *x* is physical. This is true if 'separated from one another' means 'spatially separated' but the presupposition is otherwise not logical. Sense might be made, for example, of God causing my thoughts to occur at randomly different times, so that they ceased to posses that unity and coherence necessary for their counting as anyone's. So (5) does not follow from (a).

The conjunction of (a) and (4) is clearly analytic, so (4) follows from (a) and the inference need not be considered further. The putative derivation of (6) from (a) is more contentious. If *x* is conscious then, rather than a person, *x* is a person's self (or what a person essentially is). On any reasonably complex definition of 'person' a person will be capable of much besides consciousness: rationality, self-reference, language use, memory, social interaction and perhaps more. It follows that even if the truth of '*x* is conscious' is a necessary condition for ' *x* is a person', '*x* is a person' does not follow from ' *x* is conscious'. *A fortiori* '*x* is conscious' cannot yield as a logical consequence '*x* is what a person essentially is'. So, (6) does not follow from (a).

Does any of (1)-(6) follow from (b)? Clearly the only inference that obtains here is from (b) to (3), because the conjunction of (b) and (3) is analytic. From the truth of '*x* persists over time' it does not follow that *x* is immaterial, spiritual, conscious, immortal, or what a person essentially is.

The putative inference from (c) to (1)-(6) is more interesting. The valid derivation of

' *x* is immaterial' (1)

from

' *x* is conscious of persisting over time,

depends, like that from (a) to (1), on the solution to the mind body problem. If Cartesian dualism is true then the inference goes through, if materialism is true it fails.

(2) cannot follow from (c) so long as no description of the world in which theological predicates feature may be logically derived from any description in which they do not.

The validity of the move from (c) to (3) rests on an equivocation on 'conscious of'. If ' *x* is conscious of [...]' is partly equivalent to

' *x* is conscious that [...] p'

where 'p' for example takes the value ' *x* is conscious over time'

and if it is further true that if

' *x* is conscious that p' is true then 'p'

$C(x, p) \supset p$

then the inference is logically legitimate. On this interpretation of ' *x* is conscious that p' 'conscious that' has a logic analogous to 'knows that' such that it would entail a contradiction to assert that *x* is conscious that p, but not p.

On a separate and contrasting interpretation,

'*x* is conscious of [...]'

is not partly equivalent to '*x* is conscious that p' and it might consistently be the case that '*x* is conscious that p' but 'not p'. On this interpretation of 'conscious of', 'conscious of' has a logic analogous to 'believes that'.

This distinction yields either

(i) ' *x* is conscious of existing over time, therefore *x* persists over time'

$C(x , y) \supset y$

or

(ii) ' *x* is conscious of existing over time, but *x* does not thereby persist over time'

But are there independent grounds for doubting (ii)? If it were the case that ' *x* is conscious of persisting over time but *x* does not persist over time', then it would have to be the case that that which did not itself possess any duration, not even a punctual or instantaneous existence, would nevertheless be capable of consciousness, and the mistaken belief that it itself is temporal, that is, had minimal duration. Suppose

' *x* exists'

entails

' *x* lasts'

so if *x* has no duration then *x* does not exist. If that is right then if *x* has no duration then *x* cannot be conscious, still less conscious of its own existence in time, because *x* does not exist. In order to meet this, an account of a non-temporal reality in which mistaken beliefs about time could exist would have to be shown to be coherent. If that project is in principle impossible then if *x* is conscious of existing over time then *x* exists over time, not just because of the reading of 'conscious of' that yields (1), but because of what 'exists' means.

If (c) is true then (4) is true because, clearly, if *x* is conscious of persisting over time then *a fortiori* '*x* is conscious' is true.

(5) does not follow from (c) because even if the inference from (c) to (3) is valid, and it is true that *x* persists over time, it still does not follow that *x* will continue to exist through time: eternally. Obviously, from

' *x* lasts'
WP OR SP

it does not follow that

' *x* lasts eternally'

$\neg\exists t \neg(E(x ,t)$

even if it is true that

x exists and *x* lasts at least so long as it is true that *x* is conscious of persisting over time.

So, if (c), *x* is conscious of persisting over time, it does not follow that (5), that *x* is immortal.

If the first premise of the Third Paralogism is true, then it may be validly inferred that (6),

x is a person's self

or

x is what a person essentially is

from (c)

x is conscious of persisting over time.

This is because the definition of '*x* ' is a description of the essence of '*x* '. This inference alone is not of much value to the rational psychologist because he wants a person's essence to be described using predicates in (1) to (5) and not just (6). All that follows, though, is that if it is true by definition that '*x* is conscious of persisting over time' is both a necessary and a sufficient condition of being a person on that account. That concept of a person is too poor for the rational psychologist's purpose in proving the truth of a substance dualist ontology.

If (D), *x* has memory, (1) and (2) do not go through for the reasons they do not follow from (a).

(3) follows from (d) just so long as it is logically impossible that there should be a-temporal consciousness.

(4) follows from (d) just so long as memory is a kind of consciousness.

(5) does not follow from (e) because on the basis of knowing that *x* has memory it cannot be concluded that *x* will not cease to exist.

From ' *x* has memory' it might follow that '*x* is what a person essentially is' if the definition in the first premise of the Third Paralogism is correct and 'memory' here is taken to include memory of oneself. If '*x* has memory' is read as consistent with '*x* has no capacity to remember himself' then that inference from (d) to (6) fails. This is because *x* could be capable of some forms of memory and yet '*x* ' not meet the definition in the first premise.

Finally, if (e) is true, so *x* has a concept of himself, is it true that (1) - (6)? If having a concept of itself is a linguistic capacity for self reference then there would seem no *a priori* obstacle to this concept being possessed by an entirely physical

object, in which case (1) does not follow, nor (2). If there logically could be a non-temporal reality, then it is hard to see how '*x* has a concept of itself' would be any obstacle to that. This is because

'has the concept of [...]'

is not an occurrent mental state that could be timed like a sensation. If there is a problem about

'*x* has a concept of itself but is not temporal'

then this is a general problem about something existing but not lasting.

So, if non-temporality is thinkable then (3) does not necessarily follow from (e). If having a concept of oneself is a linguistic capacity then it does not follow from '*x* has a concept of himself' that '*x* is conscious' is true. This is because consciousness is a capacity for having experiences: sensations, perceptions, images, etc. These kind of awareness are not at all the same thing as linguistic capacities. Further, the fact that computers posses linguistic capacities but no consciousness shows that consciousness is not required for concept acquisition and use in that rather austere sense. There is no additional reason relying on consciousness, why a being should not make reference to itself without any experience or awareness of itself. So, from (e) it does not follow that (4).

Obviously, from just:

' *x* has a a concept of himself'

it does not follow that

x is immortal.

There is no reason why that which has a concept of its own existence should not cease to exist.

Finally, does it follow from (e) '*x* has a concept of itself' that (6) '*x* is a person's self, or what a person essential is'? This depends on the definition of 'person'. If that expressed by the first premise is acceptable then (6) does follow from (e). But if the considerations about computers are relevant, if for example a computer may make records of its earlier machine states and presently review them, then a choice is presented: Either computers are persons, or there is more to a person than a concept of self. Arguably, however, the first premise captures something of this 'more' by including 'conscious of'. If this partly means 'capable of experiencing', then computers are not persons, even if the definition is acceptable: If having experiences is a necessary condition of being a person, and computers do not have any experiences then computers are not people.

The conclusion of the Third Paralogism follows from the first and second premises, but this is rather an empty inference because the problematic step has already been taken for granted: that the soul fits the description of 'person' given in the first premise. If disembodied existence is coherent, a rational psychology of the soul as a person seems not incoherent. (2)

Three Kinds of Identity: Kant's Attack on the Third Paralogism

To understand and appraise Kant's 'critique' of the Third Paralogism we need to make sense of a distinction he draws at A362-A33, CPR 341-2 between three sorts of case in which the question of identity over time might arise. These are the cases of a physical object, oneself, and a person other than oneself.

Physical Objects

Kant's term at CPR 341, A362, is 'an external object' (*eines äußeren Gegenstandes*) but I shall take it he is interested in physical objects other than the subject's own body. He makes an epistemological comment about how we should establish the identity of a physical object at a later time t^2 with a physical object at an earlier time t^1:

'If I want to know through experience, the numerical identity of an external object I shall pay heed to that permanent element in the appearance to which as subject everything else is related as determination, and note its identity throughout the time in which the determination change.' (CPR 341, A362)

'Wenn ich die numerische Identität eines äußeren Gegenstandes durch Erfahrung erkennen will, so werde ich auf das Beharrliche derjenigen Erscheinung, worauf, als Subjekt, sich alles Übrige als Bestimmung bezieht, Acht haben und die Identität von jenem in der Zeit, da dieses wechselt, bemerken.' (A362)

This means that if I wish to establish the identity of some physical object ' *x* ' between t^1 and t^2 then this may be achieved by the uninterrupted observation of *x* between t^1 and t^2 and, in particular, by paying attention to an aspect of *x* that does not change between t^1 and t^2, despite changes in *x* 's properties. This is a method of verifying or falsifying answers to the question 'Is the physical object at t^1 the physical object at t^2?'

Kant identifies the 'permanent element' in the object with the subject of properties, or that which bears the object's properties. There is a danger in this, because on a well known definition of 'substance' a substance is just the bearer of an object's properties. The problem is that the subject or substance in this sense is in principle unobservable. Kant needs to be able to observe the unchanging element so it will not do to identify it with what cannot be observed. But Kant need not be taken

this way. Other accounts of what does not change over time may be provided, which do not commit him to a quasi-Lockean substratum. For example, it might be that if a physical object is not subject to fission or fusion then it possesses a spatial unity over time which is an observable constant which could exist despite changes in primary or secondary properties (shape, hardness, colour, etc). Kant could then retain his identification of the observable constant with the 'subject' of properties.

Although Kant's account is epistemological, it presupposes a specific criterion for the identity of physical objects over time:

An object ' *x* ' at t^1 is numerically identical with an object at t^2 if and only if there exists an object with a single unified spatio-temporal history between t^1 and t^2.

I intend 'single' and 'unified' to capture the most plausible construal of Kant's 'permanent element' in the object. The spatio-temporal continuity criterion is presupposed in the sense that a person who successfully executed the verification procedure Kant recommends would thereby have verified the proposition '*x*, an object at t^1, is numerically identical with *this* specific object at t^2'. Although a verification procedure is not exactly the same thing as a criterion for the truth of a statement, there obtains a close conceptual dependence between the two. On the one hand, there exists a distinction between what the truth of a statement consists in and the methods by which we may discover that truth. On the other hand, the verification procedure works only if some specific criterion for the truth of the statement open to verification exists. So when Kant says 'If I want to know [...]' at CPR 341, A361, we should take it that he thinks the procedure he recommends will allow us to know the truth of identity statements about physical objects. This follows if we assume, uncontroversially, that if *p* is known then *p* is true.

Oneself

Kant thinks there is an important difference between the sort of logical status possessed by one kind of third-person statement about the identity of physical objects over time, and one kind of first-person statement about one's own identity over time. The first are informative (and thus presumably empirical, synthetic, and *a posteriori*), the second are merely tautologous (and so non-empirical, analytic and *a priori*). He has an argument at CPR 341, A362 designed to establish that 'I am numerically identical over time' falls into this second logical category:

(1) 'I am an object of inner sense.'

(1) Nun aber bin ich ein Gegenstand des innern Sinnes.

(2) 'All time is merely the form of inner sense.'

(2) Alle Zeit ist bloß die Form des innern Sinnes.

(3) 'Consequently I refer each and all of my successive determinations to the numerically identical self, and do so throughout time, that is, in the form of my inner intuition of myself.'

(3) Folglich beziehe ich alle und jede meiner sukzessiven Bestimmungen auf das numerischidentische Selbst, in aller Zeit, d.i. in der Form der inneren Anschauung meiner selbst.

(4) '"The personality of the soul" says nothing more than that in the whole time in which I am conscious of myself, I am conscious of this time as belonging to the unity of myself.'

(4) Denn er sagt wirklich nichts mehr, als in der ganzen Zeit, darin ich mir meiner bewußt bin, bin ich mir dieser Zeit, als zur Einheit meines Selbst gehörig, bewusst.

(5) 'It comes to the same whether I say that this whole time is in me, as individual unity, or that I am to be found as numerically identical in all this time.'

(5) Es ist einerlei, ob ich sage: diese ganze Zeit ist in Mir, als individuelle Einheit, oder ich bin, mit numerischer Identität, in aller dieser Zeit befindlich.

(6c) '"The personality of the soul" has to be regarded not as inferred but as a completely identical proposition of self consciousness in time; and this indeed is why it is valid *a priori*.

(6c) 'Auf diesen Fuß müßte die Persönlichkeit der Seele nicht einmal als geschlossen, sondern als ein völlig identischer Satz des Selbstbewußtseins in der Zeit angesehen werden, und das ist auch die Ursache, weswegen er a priori gilt.

I take it (1) is analytic because all I can ever hope to intuit through inner sense is myself, or some property of myself.

(2) is also true by definition. If we accept these two claims as conclusions of the Transcendental Aesthetic then we may accept them as premises here.

(3) means that if I am acquainted with some aspect of myself through inner sense then I ascribe any such properties of myself to myself as a being who endures over time (and, by implication, despite gaining or losing properties). The second conjunct

of (3) asserts that this process of inner intuiting and self-ascription is itself a temporal process.

In (4) the expression 'the personality of the soul' is elliptical for the argument of the First Paralogism. We could think of the third paralogism as a complex sentence embedded in (4) where 'the personality of the soul' features. Thus (4) gives a specific interpretation to the Third Paralogism. This interpretation is a reductivist one, in two ways. Firstly, 'soul' with all its dualist metaphysical connotations is reduced to 'self', one component of 'soul', but one which could on Kant's view have a use without the missing connotations. Secondly, 'nothing more than' implies that we are to restrict the meaning of the Paralogisms to just 'the whole time in which I am conscious of myself'. This is reductivist if it means 'just those times when I am occurently aware of myself in inner sense'. It is less reductivist if it means 'any stretch of time during which there occur introspective episodes through inner sense'. We could call this non-reductivist reading the 'dispositional' reading and the other the 'occurrent' reading. On the occurrent model, there is consciousness of time 'as belonging to the unity of (the) self' just so long as there lasts a specific act of inner sense. While there exists some mental state, but no self-conscious state, there is no consciousness of time as belonging to the unity of the self. On the occurrent model there exist self-conscious states only intermittently. On the dispositional model, there is perpetually consciousness of time as belonging to the unity of self. The first is occurrent because time consciousness occurs only and always when inner sense is exercised. The second is dispositional because there exists inner time consciousness so long as there is a disposition to self-consciousness. Indeed a dispositional interpretation may be given to inner time consciousness also: Just so long as there exists a disposition to the exercise of inner sense there exists a disposition to inner time consciousness. Although the time-consciousness disposition exists just so long as the disposition to inner sensing also exists, the disposition is only intermittently and not perpetually exercised.

What does this inner time consciousness amount to? In other words, what does this sentence mean: 'I am conscious of this time as belonging to the unity of myself'? Suppose 'I' is a word each person uses to refer only to himself. (I say 'a' not 'the' because other words are also used in this way: 'me' for example.) 'Conscious of' here could mean 'knows that' or else it could mean 'has direct experiential acquaintance with'. If it is equivalent to 'knows that' then if I am conscious of (this) time as belonging to the unity of myself then (depending on the analysis of 'know') I believe this, have reasons for believing it and it is true, or, perhaps posses this information in a waqy that is not belief entailing. If I am experientially acquainted with time as belonging to myself then this acquaintance could in principle (presumably) be non-veridical if, for example, non-Russellian. Next, what does 'this' refer back to in 'this time'? It refers back to 'the whole time during which I am conscious of myself'. This again admits of a dispositional or an occurrent analysis. If we read it dispositionally, then it means 'the whole time I have ever had a capacity for self-consciousness' (say, from an early age). If we read it occurrently, it means 'just that time I am in any self

conscious state, and only so long' (say, for a minute or so). By 'unity of myself' here I shall just understand the formal unity which is a condition for experience (above) so *x* is part of the unity of (myself) if and only if *x* logically could not exist unless a part of me, and I logically could not exist if *x* were not part of myself.

Now we may ask: Is (4) true? Taken one way, it does not seem that it possibly could be because it makes a person's acquaintance with a certain conclusion of the Transcendental Aesthetic a condition of self-consciousness. A weaker but more plausible claim would be merely to claim the truth of that doctrine as such a condition. Its putative truth needs to be examined in two stages: by deciding whether the Third Paralogism really 'says nothing more than' what is expressed by the sentence introduced by 'that', and by deciding whether that sentence is itself true. I'll take the second first as it is the epistemically prior question. If it is intended to be an empirical claim about 'I' users then it is unlikely to be true that 'in the whole time in which I am conscious of myself, I am conscious of this time as belonging to the unity of myself', whether 'whole time' [is] us given an occurrent or a dispositional reading. This is because there exists little empirical evidence that 'I' users (outside philosophical circles) think much about the relation between time and the self. But perhaps this is not quite what Kant intends. A plausible interpretation is: When I am conscious of myself, part of what I am thus conscious of is time as belonging to the unity of myself. Kant's criticism is overly reductivist when he says the Third Paralogism asserts nothing more than his (transcendental idealist) interpretation, because in (5) he will go on to claim that both interpretations are plausible, indeed complementary. If that is right, then the whole of (4) cannot be true if (5) is true, even if the sentence introduced by 'that' in (4) is true. Independently of this, the truth of (4) depends on the plausibility of each of the two interpretations of (4), and their mutual consistency. But those questions are best addressed by turning to (5).

(5) is the claim that these two sentences 'come to the same':

(i) 'This whole time is in me.'

(i) diese ganze Zeit ist in Mir

(ii) 'I am to be found as numerically identical in all this time.'

(ii) ich bin, mit numerischer Identität, in aller dieser Zeit befindlich.

If it 'comes to the same' which is asserted, then this might mean that (i) and (ii) (are logically equivalent but *prima facie* it looks implausible that (i) is true if and only if (ii) is true, and (i) is false if and only if (ii) is false. Even more extravagantly, 'comes to the same' might not just mean that (i) and (ii) have always the same truth values, but might also imply that (i) and (ii) are synonymous. (I bracket Quinean concerns about synonymy and just assume here that the logical equivalence of two propositions is not a sufficient condition of their synonymy, even though synonymy

is logically sufficient for logical equivalence.) More plausibly, 'comes to the same' could mean at least that a proposition with the same truth value is expressed whichever of (i) and (ii) is asserted, even though there exist comparatively surface or superficial semantic differences between (i) and (ii). The best way to decide Kant's meaning here is to examine the logical and semantic relations between (i) and (ii) and then put the most plausible construal on 'comes to the same'.

Suppose (i) and (ii) entail respectively:

(i) 'Time is in me

(ii) 'I am in time

and these 'amount to the same'. If this is part of what is expressed by (5) then some sense may be attached to it deploying two distinctions: the transcendental idealism-empirical realism distinction, and the subjective-objective distinction. Then 'Time is in me' is a sentence construed as expressing two propositions: a true proposition of transcendental idealism and a false proposition of empirical realism. 'I am in time' expresses a true proposition of empirical realism and a false proposition of transcendental idealism. Kant's disambiguation is only as strong as his entitlement to his distinction: So long as we do not confuse empirical realism and transcendental idealism then there is no consistency between the two sentences. We are compelled to accept 'Time is in me' to make sense of the possibility of our experience. We are compelled to accept 'I am in time' by the way things appear empirically. But if 'comes to the same' means 'logically equivalent', then it is clear that they do not 'come to the same':

'Time is in me.'	'I am in time.'	
T	F	Transcendental Idealism (TI)
F	T	Empirical Realism (ER)

So if (1)' is true [if and only if (as it is in the model above)] and (11)' false, or if (1)' is false and (ii)' true then (i)' and (ii)' cannot possibly be logically equivalent. But if it is true that synonymy is a sufficient condition for logical equivalence then logical equivalence is a necessary condition for synonymy. But if (i)' and (ii)' are not logically equivalent then they cannot be synonymous either. So on the transcendental idealist-empirical realist interpretation (i)' and (ii)' do not 'come to the same' whether this means 'are logically equivalent' or 'are synonymous'.

Although we have just deduced the non-synomymy of (i)' and (ii)' from their logical non-equivalence, there are independent grounds for thinking them non-synonymous. Firstly, even if two sentences turn out to be logically equivalent in truth

functional logic this is not a sufficient guide to their synonymy or lack of it interpreted semantically because, clearly, many non-synonymous pairs of sentences have the same truth value. I offer an independent argument for the non-synonymy of (i)' and (ii)' in case the Kantian thinks an argument can be mounted to show that (i)' and (ii)' are logically equivalent after all. There are strong logical reasons for doubting that possibility, but any Kantian will be motivated to try it because any Kantian believes transcendental idealism and empirical realism are mutually consistent. But this course is not open to the Kantian:

'Time is in me' is true in (TI)

'I am in time' is true in (ER)

Therefore 'Time is in me' and 'I am in time' are logically equivalent and

'I am in time' is false in (TI)

'Time is in me' is false in (ER)

Therefore 'I am in time' and 'time is in me' are logically equivalent. They may also be read so that (i)' and (ii)' have opposite true values: One is true only when the other is false. It follows that either (TI) and (ER) are wholly independent languages in which each of (i)' and (ii)' do not have the same semantic functions, or else (TI) and (ER) are mutually inconsistent. Neither of these suits Kant's purpose. For the two sentences to be either logically equivalent or synonymous across (TI) and (ER) each must have the same semantic and logical role in (TI) as (ER). On the other hand, if each sentence means the same whether it features in (TI) and (ER) then there exist at least two mutually inconsistent pairs of sentences across (TI) and (ER). Then (TI) and (ER) are mutually inconsistent. If so, then transcendental idealism and empirical realism are mutually inconsistent because two languages [L' and L''] are mutually inconsistent if and only if there is at least one sentence that is true in L' that is false in L" or at least one sentence that is false in L' that is true in L''. If that is right then at least one of the sentences in one of transcendental idealism or empirical realism is false, because if two sentences are mutually inconsistent then that is a sufficient condition for the falsity of one of them.

To show that t(i)' and (ii)' are not synonymous it suffices to show that they have different verification conditions, for example if the means of verifying one could not be used to verify the other. Construed empirically, 'I am in time' is verifiable through outer sense, or by a combination of inner and outer sense. Also, construed empirically, 'Time is in me' is verifiable through inner sense but not outer sense. Construed transcendentally 'Time is in me' is verifiable by taking a philosophical decision about the intelligibility of our experience based on the logic of transcdental arguments, and 'I am in time' is similarly verifiable (or falsifiable) according to

whether it has to be assumed to make the experience of a self-conscious rational being intelligible. If two sentences have different verification procedures then that is a sufficient condition for their differing in meaning. So, if it is true that (i) and (ii)' 'come to the same' then this cannot mean they mean the same.

Is there any interpretation on which the two sentences come out as synonymous, or at least with the same truth values? There is one on which they are mutually consistent. To show that the truth or falsity of one does not preclude the truth or falsity of the other, we need to draw a subjective-objective distinction:

A sentence is *subjective* if an only if it is verifiable only by reference to an inner state of the subject.

A sentence is *objective* if and only if it is verifiable only by reference to what is outer ie, no part of any subject's mental state.

Then we obtain:

I am in time	Time is in me	
F	T	Subjective
T	F	Objective

The truth values are distributed in this way because 'I am in time' has no subjective use, and 'Time is in me' has no objective use. Now we may take 'comes to the same' in a new way. If I say 'Time is in me' then I am expressing a fact about how time *appears* (to be) subjectively. If I say 'I am in time' then I am expressing a fact about how I think of myself objectively. On the subjective conception of time, I think of it indexically. [Time is subjective; the present, the time when my experiences happen, as, past as what happened before them, as future, as what will happen after them.] On the objective conception of time I think of myself as an enduring spatio-temporal particular located in a chain of events some of which are antecedent to my existence, some simultaneous with it and some subsequent to it.

But why should these two ways of thinking 'come to the same'? I think the solution is that unless one were at least thinkable, as embedded in (5), then the other would not be thinkable. Unless my mental states had a temporal 'unity', took place in a relatively non-chaotic sequence ordered by a ‘before-after’ relation and a ‘simultaneous-with’ relation, then there could be no talk of a single, unified, self-same consciousness that is mine. But this conception is needed if 'I' is to obtain a grip on the second part of (5) where it is asserted that I am to be found as numerically identical in time. The temporal unity of the self is, on this view, a condition for its numerical identity across time

But if (6c) is true, then 'thinkable' in the previous paragraph has to be read strongly. This is because Kant thinks

'I am numerically identical over time'

is analytic as putatively obtained from (5) and the other premises. So, unless I was a temporally unified self, unless the sentence describing that feature of myself were true, then 'I am numerically identical over time' could not be true either. Yet Kant says the 'personality of the soul' is ‘not inferred’. This means that it need not be regarded as the conclusion of an argument after all. This seems reasonable to the extent that it is a 'completely identical proposition', because if a sentence expresses a tautology, or is analytic, then its truth may be determined by an examination of the terms it contains, independently of the truth or falsity of other sentences.

Despite this, 'I' has to be given a special reading in (6c) to in make 'I am numerically identical over time' analytic, and this reading is spelled out by (1) to (5). In that sense, (1) - (5) are an argument for (6c): they tell us to read 'I' as 'unified temporally ordered self'. We know from the Transcendental Aesthetic that Kant thinks that there is only one time: Putatively numerically distinct times will turn out to be temporally related and so be part of one and the same time. Using this principle, it has to be the case that the time that is 'in me'; the unified time of my subjective mental state, is not numerically distinct from the time that 'I am in'; the objective time in which my body and other objects are located. And there seems no reason why they should be numerically distinct. It makes sense to say (as we do) that some subjective mental state of mine was preceded by, or simultaneous with, or followed by, some objective occurrence in the 'external world'. So, it is perhaps rather misleading to talk about two sorts of time: subjective time and objective time, or inner time and outer time. Rather there are two sorts of state or occurrence: subjective and inner, and objective and outer, but these two sorts of occurrence take place in one and the same time. It is this single time that gives unity to my mental states and is amongst the conditions for them being called 'mine' and which provides the temporal order in which I and my states am located.

Does this make 'I am numerically identical over time' a tautology? If 'I' is given the meaning derived from the premises (1) - (5) then it does. But there is another reason why it is tautologous: If *x* exists then it is a modal fact about *x* that *x* is numerically identical with *x*, whatever *x* is and however *x* changes. If it is a logically necessary condition of *x* 's existing that *x* have temporal properties, that *x* be 'in time', then it is a further modal fact about *x* that *x* is numerically identical over time if *x* exists. Thus sentences of the form ' *x* is numerically identical over time' will turn out to be analytic truths and if *p* is an analytic truth then *p* is a tautology, or as Kant puts it, a 'completely identical proposition'. On this interpretation, my identity over time is described by a sentence that is tautologous not because of any special features of the self, but because of conceptual facts holding for anything that has temporal properties, and general modal facts pertaining to any individual existent.

If 'I am numerically identical over time' is a tautology, then certain of Kant's additional conclusions follow directly. Part of (6c) is that the sentence is also 'valid *a priori*'. For 'valid' I shall just substitute 'true' here. If *p* is a (decidable) tautology, then it follows that *p* is *a priori*, because a tautology is decidably true just in virtue of the meanings of its consistent terms.

It follows, as Kant requires, that 'In my own consciousness, therefore, identity of person is unfailingly met with' (CPR 341, A362) partly because in inner sense I never come across any person, or any state of any person, other than myself, but also because I am numerically identical over time. If *per impossibile* this modal condition were not met, then the sentence 'I am my own consciousness etc' could not be true either. Given that it is met, and that I exist, in conjunction with a truth about what inner sense is, it follows that I never am directly aware of anyone other than myself 'in my own consciousness'. Kant intends this in a sense that does not preclude my being conscious of other persons, just my meeting them as aspects of or subjects of my consciousness. It is that that is logically impossible. (3)

Kant and Parfit on Personal Identity

Kant and Parfit each entertain a logical possibility about the identity of persons over time, and these deserve mutual comparison. Parfit explains his version through a science fictional illustration. It is couched in the first person singular in the original (Parfit, 1984, pp. 199-200) and runs like this: A machine called a 'Teletransporter' is used to transport me from earth to the planet Mars using the following method.

'I press a button on earth lose consciousness, and wake up […] on Mars. Meanwhile, the machine on earth destroys my body after recording all the information in its cells, it transmits this information to Mars where a Replicator creates a new body (using this information) which is qualitively identical to my old body on earth, but of course numerically distinct. I wake up in the new body, and cannot detect any change in it.

A variant on the story is that the old body on earth is not destroyed but the blueprint is taken from it for a replica to be built on Mars, then, if I wish 'I can use the intercom to see and talk to myself on Mars'' (Parfit, 1984, pp. 199).

Now, consider Kant's thesis that

'The identity of the consciousness of myself at different times is '[[...]] only a formal condition of my thoughts and their coherence, and in no way proves the numerical identity of my subject. Despite the logical identity of the 'I', such a change may have occurred in it as does not allow of the retention of its identity, and yet we may ascribe to it the same sounding 'I' which in every different state, even in one involving change in the (thinking) subject, might still retain the thought of the preceding subject, and so hand it over to the subsequent subject' (CPR 342, A363).

Es ist also die Identität des Bewusstseins Meiner selbst in verschiedenen Zeiten nur eine formale Bedingung meiner Gedanken und ihres Zusammenhanges, beweiset aber gar nicht die numerische Identität, meines Subjekts in welchem, ohnerachtet der logischen Identität des Ich, doch ein solcher Wechsel vorgegangen sein kann, der es nicht erlaubt, die Identität desselben beizubehalten; obzwar ihm immer noch das gleichlautende Ich zuzuteilen, welches in jedem andern Zustande, selbst der Umwandelung des Subjekts, doch immer den Gedanken des vorhergehenden Subjekts aufbehalten und so auch dem folgenden überliefern könnte.

There is a clear parallel and a clear contrast between Kant and Parfit here. The parallel is that both philosophers entertain the logical possibility that, in some sense, the same person may continue to exist despite a change in the numerical identity of the person's body (Parfit) or subject (Kant). What this amounts to in each case is similar: Parfit speculates that his replica might [believe he is he, that is the replica] believe himself to be identical with the person left on earth, in fact, to be that very same person but teletransported to Mars. Kant says the thoughts of each subject are passed in to the subsequent subject. So, what is preserved in both these examples is a certain kind of memory. Parfit's teletransported person remembers (or seems to remember) being the person on earth up to teletransportation. Kant's person remembers, or seems to remember, or seems to remember being, the predecessor subjects.If we call Parfit's teletransported person and Kant's subsequent subject a later self, the later self remembers, or seems to remember being, an earlier self if and only if the later self remembers 'from the inside' at least one experience or thought of that earlier self. I say that A remembers O if and only if: O happened, O is believed by A to have happened and the cause of O's being believed to have happened is that O was experienced to happen (directly or indirectly) by A. A remembers O 'from the inside' if and only if the above conditions hold but 'directly' is added. I say an experience of O is direct if O is either perceived or undergone. I say an experience of O is indirect if O is not perceived or undergone, yet some experience of perception causes a belief in the existence of O. Now, the question arises whether Kant and Parfit's later selves, because they remember earlier selves, 'from the inside' are the same person as the earlier person. To decide this, we need to be clearer about the roles of 'body' and 'subject'.

Kant means by 'subject' ‘that which experiences' or 'that which has experiences', and the body, in Parfit’s considerations, is one sort of thing which may fall under this Kantian description. So, then we have Parfit producing a *prima facie* plausible candidate for this role: the body. The question whether it is true that the body is what has experiences may be put on one side for the moment because we are interested in this question: If some owner of experience at t^1 is not *ex hypothesis* numerically identical with some owner of experience at t^2, then is it logically possible that a person claiming to exist because the owner at t^2 exists may be identical with a person which exists because the owner at t^1 exists? This question may be addressed quite independently of the question: What is the owner of experiences? For example, we

do not have to know whether the owners at t^1 and t^2 are physical objects or minds. (Parfit has in mind bodies, Kant perhaps minds if 'subject' means 'mental subject here).

The asking of this question does not require a sophisticated or very rich concept of a person. Kant and Parfit work with rather a minimalist one in these passages. On Kant's and Parfit's quasi Humean ontology, a person is a set of experiences at least some of which seem to be memories. Also involved here in the concept of a person is the consciousness of, or at least the belief in, one's continued persistence over time. So, a person counts as such not just in virtue of experiencing, or even remembering but also in believing himself to persist over time (or seeming to remember himself as persisting over time). 'Person' means 'that which experiences, whatever that is'. This is because if we identified the person with the subject in Kant's case, then it would come out as trivially false that that person was numerically identical with some subject at t^2 if identical with some subject at t^1, if *ex hypothesis* the subject at t^1 is numerically distinct from the subject at t^2. Similarly, if we identify the person with the body in Parfit's case then it comes out as trivially false that that the person is numerically identical with the body at t^2 if, *ex hypothesis*, the body at t^2 is numerically distinct from the body at t^1.

There is a particularly acute point to Kant and Parfit's formulation: If we answer 'Yes' to the original question, if it is true to say that the numerical identity of the owner of experiences over time is not a necessary condition for the persistence of the person over time then one particular remark about the nature of personal identity is false. This is Butler's remark that far from its being the case that memory constitutes personal identity, it presupposes it. Kant and Parfit have developed a senario whether by it might *prima facie* be coherently supposed that memory might be at least partly constitutive of the identity of the person over time, but where this does not presuppose an enduring memory of a physical subject of experience. However, identity, rather than continuity requires something utterly unchanging between t^1 and t^2. Anything physical or psychological is changing. The unchanging is the soul.

Whether the minimal person at t^2 is numerically identical with the minimal person at t^1 is not wholly a matter of stipulation. As Parfit says, if not a criterion, then at least a sound intuition about the answer may be obtained by asking: Would I survive? Suppose you are a minimal person who exists at t^2, associated with a subject numerically distinct from the subject at t^1. Then one's intuition is that one would be numerically identical with the person at t^1, if the person at t^2 could remember the experiences at t^1 'from the inside'. This is because 'experiencing from the inside' is a component of our pre philosophical concept of a person, and specifically the person who one is. Suppose at t* I remember doing something at t^1, in the non-question begging sense of 'remember doing something at t^1', then arguably, it is impossible for me not to identify myself with the person at t^1. This is because it does not make much sense to say: I remember that pain but I doubt it was I who felt it. This perhaps makes some sense: I remember that someone was in pain, but not who was in pain. It is in principle possible for someone to remember that

there was a pain but not whether it was himself or another who felt it. This is no doubt very rare but is not in principle impossible. For example, the pain might be the cause of some event which looms so large in the memory as to render irrecoverable the answer to: Whose pain was it? But suppose now the person remembers from the inside. Then that person remembers the quality of the pain: what it was like to have it. This memory of what the pain was like, if veridical, may only be accessed from a first person singular point of view so, in some sense, if I remember from the inside some experience then I am the owner of that experience. Or, at least, I would find it psychologically impossible to disbelieve that I am the owner because it would be manifestly false to assert that one remembered an experience from the inside, but also that one never had it.

The Objective Critique of Personality: Kant on Oneself and Other People

At CPR 341, A362 Kant works with a tripartite distinction, the last part of which logically presupposes the first two. It is:

1. self-consciousness

2. consciousness of another person

and

3. thought about oneself 'from the standpoint of another person' but by oneself.

He could have added for the sake of completeness:

4. thought about another person 'from the standpoint of oneself' by oneself

and for the sake of precision drawn a distinction between

(SCIS) self consciousness in inner sense

and

(SCOS) self consciousness in outer sense.

Both of these additions are to his immediate purpose. Kant uses this distinction to present an argument for the conclusion that one person may conclude that a person at t^2 is identical with a person at t^1 without thereby being committed to the thesis that there is a permanent or enduring element in the person which sustains this identity. The argument is:

(1) 'If I view myself from the stand point of another person (as object of his outer intuition), it is this outer observer who first represents me in time'

(1) Wenn ich mich aber aus dem Gesichtspunkte eines andern (als Gegenstand seiner äußeren Anschauung) betrachte, so erwägt dieser äußere Beobachter mich allererst in der Zeit.

(2) '(for) in the apperception time is represented, strictly speaking, on in me'

(2) denn in der Apperzeption ist die Zeit eigentlich nur in mir vorgestellt.

(3) 'He admits therefore the 'I' which accompanies, and indeed with complete identity, all representations at all times in my consciousness.'

(3) Er wird also aus dem Ich, welches alle Vorstellungen zu aller Zeit in meinem Bewusstsein, und zwar mit völliger Identität, begleitet, [ob er es gleich einräumt, doch noch nicht auf die objektive Beharrlichkeit meiner Selbst schließen.]

(4) 'The time in which the observer sets me is not the time of my own but of his sensibility'

(4) Denn da alsdenn die Zeit, in welche der Beobachter mich setzet, nicht diejenige ist, die in meiner eigenen, sondern die in seiner Sinnlichkeit angetroffen wird …

(5) 'The identity which is necessarily bound up with my consciousness is not bound up with his, that is, with the consciousness which contains the outer intuition of my subject'

(5) […] so ist die Identität, die mit meinem Bewusstsein notwendig verbunden ist, nicht darum mit dem seinigen, d.i. mit der äußeren Anschauung meines Subjekts verbunden.

(6C) 'He will draw no inference from this (3) to the objective permanence of myself' (CPR 341-2, A362-3)

(6C) [ob er es gleich einräumt, doch noch nicht auf die objektive Beharrlichkeit meiner Selbst schließen.]

In (1) Kant need not have used the restrictive 'myself', and 'I'. 'Oneself' and 'one' would have done just as well because he is concerned with an asymmetry between the sort of knowledge a person may have of his or herself, and the sort of knowledge

another person may have of that person. We may call this asymmetry a difference between a first and third person perspective on the person (although in this case of first person knowledge this metaphor is potentially misleading).

Then (1) expresses this thought: If one person observes another then this is achieved through the exercise of outer sense, so that person's experience of the other is necessarily spatial. But that person's sense experiences are also necessarily temporally ordered and, in particular, those sense experiences are (re)presentations of the observed person which stand in temporal relations to other experiences of the observer. They are, I take it, necessarily either before, after or simultaneous with some other experiences of the observer.

Now, this sort of observation though outer sense is to be contrasted with the sort of thought about time which is possible only through apperception, and this is necessarily first person singular. That is the force of (2). (2) precludes the possibility of any third-person use of apperception. That is why Kant puts in 'only in me', meaning 'only in the subject' *qua* subject. Also precluded, *a fortiori*, is the possibility of a person using apperception to detect the states of another person. That role can only be performed by outer sense. Indeed, any putative observer could not become acquainted with any mental or temporal state that was not one of his own through the use of apperception.

Despite this, and quite consistent with it, (3) is the claim that any person observing another holds the central fact contained in the doctrine of pure apperception to be true of the other, *viz*: the formal truth that all the other's experiences are the other's. This extrapolation from knowledge which is primarily first person to a third person case does not follow from (1) and (2) in the way that Kant's use of 'therefore' in (3) suggests. But it is made possible by the truth of (1) and (2) and by the doctrine of pure apperception. A corrolary of this extrapolation is that that observer assumes that one and the same subject of experience is the object of that observer’s outer sensing of the other.

(4) is the thesis that, as object of a person's outer observations, a person *qua* object of outer observation is located in the time order of the observer's experiences, rather than in the time order of his own experiences. So, if a person is observed by another then the observed person is perceptually presented to that person before after or during other perceptions made by that person.

(5) introduces a distinction between the identity of two consciousness; that of the observer and that of the observed. The implication is that no conclusions about the unity of consciousness of a putative person may be logically derived from any set of sentences describing the unity of consciousness of some person other than that person. In particular, from the fact that a putative person is the object of another's perceptions it does not follow that that putative person is a unified consciousness, numerically identical though the time the person is observed.

This gives Kant the conclusion, (3), that from the fact, the purely formal fact, that all a person's experiences belong, logically, to that person it cannot be validly deduced that there exists an enduring single subject of experience that persists during

the whole time a person is observed. Kant thinks, *a fortiori*, it cannot be validly concluded that a person is essentially a soul. Kant thinks there is no need to postulate a subjective substance to account for the observations of a person over time being of a self-same person over that time. I am inclined to disagree. If the other is an enduring person the subjectivity of the other endures and the soul is enduring subjectivity.

The way Kant couches the conclusion (6C) does not preclude the possibility of a spatio-temporal criterion for the identity of a person over time for the person *qua* the object of another person's observation. Because the 'objective permanence of myself' cannot be logically obtained from the fact of a person's observing another over time it does not follow that that putative person does not possess the sort of spatio-temporal unity exhibited by physical objects. Indeed, two considerations suggest any such putative person does possess such unity. First, if a person is an object of outer sense then it is hard to see how *qua* object of outer sense such a person could appear to be anything but a physical object (or physical event, or constellation of space-time processes etc.) *viz* a human body. Secondly, the way Kant has devised the problem, built into it is the idea of a person perceiving another over a period of time. This presupposes there is something the observed body's identity over time consists in. What it does not presuppose is that a particular account be given of that observed body's belonging to a subjective unified consciousness with an enduring psychological history. Kant does not spell out what the criteria are for the numerical identity of a human body over time, but this is perhaps because he rather assumes the reader will not take them to be different from those for any other physical object. But the conclusion (6C) trades on the difference between two sorts of criterion for the identity of person over time, one objective and physical, the other subject and mental. That the first sort is met if a putative person, in this case a human body, is observed over time is no guarantee that the second is met. Kant is quite prepared here to allow a logical gap between a subjective and an objective mode of conceiving the person. We could make this explicit in this way: From no sentence or set of sentences which truly describe the continued existence between t^1 and t^2 of a physical object with the exact shape of a human body observed by a person can there be logically derived any sentence, or set of sentences, which truly ascribe a unified subjective consciousness to that object between t^1 and t^2. We could call this the ‘irreducibility of the subjective to the objective’.

We should not be misled by Kant's use of 'objective' in (6C) in 'objective unity of myself'. To understand this use, suppose I am observed by another person (ie, I am an object of his/her outer sense). Then, my unity is 'objective' just in case it exists or endures independently of the other's perceptions of me, ie: It exists whether or not the other perceives me, and it exists however his observations of me might change. This sense of 'objective' is not the same as that in which some entity may be correctly termed 'objective' if it is an object of (veridical) outer intuition. There is value in Kant's using 'objective' in (6C), however, because if the argument for (6C) is sound then Kant has established there is no objective criterion for the identity of a person

(ie a unified self, not just a human body) over time that is available to the third person perspective. By 'available' here, I mean 'that may be deployed to conclusively verify or falsify claims about the continuing identity of the person'. Because it is deployed to criticise what is available from a third person point of view it is appropriate to call the argument the Objective Critique of Personality. Kant has another, more radical critique of the Third Paralogism, and this makes use of the first person 'perspective' on the person only. I call this the Subjective Critique of Personality.

The Subjective Critique of Personality

(1) 'The identity of the consciousness of myself at different times is [[...]] only a formal condition of my thoughts and their coherence'

(1) Es ist also die Identität des Bewußtseins Meiner selbst in verschiedenen Zeiten nur eine formale Bedingung meiner Gedanken und ihres Zusammenhanges [beweiset aber gar nicht die numerische Identität, meines Subjekts in welchem, ohnerachtet der logischen Identität des Ich, doch ein solcher Wechsel vorgegangen sein kann, der es nicht erlaubt, die Identität desselben beizubehalten]

(2) 'The same sounding 'I' which in every different state, even in one involving change in the (thinking) subject, might still retain the thought of the preceding subject and so hand it over to the subsequent subject'

(2) Obzwar ihm immer noch das gleichlautende Ich zuzuteilen, welches in jedem andern Zustande, selbst der Umwandlung des Subjekts, doch immer den Gedanken des vorhergehenden Subjekts aufbehalten und so auch dem folgenden überliefern könnte.

(3) 'Despite the logical identity of the 'I', such a change might have occurred in it as does not allow of the retention of its identity.'

(3) [...] ohnerachtet der logischen Identität des Ich, doch ein solcher Wechsel vorgegangen sein kann, der es nicht erlaubt, die Identität desselben beizubehalten.

(C) '[(1)] in no way proves the numerical identity of my subject.' (CPR 342, A363)

(C) [(1)] beweiset aber gar nicht die numerische Identität, meines Subjekts.

The first premise, (1), reiterates Kant's clarification of the doctrine of the transcendental unity of apperception as a purely formal condition of there existing

persons. It is worth making explicit the thoughts expressed by this particular version of it.

It entails, firstly, that when I am self conscious over time (and we may read this dispositionally or occurrently here), if I am conscious at two times, t^1 and t^2, at t^1 I am identical with that which is conscious at t^2. This is clearly not to assert that that which has the property of being conscious at t^1 has the property of being conscious at t^2 at t^1, nor that that which has the property of being conscious at t^2 has the property of being conscious at t^1 at t^2. It is just the claim that that which is conscious at t^1 is what is conscious at t^2.

Secondly, Kant is making an additional identification: That which is conscious at t^1 is what its consciousness is of at t^1, and that which is conscious at t^2 is what its consciousness is of at t^2. If not, we could not talk of self consciousness.

There is a third thought which is not made explicit by Kant here but which is entailed by the first two and which has to be true if the doctrine of apperception is to work: That which is what consciousness is of in self consciousness at t^1 is identical with that which consciousness is of at t^2. If not we could not talk of 'myself'. When he says this in only a formal condition of coherent thought he means we should draw no ontological conclusion from what is a conceptual or logical truth about persons, or an unpacking of some of the logical of 'person'. Kant does not realize that his claim, despite his insistence on the purely form, is blantantly ontological. Kant, in common with many thinkers about personal identity needs to take seriously the notion of identity. Identity is a much stronger relation than continuity. If what is at t^1 is what is at t^2 then something is utterly unchanging. Anything mental or physical is changing so the only plausible candidate is the soul.

The second premises, (2), is designed to show that the formal doctrine of experience, expressed by (1), is consistent with an ontology of the person other than that recommended by the rational psychologist: There does not have to exist an enduring subject of experience and, *a fortiori*, there does not have to exist a soul, between t^1 and t^2 in order for 'I' to have a use for the same person between t^1 and t^2. This is because persons might be episodic, that is, they might consist in a series of mental episodes or states where none of these is a state of a persisting mental substratum. If we then ask what the continuation of such a series of episodes consists in, what makes it a series constitutive of a person, or indeed a mental series at all, then Kant's answer is to invoke a memory criterion. Consider two mental episodes e and e', and suppose e obtains at some time t^1 and e' obtains at some later time t^2, then Kant's view is that part of what e' may consist in is a memory of e. Further, any subsequent mental episode e'' may partly consist in a memory of either e or e'. This is what Kant means by thoughts being ‘handed over’. Now, this account needs supplementing if it is to suit Kant's purpose. Clearly, the fact that any countable set of episodes takes place does not ensure that they are a series, because simultaneous episodes could be individuated. Nor is it enough for Kant's purpose to merely restrict his consideration to episodes that are only temporally related to one another by a before/after or earlier/later ordering, because one mental episode might be

chronologically subsequent to another yet they might each belong to a different person. I take it Kant is aware of both these points and one reason the memory criterion is invoked is to try to meet them. But even if the occurrence of a later episode *e′* partly consists in the remembering of an earlier episode *e* it might still be the case that *e* and *e′* are episodes in numerically distinct persons. This is because it makes sense to talk of one person remembering the experience another person had. Kant needs a distinction between two sorts of memory: remembering from the inside, and remembering from the outside. There is a danger of circularity here, but the distinction between the two sorts of remembering may be stated using 'person' just for purposes of clarification: A person remembers an experience from the inside only if that persons had that experience, but a person remembers an experience from the outside only if that person did not have that experience. By 'remembers' I mean: a person A remembers *x* if and only if A is caused to think a true sentence about *x* and is thus caused by the occurrence of *x* , where the occurrence of *x* is chronologically antecedent to the thinking of the sentence. (If memories can be non-veridical then we can allow for this by inserting 'or false' after 'true'). What Kant needs is a version of 'remembering from the inside' that does not use 'person' with circularity. To show that is possible I use two notions; that of 'remembering what it was like', and that of 'remembering that'. If *e′* is partly the memory of what *e* was like then *e′* is partly a memory of *e* from the inside. But if *e′* is partly a memory that *e* occurred but is not and could not be a memory of what *e* was like then *e′* is partly a memory from the outside. We could call these 'subjective' and 'objective' modes of remembering. The distinction provides Kant what he needs so long as the enduring substratum concept of a person does not have to be invoked to draw the subjective/objective distinction used here. It is between what is epistemologically private to a state, ie the phenomenological quality of an earlier state, and what is epistemologically inaccessible to a state, the quality of a different state yet where knowledge of the bare occurrence of that last state is open to the first state. (4)

It seems to me the enduring substratum concept of person does ultimately have to be invoked to draw this distinction. (5) This is because my experience of my own mental states *qua* my own, including my memories of my earlier states, requires the truth of the conclusions of rational psychology. Kant's ‘passing on memories’ account depends on the existence of the soul for the identity of who the memories are passed by and to. Also, one of my states could not consist in a memory of a state that was not mine, ie be the memory that it happened, without its thereby being true that that state had a substantial owner. I prefer 'episode' to 'state' because part of the logic of 'state' is 'state of' but this need not beg any questions, nor need the use of 'I' here.

If Kant's alternative model of personal identity works, then (3) follows from (1) and (2). *Prima facie* (3) is illegitimate because it may be read as entailing a contradiction. It can easily be read however so that the contradiction is only apparent. The problem is that Kant has asserted that 'the logical identity of the 'I'' obtains, yet the 'retention of its identity' is not 'allow(ed)', which *prima facie* suggests what is logically impossible, *viz* that something be the same and different in the same

respect. This problem disappears if we invoke Kant's purely formal reading of 'I', and use that to interpret 'the logical identity of 'I', but then give an ontological reading to 'its' in 'its identity'. Then we obtain the claim that the formal identity may be retained whilst the ontological identity of what the word 'I' putatively denotes is not retained. That is what Kant intends, and it clearly yields the conclusion he wants but I contest, (6C), because that is the claim that, ontologically speaking, I need not be numerically identical over time, that is, there need not be any substance which endures between t^1 and t^2 in order for it to be true that I endure between t^1 and t^2. But: who endures? Kant is not able to dispense with identity because if I endure between t^1 and t^2 something t^1 is identical with something at t^2. In particular, Kant thinks the theory of pure apperception, if true, does not commit one to the postulation of any such substance. But: Who thinks 'I think'? So, even if we accept the various supplements and interpretations given above, it does not have to be admitted that the Subjective Critique of Personality is successful against rational psychology. Clearly it does not prove that people do not have souls. It does not show that one sort of reason for thinking that they do is not a good one.

The Third Paralogism in the Second Edition

Kant's argument against the rationalist proof that the identity of a person over time consists in the identity of a soul over time is:

(1) 'The proposition, that in all the manifold of which I am conscious I am identical with myself is [...] implied in the concepts themselves'

(1) Der Satz der Identität meiner selbst bei allem Mannigfaltigem, dessen ich mir bewusst bin, ist ein eben so wohl in den Begriffen selbst liegender…

(2) '(It) is therefore an analytic proposition'

(2) […] mithin analytischer Satz

(3) 'But this identity of the subject, of which I can be conscious in all my representations does not concern any intuition of the subject, whereby it is given as object'

(3) Aber diese Identität des Subjekts, deren ich mir in allen seinen Vorstellungen bewusst werden kann, betrifft nicht die Anschauung desselben, dadurch es als Objekt gegeben ist,…

(4c) '[It] cannot therefore signify the identity of the person, if by that is understood the consciousness of the identity of own's own substance as thinking being in all change of its states' (CPR 369, B408)

(4c) […] kann also auch nicht die Identität der Person bedeuten, wodurch das Bewusstsein der Identität seiner eigenen Substanz, als denkendes Wesen, in allem Wechsel der Zustände verstanden wird […]

(1) entails that it is a conceptual truth that through any time that I am conscious of the manifold the subject of that consciousness is numerically identical with itself. It is clearly a modal fact about any subject of consciousness that it is what it is: This follows from giving 'x' the value 'subject of consciousness' in the formal truth $\Box (\forall x)(x = x)$. So Kant is right that 'I am identical with myself' is true, because it is necessarily true. We can allow that this is a conceptual truth also because 'I' and 'myself' have a single referent in 'I am identical with myself', and both 'I' and 'myself' are words which a person uses knowingly to refer only to himself.

But by saying this 'is […] implied in the concepts' Kant means we should draw no ontological conclusions about the persistence of the self over time from these grammatical and conceptual truths.

We can allow (2), the claim that (1) is analytic, because if a proposition is both a necessary truth and a conceptual truth then that is a sufficient condition of its being analytic. 'I am self-identical in my consciousness of the manifold' is a necessary truth and a conceptual truth, so it is analytic.

So if it is true that I am conscious of the manifold of intuition, say between t and t^1, then it follows that just one subject of experience is conscious of the manifold of intuition between t and t^1. (Here the privacy and inalienablity of the manifold is assumed).

(3) separates two doctrines: the transcendental doctrine of apperception, the theory that any of my experiences may in principle be prefixed by 'I think […]' as a condition of its being mine, and the doctrine that I am directly acquainted with my personal identity as an object of my own introspection. Kant accepts the first of these views, but rejects the second at least in so far as this means directly acquainted with myself as a changeless substance over time. In particular, Kant insists that introspective acquaintance does not follow from, and is no part of, the transcendental unity of apperception doctrine.

There is an ambiguity in the notion that I may be conscious of the identity of the subject in all my representations. This could mean that I may be acquainted with the truth conditions of the transcendental unity of apperception theory, in which case Kant is clearly right to insist that this knowledge is distinct from any self-acquaintance in introspection. Or, Kant might mean that I can be self-conscious: If I am self-conscious then what I am conscious of does not differ numerically between the different occasions on which I am self conscious. That is necessary if each occasion is to count as an episode of *self* consciousness. In this sense, I maintain, (3) is true. Kant presupposes the changelessness of the subject but fails to specify its ontology.

The apperception doctrine includes a dispositional account of self-consciousness: the possibility that I think 'I think' when I experience, if not the persistent actuality. If Kant does mean this, then he must allow that we are each acquainted with ourselves though our own experience of ourselves. But all he need concede here is his own doctrine of inner sense: the theory that a person may be directly acquainted with their own mental states through intuition of them. Kant thinks he is not thereby committed to the rational psychological doctrine that I am directly acquainted with myself as a mental substance which persists over time. Nevertheless, I claim that we are thus acquainted with ourselves, even if not under that description. The changelessness of the soul is a presupposition of the exercise of inner sense over time. It is what the identity, as opposed to the mere continuity, of the subject consists in. If I am a soul, and if I am aware of myself, a soul is what I am aware of irrespective of whether I know that fact.

(3) includes part of Hume's theory of the self: the thesis that, in self-consciousness, a person is never acquainted with a self as one item persisting amongst the set of that person's experiences. Kant is explicitly contrasting his theory of the transcendental unity of apperception, which is a formal thesis about how experience is possible, with the thesis that Hume denies: that we are each aware of ourselves as enduring mental substances. Clearly the proposition that we are each directly acquainted with ourselves as persisting mental substances is synthetic, because it is informative about the self. That we are persisting mental substances is necessary for our personal identity. Kant is wrong to think the thesis of the transcendental unity of apperception is only formal, because it entails synthetic propositions. Kant's apperception doctrine entails propositions about the ontology of the self.

[The claim that I intuit myself is an epistemological claim, and one that is verified or falsified by reference to experience. But the transcendental unity of apperception is formal, and not either epistemological or empirical. It is not a thesis about the ways in which I am acquainted with myself so the object of my own introspection cannot be inferred from the formal unity of apperception either.]

The conclusion, (4c), only follows from the premises because (4c) says that it does not follow from the formal unity of the subject that a person introspectively comes to know that formal unity doctrine is true. On Kant's account this does not carry any ontological implication. In particular it does not yield the conclusion the rational psychologist seeks: that each self is a substance in the sense of a purely mental individual which persists over time through changes in its properties. If Hume and Kant are wrong about introspection, the conclusion of this argument is false.

Kant's "Flux " Argument

At CPR 342-3, A354 -A364 Kant has an argument to demonstrate the consistency of his non-substantial theory of personal identity with 'the dictum of certain ancient schools that everything in the world is in flux '. (6) This raises the questions of

whether Kant may be construed as even endorsing a process philosophy of the self, and whether his argument is sound.

Kant's philosophical motivation in the section of the *Critique of Pure Reason* where the flux argument appears is to offer a solution to the problem of personal identity that does not entail Cartesian dualism. Kant thinks (rightly) that the flux theory is inconsistent with the existence of substance. (CPR 65-91, A19-48, B33-73) So if he can show his theory is consistent with the flux theory then he will have opened an ontological alternative to the postulation of a Cartesian substantial self. The argument is as follows:

(1) 'We are unable from our own consciousness to determine whether, as souls, we are permanent or not'

(1) Denn wir selbst können aus unserem Bewusstsein darüber nicht urteilen, ob wir als Seele beharrlich sind, oder nicht..

(2) 'We reckon on belonging to our identical self only that of which we are conscious'

(2) …weil wir zu unserem identischen Selbst nur dasjenige zählen, dessen wir uns bewusst sein[...]

(3) 'We must necessarily judge that we are one and the same throughout the whole time of which we are conscious'

(3) …, und so allerdings notwendig urteilen müssen: dass wir in der ganzen Zeit, deren wir uns bewusst sein, eben dieselbe sind.

(4) 'The only permanent appearance which we encounter in the soul is the representation 'I' that accompanies and connects them all'

(4) [In dem Standpunkte eines Fremden aber können wir dieses darum nocht nicht vor gültig erklären, weil] da wir an der Seele keine beharrliche Erscheunung antreffen, als nur die Vorstellung Ich, welche sie alle begleitet und verknüpft [...].

(5) 'We are unable to prove that this 'I', a mere thought', may not be in the same state of flux as the other thoughts'

(5) […so können wir niemals ausmachen, ob dieses Ich (ein bloßer Gedanke) nichteben sowohl fließe, als die übrige Gedanken, die dadurch an einander gekettet werden.]

(6c) 'Although the dictum of certain ancient schools, that everything in the world is in a flux and nothing is permanent and abiding, cannot be reconciled with the admission of substances, it is not refuted by the unity of self-consciousness'

(6c) [Wenn gleich der Satz einiger alter Schulen: dass alles fließend sei, nicht statt finden kann, sobald man Substanzen annimmt, so ist er doch nicht durch die Einheit des Selbstbewusstseins widerlegt.]

Premise (1) is the claim that introspectively based claims cannot decide the ontological issue between Kant and the 'rational psychologist'; anyone who thinks the existence of the soul *qua* substance can be proved *a priori*. 'As souls' here just means 'as selves' and not 'as Cartesian substances', so from no set of first person singular psychological claims not already mentioning the Cartesian soul can any sentence of the form 'I am a Cartesian soul' be logically derived. Here Kant rightly takes permanence to be a defining characteristic of 'soul' in the Cartesian sense (or necessary condition for being a substantial soul). He argues validly that if a soul is permanent (except for vulnerability to divine anihilation), and if I am not permanent, then clearly I am not a soul. Although this quasi-Humean tenet is widely endorsed, it is not obvious that it is true. Something has remained utterly changeless in one's own experience. It is not clear that, in introspection, one is acquainted with something that could disintegrate into parts. Consciousness, as opposed to the episodes in it, is presented as indivisibly one.

Premise (2) stipulates a content for the self. As expressed in (2), this content is too wide because I can be conscious of things other than myself. We therefore need to read 'conscious' as 'self-conscious' or 'conscious only through inner sense and empirical apperception'. Then (2) amounts to the claim that a person counts as himself only what falls within the scope of his capacities for self awareness. I think we have to insert 'could in principle' here so that we have: A person believes him or herself to be essentially made up of those aspects of him or herself with which they could in principle be aware. We need this because it is an empirical truth that persons believe themselves to be partly constituted by aspects of themselves of which they are not aware.

There is another way of taking (2) which does not require restricting 'conscious' to 'self-conscious'. Then, the content of my consciousness, broadly conceived as my 'intuitions and my (re)presentations', is what I identify myself with. This would yield a view of the self as apparently nothing over and above its experiences (if we allow the plausible assumption that one's experiences may be objects of one's consciousness).

(3) entails that it has to be the case that if a person is conscious between t^1 and t^2, then during that time: that person must judge that the person conscious at t^1 is identical with the person at t^2, persists uninterruptedly between t^1 and t^2, and is

(numerically identical with) that person. This requirement looks too strong. It is not a precondition of a person being conscious between t^1 and t^2 that that person be in the occurrent and conscious mental state of judging between t^1 and t^2 that he is self-identical between t^1 and t^2. On the other hand, there is some plausibility in suggesting that it is a condition of a person being conscious between t^1 and t^2 that that person does not disbelieve that he is identical with the person at t^2 and the person at t^1 and that they are identical with each other. If someone did disbelieve this then such a putative person would have to disown any memory links that seemed to obtain between him or herself at t^1 and him or herself at t^2.

Prima facie there might be a difference between what a person is and what a person thinks they are, and between who a person is and who a person thinks they are. If memory is partly constitutive of being a person, but a person systematically disavows memories of their past state, then one important conceptual component of 'person' is broken, at least on a Kantian analysis. I take it this is why Kant includes 'necessarily' in (3). My belief that I am self-identical over time is a necessary condition of my being an 'I' user at all, so if I am an 'I' user then I believe I am self-identical over time. Here 'believe' need not be given a strong reading. On Kant's theory if A judges *x* to be *F* then A believes *x* to be *F*. Believing and judging may be given dispositional rather than occurrent analyses, so we are not forced to the rather implausible view that persons persistently judge themselves to be identical over time.

I have taken 'time of which we are conscious' to mean 'time that we are conscious' but, if we read this literally, Kant is talking about consciousness of time. It is a thesis of the Transcendental Aesthetic chapter of the *Critique of Pure Reason* that all consciousness is temporal: that all outer and inner intuition is temporal. Time is transcendentally ideal but empirically real, so in just that sense it pertains to the sensibility of the subject. (2)

Bearing this thesis in mind we can reread (3) so that it says: When I am in any sense conscious of time I am necessarily conscious of my self-identity through that time. 'Conscious of the whole of time', could mean the awareness a person has of being located within the time series as a whole. Or, it could mean the sort of direct acquaintance with time, as a series of temporally ordered mental states, that Kant thinks each person possesses through inner sense. Or, finally, it could mean roughly 'lifetime', or the time-slice of events of which I am conscious so long as I persist. These interpretations are interestingly different, and nothing in the text establishes whether Kant has all of them in mind. But nothing about (3) precludes its applying to each of them.

So, then the claim is; If I am conscious of time then I judge myself to be self identical during that same time. We should draw a distinction here between two temporal series (which is consistent with the thesis of the Transcendental Aesthetic that there is only one time). There exists the temporal series that I am conscious of. But there exists also the temporal series of which my consciousness of time is a part. Now, for many purposes these could be treated as one and the same: For example, if I observe a changing physical object. But, if part of my consciousness of time is an

awareness of the whole of time then that awareness of the whole can only be *via* an awareness of a part. Or I would have to last the whole of time. Then it cannot be true that so long as I am aware of time I judge myself to be self-identical during the whole of the time I am aware of: I only judge myself to be self-identical during that time; however extensive the time of which I am aware. This interpretation of (3), which draws on theses about time, needs to be supplemented by the same remarks, above, about dispositional and occurrent belief and judgement.

(4) seems on the face of it to contain a contradiction, if we adopt a purely formal interpretation of the transcendental unity of apperception. This is because if the 'I' is merely formal it could not possibly be a 'permanent appearance which we encounter in the soul'. But the contradiction can be avoided if we do not take 'appearance' to mean 'appearance to inner sense'. Then Kant can avoid the suggestion that I can be introspectively aware of one item amongst others called 'I'. Kant strongly rejects this suggestion on Humean grounds. Similarly, 'permanent' must not be taken to suggest an occurrent mental state which endures throughout the whole series of episodes making up a person's life. We can give 'permanent' a more dispositional reading so that whenever I am self conscious I think (or may think) of my thoughts as my thoughts. If I could not do this we could not talk of self consciousness at all in this case. I do not always have to be thinking this thought: 'This is a thought of mine' but it must be a permanent fact about me that I am capable of thinking such thoughts if it is true of me that I am a self-conscious being. This capacity must last as long as I do.

Part of what (4) amounts to is that the 'I' is just one thought amongst others. More precisely, some of my thoughts may be couched in the first person singular. (5) makes this explicit in 'this 'I'' is 'a mere thought'. But Kant makes an important further claim by (5): the 'I' itself might change. This can be taken in at least three ways.

Firstly, the 'I' is just one thought amongst others and thoughts come and go. Some thought comes and goes in this sense if and only if numerically distinct tokens of some thought type occur separated by time intervals. If the dispositional analysis of 'I think' is right then a person does not always have to be thinking 'I think [[...]]' in order to lead a conscious life so 'I thoughts' may come and go. They might, as he puts it, be 'in a state of flux '.

Kant also leaves room for the idea of a changing content to 'I' thoughts. In order for them to remain ‘I’ thoughts they have to retain essentially the indexical function of the first person singular pronoun. Consistently with that, I can self consciously think different propositional contents; I think that *p*, I think that *q* etc. First person singular psychological ascriptions may change in semantic content while remaining essentially prefixed by 'I think […]'.

The other way of understanding how the I may be in flux is to accept that 'I' denotes something more than thought. For example, it denotes the person who I am, or, on rational psychology, it denotes an immaterial substance. Kant is saying that if the 'I' denotes the person who uses it then what it denotes can change. So, suppose 'I' is the word each person uses to refer only to himself, then quite consistently with that

on Kant's view, what 'I' thereby refers to may change its nature over time, and threby change what it (qualitatively) is over time. Just as on the semantic interpretation the indexical function of the first person singular pronoun had to be left intact to make sense of what 'I' prefaces changing so, on the denotative reading, some denotation of 'I' has to remain numerically identical over time in order for us to talk about an 'I' being in a state of flux. Kant can clearly allow this on his episodic theory of personal identity. What makes a person at an earlier time, t^1, identical with a person at a later time, t^2, is that they are each parts, time slices, of the same episodic person. This is equivalent to claiming that a person as a whole is a process.

There is another way of taking (3) that bears scrutiny. It involves emphasising that Kant thinks an 'I thought' 'might be in the same state of flux as the other thoughts. So, not only is the 'I think [...] ' one thought amongst others but all thoughts, including the 'I' thoughts, are in flux. On this reading Kantextends Hume's 'bundle of perceptions' theory of the self, so that the 'I' is just one more member of the bundle. The contents of the bundle change over time so sometimes there is an 'I' thought in it and sometimes there is not. The 'I' thought is as episodic and intermittent as any other thought. This is quite consistent with the 'permanent appearance' doctrine in (4) so long as we give that the dispositional and formal reading. Then it is a different version of his doctrine that the ontological non-continuity of any permanent substantival self is consistent with the actual logic of 'I'.

Granted (1) to (5), does the conclusion, (c), follow? Clearly it does, because we have to read 'unity of consciousness' as 'formal unity of consciousness'; that expressed by the analytic truth 'all my thoughts are mine'.

On the quasi-Heraclitean account, with which Kant thinks his views consistent, nothing exists that does not change and this is an ontological doctrine, not a sort of conceptual truth. But Kant's theory of the unity of consciousness as mentioned in (6c) is intended to avoid any strong metaphysical implications. Minimally, the 'I' is just one thought amongst others, even if 'I think [...]' being thinkable is a condition for some thoughts being mine.

So, if the view that everything is in flux is consistent with any number of conceptual or logical truths then it is consistent with the transcendental unity of apperception. As Kant wished to prove, if the transcendental unity of apperception and the episodic theory of personal identity are consistent with the Heraclitean flux doctrine, but the Heraclitean flux doctrine is inconsistent with the existence of substance, then the transcendental unity of apperception and the episodic theory of personal identity are inconsistent with the existence of mental substance.

Kant thinks the existence of the soul does not follow from a correct theory of the person and he thinks the view of the self as a Cartesian substance is logically dependent upon the dispositional account of the self. This is made clear when he says:

'Permanence however is in no way given prior to that numerical identity of our self which we infer from our identical apperception, but on the contrary is inferred from the numerical identity.' (CPR 363, A365)

Aber diese Beharrlichkeit ist uns vor der numerischen Identität unserer Selbst, die wir aus der identischen Apperzeption folgern, durch nichts gegeben, sondern wird daraus allererst gefolgert.

He thinks a transcendental illusion is produced by the use of 'I' by a person. The illusion is that 'I' denotes a persisting substantival self that the person essentially is. [Suppose *per impossibile* a rational proof of the existence of the soul were possible, then Kant thinks it would not follow that there would exist a permanent consciousness, a kind of uninterrupted continuing awareness, but the possibility of that would have been proved.]

This is not correct, even though Descartes, for example, is unsure whether the soul always thinks, even though he is sure that it exists. If a substantial soul exists then it follows on the rationalist account that it is capable of thought: It is what thinks. Kant's point is that 'the possibility of a continuing consciousness is already sufficient for personality'. This means that if *x* is to count as a person then it is not necessary that *x* perpetually be conscious, merely that *x* posses a disposition to consciousness. This clearly requires only that *x* sometimes be conscious, if a disposition need only be intermittently exercised or even merely that *x* be capable of consciousness, if a disposition need never be exercised. So, if there are times when I am not conscious, as there are when I am unconscious, I do not thereby cease to be a person. This is the force of 'personality does not itself cease because its activity is for a time interrupted'. Conceptual room is left for a person who is never conscious.

'Substance' is a category, and categories have only an empirical use, so 'substance' has no metaphysical use, and so cannot be coherently conjoined with the predicate 'mental' or 'spiritual'. But its use, like that of the other categories, presupposes the truth of the transcendental unity of apperception. In this sense, Kant thinks the rational psychologist's theory of the self may only be formulated on condition that Kant's theory of the self is true. In fact the reverse is the case. If we ask: Who or what posseses the disposition to self consciousness? the only plausible answer is: The soul. Kant thinks it is philosophically harmless to retain the concept 'person' so long as we restrict it to its empirical uses and do not try to use it to extend self-knowledge in a the direction of substance metaphysics. Indeed, he thinks 'person' indispensable for practical purposes. (2)

I conclude that it is legitimate to read Kant as a process philosopher, at least in his account of the self. An essential prerequisite for being a person, and so for being anyone, is the intermittent exercise of a disposition to self-consciousness. That intermittent exercise is a process. Self-consciousness partly consists in the fact that some thoughts may be couched in the grammatical first-person singular, partly in the fact that some thoughts are about others in an internalist and unmediated way. Kant

is at pains to repudiate the view that the self is a mental substance. On at least one interpretation Kant thinks a person as a whole is a process.

Kant's argument that his account of personal identity is consistent with the process ontology of certain 'ancient schools' is valid. It is less clear that it is sound because that would require showing that the premises are ultimately beyond dispute.

Kant's Attack on Mendelssohn

Kant's considerations of immortality and annihilation are concentrated in the section of the *Critique of Pure Reason* called 'Refutation of Mendelssohn's Proof of the Permanence of the Soul' and in the chapter on the postulates of pure practical reason in the *Critique of Practical Reason*. (1) I shall deal with the first critique then the second.

As reconstructed by Kant, Mendelssohn's position on death is that there is a distinction between two sorts of destruction. One, called 'dissolution', consists in something being fragmented into parts but each of these parts does not cease to exist. The second, called 'vanishing', consists in the whole of something entirely ceasing to exist, so that not even any part of it continues to exist. Mendelssohn has an argument by which he tries to demonstrate the impossibility of the soul being destroyed in the second of these two ways. Clearly, if the two sorts of destruction distinguished by Mendelssohn are jointly exhaustive, and if Mendelssohn's argument is sound then the soul is immortal and death does not consist in the destruction of the soul. This is Mendelssohn's argument as presented by Kant:

(1) 'The soul cannot be diminished, and so gradually lose something of its existence being by degrees changed into nothing'

[Die Seele] könne gar nicht aufhören zu sein, weil, da es gar nicht vermindert werden und also nach und nach etwas an seinem Dasein verlieren, und so allmählich in Nichts verwandelt werden könne

(2) 'Since it (the soul) has no parts it has no multiplicity in itself'

Indem es (die Seele) keine Teile, also auch keine Vielheit in sich habe.

(3) 'There would be no time between a moment in which it is and another in which it is not-which is impossible'

Zwischen einem Augenblicke, darin es ist, und dem andern, darin es nicht mehr ist, gar keine Zeit angetroffen werden würde, welches unmöglich ist.

(C) 'A simple being cannot cease to exist'

Ein einfaches Wesen könne gar nicht aufhören zu sein. CPR 372-3, B413-4)

The first premise is the assumption of the simplicity of the self. Kant does not criticise this proposition directly at CPR 372-3, presumably because he thinks it vulnerable to all the objections brought against the conclusion of the Second Paralogism. It is the thesis that the soul is not even in principle divisible because it has no parts. If we read 'since' in (1) as 'if [...] then' then (1) is necessarily true, because all it amounts to is that if the soul has no parts then it is not an aggregate or a collection of components. This is analytic, but does not prove that there exist any simple souls.

The second premise is that the soul, as a simple substance, cannot gradually cease to exist. The claim is that if something gradually ceases to exist then part of it must continue to exist while some other part has ceased to exist. But the soul has no parts, so it is logically impossible that part of it should cease to exist while part of it continues to exist. For this premise to have any force against the possibility of the annihilation of the soul Mendelssohn needs to prove independently that if something is annihilated then it is necessarily annihilated part by part and not all at once. Otherwise the clear possibility remains open that the soul be wholly and instantaneously annihilated, even if it could not have been annihilated part by part in a chronological sequence.

Premise (3) should be read as providing further grounds for accepting (2). Mendelssohn thinks that if something is annihilated then it must be annihilated over time so there exists during that time some part which has not be annihilated, but some part has been annihilated. This gradual (or temporally extended) destruction must obtain if there is any annihilation, on his view, otherwise there could not exist any time between the time when the soul exists and the time when it does not exist. Without a time chronologically located between those times there could not be a time at which the soul is destroyed, and if there is no time at which the soul is destroyed then the soul cannot be destroyed.

I think this premise is false, because it relies on a questionable punctual notion of time, one on which time is objectively divided into discrete 'moments' or temporal durations. Mendlesohn has not shown that the destruction of the soul cannot be instantaneous. It is not incoherent to maintain that the soul is destroyed but does not take any time to be destroyed.

If there is a problem it arises for the destruction of the putative parts of the soul also, if (on Mendelssohn's theory *per impossibile*) the soul were composite, and it were destroyed piecemeal, some account of how this is possible would have to be given. Must the parts have parts so that they might be destroyed piecemeal? and so on *ad infinitum* ?

Secondly, arguably no change would be possible if, as Mendelssohn implies, there has to exist some time between some state S and some state S' for S to cease to exist and for S' to obtain instead. Suppose S is 'the fly is on the table' and 'S is 'the fly is off the table' then it would be contradictory to postulate some intermediate time

when the fly is both on and off the table, because, at any time, the fly is either on or off the table. Similarly, in the case of the soul, the soul at any given time either exists or does not exist and there is no need to postulate, with dubious coherence, some intervening time which is the time when it ceases to exist. So far from (3) being true, it is arguably impossible that there should be a 'time between a moment in which it is and another in which it is not'.

The conclusion that a simple being cannot cease to exist, does validly follow from the premisees, but because at least (3) is false, and (1) and (2) may be false, the argument cannot be sound.

Nor are there independent grounds for believing that '*x* cannot be annihilated' follows from ' *x* is simple'. Certainly, if ' *x* is simple' is true then it follows that *x* is indivisible, and if *x* is indivisible then it follows that *x* may not be destroyed in the first of the two senses distinguished by Mendelssohn. *x* cannot be subject to dissolution because *x* 's putative dissolution would necessarily consist in the distribution of *x* 's parts, but, obviously. if *x* has no parts, *x* cannot be subject to dissolution.

But it does not follow from any of this that *x* may not be subject to destruction in the other of Mendelssohn's two senses. *x* may vanish, or as I should say, *x* could be annihilated or utterly cease to exist whether or not *x* consists of component parts. This is because it is not true that the only way in which something may be annihilated is by the gradual destruction of its parts; there is no *a priori* objection to anything's entire and instantaneous annihilation.

Kant's criticism of Mendelssohn's argument is that even if it is true that the soul is simple it still does not follow that it cannot cease to exist gradually. He makes the supposition (for the sake of argument) that the soul is simple in the sense of containing 'no manifold of constituents external to one another' (CPR 373, B414). This means 'not composed of numerically distinct (and so in principle separable) parts'.

But even if the soul is non-composite in that sense, it might nevertheless be the case that it is admits of 'a degree of reality in respect of its faculties' “einen Grad der Realität in Ansehung aller ihrer Vermögen” (CPR 373, B414). This is ambiguous. It could mean that the soul, although simple in the sense supposed above, is composite in a different sense *viz* by comprising certain mental faculties; perception, imagination, memory etc. But on a second interpretation, which is born out by Kant's 'Consciousness itself has always a degree which always allows of diminution', the soul might, to use a metaphor, fade out of existence, rather as the light of a light bulb might diminish in intensity. The two interpretations are mutually consistent, and there is nothing to preclude our ascribing both to Kant's text.

If we consider the 'mental faculty' view, then clearly the soul may cease to exist gradually through the successive annihilation of its faculties, although any such account would have to respect the mutual dependence of mental faculties. The 'diminution of intensity' interpretation is perhaps more effective against Mendelssohn than the 'mental powers account'. Kant concludes:

'In the manner of the supposed substance - the thing, the permanence of which has not yet been proved - may be changed into nothing, not indeed by dissolution but by gradual loss (remission) of its powers.' (CPR 373, B414).

Und so die vorgebliche Substanz obgleich nicht durch Zerteilung, doch durch allmähliche Nachlassung (remissio) ihrer Kräfte in Nichts verwandelt werden könne.

The difficulty for deriving this conclusion on the 'mental powers' interpretation is that if we make the supposition that the soul is a substance (which we must for sake of the argument) then it does not follow from the fact that certain properties of a substance cease to exist (in this case the powers of the soul) that the substance itself should cease to exist. In rationalist psychology, a substance is that which bears properties and is not itself a property so it would be mistake to think that a substance could be annihilated by the destruction of its properties. It might not still warrant the name 'substance' but that verbal point does little damage to the rational psychologist's ontology of the self.

So, from the fact that a soul's memory, imagination, etc are successively destroyed it cannot be concluded that the soul itself is destroyed, so Kant is wrong to claim on the basis of the argument so far that the soul 'may be changed into nothing'. We would have to extend the 'diminution of intensity' interpretation so that it applies to the substance as well as its properties in orer for Kant's argument goes through. If we think of the soul, to pursue the metaphor, as the light itself (rather than, say, the light bulb), then even the diminution of the light to what seems like total darkness presupposes a kind of inner light: the dark night of the soul described by St. John of the Cross. There is no imagining of the gradual non-existence of the pure consciousness that the soul consists in because it is the acknowledged or unacknowledged ground of the imagining itself. This helps us to understand the falsity of Kant's claim that a soul may cease to exist gradually even if it is not composite in the sense that, say, a physical object is composite. (2)

In a sense the rational psychologist seeks to prove too much in trying to establish the *logical* impossibility of the destruction of the soul, for it is a traditional tenet of religious metaphysics that God may annihilate the soul.

The Moral Argument

Although Kant thinks there exists no valid argument for the immortality of the soul he thinks we have to postulate the immortality of the soul in order to make sense of our moral obligations. That there does exist an immortal soul is a postulate of pure practical reason. He defines 'postulate of pure practical reason' at CPR 127:

'By a postulate of pure practical reason, I understand a theoretical proposition which is not as such demonstrable, but which is an inseparable corollary of an a priori unconditionally valid practical law.'

By 'proposition' I shall understand here, straightforwardly, what is expressed by a sentence capable of truth or falsity. By 'theoretical' I understand 'non-empirical', by 'not as such demonstrable' I understand that any such theoretical proposition could not even in principle feature in the conclusion place of any deductive or inductive argument from completely factual premisees. So postulates are unprovable non-empirical propositions. Yet that they be true, or at least that we believe them to be true, is a necessary condition for the existence of moral obligations.

The *a priori* unconditionally valid practical law mentioned in the definition is the categorical imperative: the obligation to act only in such a way that one's behaviour could be universalised so as to conform to a general rule; a rule that all similarly placed persons should also act in that way. Kant thinks we are each under a pure moral obligation to act from duty, and the fact of this moral obligation cannot itself be derived from any empirical premisees. So, the postulation of the immortality of the soul makes intelligible the real nature of our moral obligations, and the existence of the immortal soul makes those very obligations possible. Kant has an argument for this to which I now turn:

(1) 'The achievement of the highest good in the world is the necessary object of a will determinable by the moral law'

(2) 'In such a will, however, the complete fitness of intentions to the moral law is the supreme condition of the highest good'

(3) 'This fitness therefore must be just as possible as its object, because it is contained in the command that requires us to promote the latter'

(4) 'But the complete fitness of the will to the moral law is holiness, which is a perfection of which no rational being in the world of sense is at any time capable'

(5) 'But since it is required as practically necessary, it can be found only in and endless progress to that complete fitness'

(6C) 'On principles of pure practical reasons, it is necessary to assume such a practical progress as the real object of our will'
(CPR 127)

The first premise is the claim that if a person's will (ie. intention) is determined by the categorical imperative then it necessarily tends towards the highest good possible.

(1) does not contain any argument for the conclusion that we rational self-conscious beings are under an obligation to act in accordance with the categorical imperative. That conclusion is derived from moral premises which we need not consider here. (1) makes explicit one consequence of what Kant assumes to be an ethical fact about us: We are under a moral obligation to do our duty.

(1) seems to contain a fallacy. From the fact, if it is a fact, that persons are under a moral obligation to maximise good and do tend towards virtue by performing the largest number of actions possible from pure duty, it seems not to follow that there exists a 'highest' good which in some sense persons thereby tend towards. The maximisation of some tendency does not logically presuppose some maximal limit to that tendency. But unless Kant assumes this then 'highest good' in (1) makes little sense. It could mean only, perhaps, ‘highest good which striving beings are capable of achieving’.

(2) at least contains the claim that we could not be under any moral obligation unless the actions we are obliged to perform could in principle be carried out. That is what he means by 'the fitness (of the will) must be possible': It must be within our capabilities to act morally if it is true of us that we are obliged to act morally. So, (2) is a version of the doctrine that 'ought' implies 'can', or if it is true that A ought to do *x* then it is thereby true that A can, at least in principle, do *x*. It also follows that if we are under an obligation to achieve the 'highest good' then we could in principle achieve the 'highest good'. However, from the fact that we are under an obligation to maximise good, it does not follow that we are capable of performing the highest good that exists, or that there exists some good higher than that of which rational self-conscious (empirical) beings are maximally capable.

Premise (3) expresses the proposition that we must be capable, in principle, of achieving the highest good if it is true that we are obliged to achieve the highest good. This premise therefore follows from (2), and it is rational to accept it so long as we accept the existence of the highest good, the fact that we are under a moral obligation to achieve it, and the 'ought implies can' doctrine. When Kant says 'contained in the command' in (3) he is insisting on the 'ought implies can' doctrine. He means 'conceptually contained’ so that the command (or imperative) 'act in accordance with the moral law' contains or implies our 'fitness' or capacity to act in accordance with the moral law.

(4) expresses two related theses. First, although we have an obligation to achieve the highest good this cannot be done 'in the world of sense'. This means we cannot achieve what is infinitely good in the finite spatio-temporal world of physical objects. Secondly, and following from this, if we are capable of achieving the highest good, but if this is not in virtue of our empirical properties, our being physical and finite, then it must be because we are also holy, or at least in principle capable of holiness. When he says the complete fitness of our will to achieve the highest good is holiness he means infinite perfection of a non-physical sort. The fact that no rational being in the world of sense is capable of holiness carries the suggestion, or at least

opens the possibility, that some rational being not occupying the world of sense might be capable of holiness.

(5) says that the truth of (4) is required as 'practically necessary'. I take it this means that that (4) is true is a condition our being under an obligation to conform to the categorical imperative. The second part of (5) is the claim that we must be engaged in an infinite or endless moral progress, not a finite one. This does not follow from the first sentence of (5), nor from (4), despite Kant's use of 'since'. This is because there seems no *a priori* objection to a rational being achieving the highest good possible, or at least Kant would have to demonstrate some such objection to show that our moral progress is infinite, does not come to an end.

(6C) does validly follow from the premisees, even if some of those premisees are false or at least not known to be true, because if we are under an obligation to achieve the highest good, and if this is only possible by infinite progress, and if we are capable of this infinite progress then we must ourselves be infinite. Any being capable of infinite progress is itself infinite.

Although the argument is valid, it requires further premises to show that it is sound. If our progress to the highest good is infinite then it might be objected that it is in principle impossible for us to achieve the highest good - we can only approach it but never reach it. But if that is true then we cannot be under any obligation to achieve the highest good either because, on Kant's own terms, I am obliged to do *x* only if I am capable of doing *x*. If it is certain that there will never be a time that I do *x* then I cannot do *x* and so am not obliged to do *x*. If we are of such a nature to achieve the highest good we must be timeless. If the highest good is an infinitely receding future goal it would take infinite time to reach. If the highest good is enternal in the sense of the eternal present then it may be achieved or disclosed if we ourselves partake of that eternal presence.

It follows that if we are capable of the 'highest good', we are obliged to do the highest good, and so Kant has adduced a persuasive reason to suppose we are engaged in an infinite progress to the highest good. It follows that Kant has provided ethical grounds for supposing we have immortal souls. If we are souls, this is because we are in a sense eternally present. (3)(4)

IV

Kant’s Refutation of Idealism

There is a variety of theses in metaphysics and the philosophy of mind called 'idealism' and it is doubtful whether all the doctrines that go by that name are mutually consistent. Here I examine some idealisms which Kant distinguishes in the

Critique of Pure Reason, then appraise for soundness two arguments he deploys against them.

The Varieties of Idealism

Kant defines 'material idealism' in this way:

'Idealism (I understand material [idealism]) is the theory which declares the existence of objects in space outside us either to be merely doubtful and indemonstrable or to be false and impossible' (B278)

'Der Idealism (ich verstehe den materialen) ist die Theorie, welche das Dasein der Gegenstände im Raum außer uns entweder bloß fur zweifelhaft und unerweislich, oder für falsch und unmöglich erklärt'

'Material Idealism' is the name for any or all of these doctrines:

1. It is doubtful that physical objects exist.

2. It is not provable that physical objects exist.

3. It is logically possible that physical objects do not exist.

4. It is false that physical objects exist.

5. It is provable that physical objects do not exist.

6. It is not possible that physical objects exist.

Kant calls any or all of 1,2 and 3 'problematic idealism', and says that Descartes is an exponent of it, and he calls any or all of 4,5 and 6 'dogmatic' idealism and says Berkeley holds this view.

It is worth examining the logical relations between problematic and dogmatic idealism because if any conjunction of any of 1-3 with any of 4-6 is inconsistent then that is a sufficient condition for some of 1-6 being false. Also if any of 1-3 entails any of 3-6, but some of 3-6 turn out to be false then then the entailing member of 1-3 is false. Conversely, if any of 3-6 entails any of 1-3 but some member of 4-6 turns out to be false then the entailing member of 1-3 is false. 1 and 4 are logically independent. 2. is ambiguous between a realist and an anti-realist interpretation because of an equivocation on 'provable'. Kant thinks that both problematic and dogmatic idealism are false so he is committed to the falsity of all of 1-6.

A Priori Subjects: Kant and the Existence of the Soul

Before we decide whether he is right about that we should decide what each of 'problematic idealism' and 'dogmatic idealism' denotes. That is best done by appraising the claim that Descartes and Berkeley each subscribe to one of them.

Kant on Descartes and Berkeley

What each sort of idealist rejects or doubts is a particular account of what a physical object is. In particular they either doubt, or deny, that physical objects are material particulars existing independently of sense experiences. This means that physical objects are not or might not be composed of matter, and they are or might be exhausted by our experience of them. Something is exhausted by our experience of it if and only if it neither predates nor postdates our experience of it and would not exist during our experience of it if that experience did not exist.

Now, it seems clear that Descartes in the Meditations does advocate a view of the 'problematic' type. He asserts there that:

'These general things viz eyes, head, hands and the like may be imaginary [...] Corporal nature in general, and its extension, are of this class of things; together with the figure of extended things, their quantity or size, and their number, as also the place where they are, the time during which they exist, and such like' (Descartes, 1974, 98)

If physical objects 'may be imaginary' they are doubtful, their existence cannot be proved, and it is logically possible that they do not exist. So, at least in the sceptical 'First Meditation' Descartes' claims fall under the description Kant has given of problematic idealism and entail (1-3).

Kant also says problematic idealism includes the claim that there is just one certain proposition or, at least, if this proposition were not certain no other proposition would be; the first person singular present tense assertion of existence:

'The problematic idealism of Descartes [...] holds that there is only one empirical assertion that is indubitably certain, that 'I am'.' (B274)

Der erstere ist der problematische des Cartesius, der nur Eine empirische Behauptung, nämlich: Ich bin, für ungezweifelt erklärt.

Kant's use of 'empirical' is puzzling here. Although 'I exist' could be given an empirical interpretation it is essential to read it non-empirically if it is to be the incorrigible foundation sentence for Descartes' epistemology because it is plausible that all empirical propositions are dubitable, as Descartes himself thinks. But clearly Descartes does argue that 'I exist' is the only ultimately incorrigible proposition. Famously:

'I am, I exist, is necessarily true, every time I express it or conceive of it in my mind.' (Descartes 1974, 103)

This is an incorrigible claim on this criterion: *p* is incorrigible just in case if *p* is believed then *p* is true. I take it this is the force of 'necessarily' here: If *p* is believed then *p* is not only true, but could not be false.

Although Kant is right to say Descartes thought 'I exist' the only ultimately incorrigible proposition it is rather misleading to make this part of what problematic idealism consists in. Rather it is a premise Descartes' version of problematic idealism rests on. If problematic idealism is the view that the existence of physical objects is doubtful then this is a logical consequence of the fact that only one proposition is certain, *viz* 'I exist'.

If Kant reports Descartes accurately, he is rather careless about Berkeley. For example, Kant thinks

'He (Berkeley) maintains that space, with all the things of which it is the inseparable condition is something which is in itself impossible.' (CPR 244, B277)

[...] der zweite ist der dogmatische des Berkeley, der den Raum, mit allen den Dingen, welchen er als unabtrennliche Bedingung anhängt, für etwas was an sich selbst unmöglich sei 8…] erklärt.

This is right if it means that Berkeley thinks absolute Newtonian space is impossible but wrong if it implies that Berkeley denies that 'space' has any referent. Berkeley has a quasi-Leibnizian, or relational, view of space as nothing over and above the extensions of and relations between physical objects (for which, of course, he provides an idealist analysis). Berkeley feels able to coherently imagine Leibnizian space but not Newtonian space:

'If we inquire narrowly we shall find we cannot even frame an idea of pure space exclusive of all body. This I confess seems impossible, as being as most abstract idea.' (Berkeley, 1977, 123).

Berkeley uses 'space' in a premise for an argument for his thesis that physical objects do not exist independently of minds: If space does not exist independently of physical objects, and if physical objects do not exist independently of minds, then space does not exist independently of minds. 'Space' has a referent but not a mind-independent referent.

Kant speaks of space as the 'inseparable condition' of the things in it, as part of Berkeley's view. This is right but Kant misses that Berkeley also holds the converse: That there exist physical objects (in his idealist sense) is a necessary and sufficient condition for there existing space.

Kant says that Berkeley

'therefore regards the things in space as merely imaginary entities'

und darum auch die Dinge im Raum für bloße Einbildungen halt. (CPR 244, B277).

but Berkeley's argument is not: Space does not exist, therefore putatively spatial objects are really mental (imaginary). Rather, Berkeley is concerned to argue the other way around: There are no mind-independent spatial objects so there is no mind-independent (Newtonian) space either. Nor is it right to say that for Berkeley physical objects are merely imaginary. Berkeley draws a clear distinction between sense-experience and imagination. For example, imagination is a power of 'variously compounding and dividing' (Berkeley 1977, 49) and it is an important part of his idealism that in sense experience we do not have this control over content.

Kant says dogmatic idealism is unavoidable if space is regarded as a property of things-in-themselves. (CPR 234, B277) This is quite wrong if it implies that the theory that space exists independently of our experience of it entails Berkeley's thesis. In fact quite the reverse is the case. If *per impossibile* (on Kant's view) space were a property of things as they are in themselves then Berkeley's idealist theory of space would be false. Kant then is hardly entitled to this conclusion about Berkeley's idealism:

'The ground on which this idealism rests has already been undermined by us in the Transcendental Aesthetic' (B277).

Der Grund zu diesem Idealism aber ist von uns in der transz. Ästhetik gehoben.

Despite his faulty exegesis of Berkeley, Kant is right to call Berkeley's idealism 'dogmatic idealism' because it does entail that physical objects do not exist.

Kant's Refutation

Kant then naturally says the refutation of idealism (of either sort) must take the form of a proof that we experience outer things, and that we do not merely imagine them. It is as well Kant does not in fact pursue this strategy because it might simply mark a sense-experience/imagination distinction which Descartes and Berkeley might be inclined to accept and which in any case would be consistent with both problematic and dogmatic idealism. Kant italizises 'experience' at (B278) but really he should italizise 'outer' because it is that physical objects exist independently of our experience of them that he needs to prove. It is this that he in fact tries to do.

The conclusion which Kant seeks to prove is:

'Even our inner experience, is possible only on the assumption of outer experience.' (B278)

Dass selbst unsere inneren, dem Cartesius unbezweifelte Erfahrung nur unter Voraussetzung äußerer Erfahrung möglich sei.

By 'outer experience' Kant means experience of mind-independent physical objects. If this sentence could be validly inferred from true premises then both problematic and dogmatic idealism would be refuted. Problematic and dogmatic idealism could be formulated just on condition that they are false. Kant's strategy is to show that the existence of physical objects may be logically derived from sentences about one's own mental states on extra premises. These premises are assumptions that he thinks the idealist will accept because they are tacit entailments of idealism.

He might also establish a weaker dependency of the inner on the outer. Three candidates are *prima facie* plausible: psychological, causal, and conceptual. On a psychological interpretation, we could not find our inner experience intelligible unless we had or had had outer experiences. This would be psychologically impossible for us. On a causal interpretation, we could not have inner experience unless we had or had had outer experience because our inner experiences are caused by our outer experiences. (For example, an empiricist might think the content of inner experience is caused by perceptions of physical objects.) On the semantic or conceptual interpretation it does not make sense to talk of inner experience unless it makes sense to speak of outer experience. There is no internal world unless there is an external world.

We may now examine Kant's two arguments against problematic and dogmatic idealism with these possibilities in mind, as well as the strong or logical one that it would be self-contradictory to affirm inner experience but deny outer experience.

Argument Against (Problematic and Dogmatic) Idealism (1):

(1) 'I am conscious of my own existence as determined in time.'

Ich bin mir meines Daseins als in der Zeit bestimmt bewusst.

(2) 'All determination of time presupposes something permanent in perception.'

Alle Zeitbestimmung setzt etwas Berharrliches in der Wahrnehmung voraus.

(3) 'This permanent cannot however be something in me, since it is only through this permanent that my existence in time can itself be determined'.

Dieses Beharrliche aber kann nicht eine Anschauung in mir sein. Denn alle Bestimmungsgründe meines Daseins, die in mir angetroffen werden können, sind

Vorstellungen, und bedürfen, als solche, selbst ein von ihnen unterschiedenes Beharrlich, worauf in Beziehung der Wechesel derselben, mithin mein Dasein in der Zeit, darin sie wechseln, bestimmt werden könne.

(4) 'Thus perception of this permanent is possible only through a thing outside me, and not through the mere representation of a thing outside me.'

Also ist die Wahrnehmung dieses Berharrlichen nur durch ein Ding außer mir und nicht durch die bloße Vorstellung eines Dinges außer mir möglich.

(5) 'Consequently the determination of my existence in time is possible only through the existence of actual things which I perceive outside me.' (B275-6)

Folglich ist die Bestimmung meines Daseins in der Zeit nur durch die Existenz wirklicher Dinge, die ich außer mir wahrnehme, möglich.

The commentators have struggled heroically to make sense of this argument.

The First Premise

Commentators failure to understand the refutation of idealism is caused by a misunderstanding of a word in the first premise. The German word that Kemp-Smith translates as 'determined' in the first premise is 'bestimmt'. This can mean 'determined' but a more perspicious translation here is 'located'. The word he translates in the second premise as 'determination of time' is 'Zeitbestimmung' which means 'location in time'. Kemp-Smith's renderings leave obscure the crucial point that Kant is discussing the concept one has of oneself as located in time. In what follows I translate 'bestimmt' and similar locutions by 'located' and similar locutions.
(1) The first premise therefore becomes:

'I am conscious of my own existence as located in time'

The following sentences are entailed by the first premise:

(i) I exist
(ii) I am conscious
(iii) I am conscious of my existence
(iv) I am located in time
(v) I am conscious of my existence as located in time.

I offer a provisional definition and a formal or semi-formal rendition of each:

(1) 'I exist'.

It is not clear what I have claimed about myself when I have claimed that I exist, but for present purposes we may assume 'I exist' is true if and only if there exists something with which I am identical.

(ii) 'I am conscious'
Nor is it clear what I have said about myself if I have said I am conscious. For present purposes I shall say I am conscious if and only if I have knowledge:

(iii) 'I am conscious of my existence'
I shall read 'I am conscious of my existence' to mean 'I know that I exist':

(iv), 'I am located in time'
(there are times before, simultaneous with, and later than x)

The first premise is, then:
I am conscious that I exist and conscious that (there are times before, simultaneous with, and later than x)))

The first premise is ambiguous between entailing ‘I am in time’ and not entailing that. On the first construal, I am conscious of myself as located in time and I am located in time. On the second construal I am conscious of myself as located in time but it does not follow that I am not in time. Kant needs the first construal for his refutation of idealism to be valid. I suggest it is not obviously true that I am located in time, even though it is obviously true that I have a concept of myself as located in time. For example, if the rational psychologist is right that I am an immaterial soul, and no immaterial souls are located in time, then I am not located in time. Of course, one good explanation of my concept of myself is that I am located in time. However, ‘in’ might not be right. It could be that all reality is present reality and the changing content of the presenet creates the illusion of being ‘between’ the past and the future. This is a third-person way of thinking imported into one’s own self conception. Once this conditioning is peeled away, there is only the timeless present.

The Second Premise

'All location in time presupposes something permanent in perception'

The German word Kemp-Smith translates as 'permanent' here is 'beharrliche'. 'beharrlich' can mean 'permanent' but here is best translated as 'constant', otherwise Kant just begs the question against the idealists.

I shall read premise (2) to mean that any knowledge of one’s location in time, requires the experience of something constant. Kant says someone's consciousness that they are located in time presupposes something constant in their perception. He does not spell out the nature of this presupposition, but it is at least *prima facie*

plausible that if someone can conceptually, or perceptually, discriminate himself from what he is not, then this is because there exists something with which he may perceptually or conceptually contrast himself: something that he is not.

When Kant says there must be something constant in perception this is ambiguous. It could mean there must be something that is perceived to be constant just so long as some perception lasts. Or, it could mean there must exist objects which are constant in the sense of existing whether perceived or not. Kant has not so far established the second of these and the first, although convincingly argued for, is too weak to do much damage to idealism. If we take the contrasive point that for *x* to be discriminable from not-*x* in thought or perception, then 'not-x ' must denote, then it is not clear that what 'not *x*' must denote must exist constantly in the sense of 'whether it is perceived or not'. There seems nothing to compel us to assert the existence of not-*x* independently of our perception of both *x* and not-*x*, or our thought of both *x* and not-*x*. Indeed, the existence of not-*x* unperceived by us (like the existence of not *x* as unthought by us) is arguably wholly irrelevant to the conditions for our making a conceptual or perceptual discrimination between *x* and not-*x*.

Kant does not draw a distinction between two senses in which it might be true that something exists independently of perception. On the first, something exists before the perception, during it, and after it, and would have still existed uninterruptedly even if it had not been perceived. On the second, something only exists at the same time as the perception of it, yet its existence does not depend upon its being perceived. Kant only needs the weaker of these to refute idealism. That the 'constant' in perception would still have existed even if not perceived is inconsistent with idealism so, if proved, refutes idealism. He does not need the stronger claim on which the permanent exists before it is perceived and continues to exist after it has been perceived, even though that is clearly sufficient for the refutation of idealism. Besides, we do not wish to make it a logical condition of the objectivity of physical objects that they exist for some stretch of time before and after they are perceived (even though that is no doubt true of most physical objects). What is essential to the anti-idealist concept of a physical object is that any such object's existence must be independent of any perception of it. Whether it is perceived must not decide the issue of whether it exists.

The Third Premise

Whatever the status of the constant element in perception, it has to be true that it is not 'in me'. If idealism is to be refuted, it must not pertain only to the psychology of the subject. Kant could have deployed a straightforward Humean argument to support this thesis: If there is nothing 'constant' within a person's own psychology with which they may be directly acquainted it follows that a person is never acquainted with himself *qua* psychological subject in introspective consciousness. It would then follow from the fact that a person's experience of themselves and of other

things is exhaustive of their experience, that if something constant is present in perception then that constant is external to the psychology of the subject.

But Kant does not pursue this line of argument. Instead he relies on a certain principle: If *y* makes possible the concept of *x* as located in time then it is not true that *y* is *y* or is any property of *x*. (Kant needs to exclude relational properties but I waive that objection.) Kant needs an argument for the conclusion that there is some condition for the existence of the concept of *x* over time other than *x*. My concept of myself as located in time is a specific case of the thesis that all location in time requires the existence of something other than that which is located in time. Nevertheless, that the argument may be couched in first person singular form, and that the first person singular case provides no exception to the thesis that if *y* is a condition for *x*'s location in time then -(*x* =*y*), is a necessary condition for the refutation of idealism on Kantian lines.

The Fourth Premise

We know that my location in time is consistent with the existence of physical objects as mind-independent particulars, but does it logically entail it?

A reason is implicit in (3) why it is impossible that any putative candidate for something constant in perception should in fact turn out to be one of my mental states. If we accept the thesis that if *y* determines *x* in time then *x* is not *y*, then if something determines myself in time then that cannot be myself. It follows that the constant in perception cannot be a representation because (*ex hypothesi*) mental states are in me. Now, if what is in me and what is not in me are mutually exclusive and jointly exhaustive, then if the constant is not in me it must belong to that portion of reality that is not me, so it exists independently of my mental states. This line of argument is sufficient to establish the objectivity of the permanent in perception, but perhaps not that it is a *thing*, if 'thing' here means 'physical object'.

If we accept Kant's reasoning so far, the logical possibility seems open that the constant in perception, which does the work of generating the concept of oneself as located in time, might in fact not be a physical object, still less a set of physical objects. Anything will do, just so long as these two conditions are met: I am not it, and, it changes no more than I do (plus, of course, it must be a possible object of my perception). No doubt the paradigm of a relatively permanent, objectively persisting, particular is a physical object but other possibilities seem open: a continuous smell, sound, or tactile phenomenon perhaps. Although this possibility is not closed by Kant's argument, certainly not by his bare use of 'thing', the argument is still strong enough to work against both the problematic and the dogmatic idealist. This is because they must at least deny or doubt the objective permanence of physical objects, ie deny or doubt that they persist in a mind-independent way. But it is just that characteristic of physical objects which Kant establishes by the argument.

The point at issue between Kant and the idealists is not the other putatively essential properties of physical objects (the primary qualities: shape, size, solidity,

etc) but whether an exhaustive analysis of 'physical object' may be provided by some description of the psychology of the subject, or whether that is impossible because they are enduring mind-independent particulars. That is really the only point of disagreement between Kant and the idealists, so it does not much matter that Kant does not close the argument by proving that the permanent element in perception is identical with the physical objects the idealists are skeptical of. This would be required for (5c) to formally go through, but enough is proved to refute problematic and dogmatic idealism, without (5C). This is because idealism entails:

(L) There are no (or might not be any) mind-independent particulars which are possible objects of perception

but Kant's refutation establishes:

(K) There exists at least one mind-independent particular which is a possible object of perception.

If we could supply just this:

(PO) Any mind-independent particular which is a possible object of perception is a physical object.

Then we would have

(I') There are no (or might not be any) physical objects.

but

(K') There exists at least one physical object.

But if Kant's refutation is sound, then at least (K) and possibly (K') is true. But (K) and (K') are inconsistent with (I) and (I') so (I) is false and possibly (L') are false.

Quasi-Solipsism

I distinguished above a sense in which it matters to Kant's case that he makes use of 'me' and a sense in which it does not. Now we need to examine more closely the significance of this first person singular formulation of the refutation of idealism.

Although there obtains a distinction between idealism and solipsism, there remains something solipsistic about even those idealisms which are not solipsisms. By ‘idealism’ I mean the doctrine that only minds exist. By ‘solipsism’ I mean the doctrine that only my mind exists. So solipsism is a restrictive version of idealism. All idealisms are quasi-solipsistic because their skeptical conclusions about the

objectivity of physical objects rest on premisees which are first person present tense psychological ascriptions. Indeed, many idealist arguments rely on denying the validity of inferences from such first person present tense psychological ascriptions to third person present tense physical ascriptions. Kant's purpose in couching the refutation in first person singular terms is to undermine this skepticism at its strongest point. This is the best way of interpreting Kant's avowed aim of showing that inner experience is only possible on condition there is outer experience. (See, for example, (B275) and (B276)).

The Second Refutation of Idealism (A 2)

Kant tries to make explicit the dependence of inner sense on the existence of outer sense, and the objective reality of its objects, by an argument at (B276):

(1) 'Consciousness of my existence in time in necessarily bound up with consciousness of the (condition of the) possibility of this time determination'

Nun ist das Bewusstsein in der Zeit mit dem Bewusstsein der Möglichkeit dieser Zeitbestimmung notwendig verbunden.

(2) 'It is therefore necessarily bound up with the existence of things outside me, as the condition of the time determination'

Also ist es auch mit der Existenz der Dinge außer mir, als Bedingung der zeitbestimmung, notwendig verbunden.

(3) 'In other words, the consciousness of my existence is at the same time an immediate consciousness of the existence of other things outside me' (CPR 245, B276).

Das Bewußtsein meines eigenen Daseins ist zugleich ein unmittelbares Bewußtsein des Daseins anderer Dinge außer mir.

The First Premise (Ref2)

In the first premise, 'bound up with' is an unfortunately vague spatial metaphor. Various possibilities are: if *x* is 'bound up with' *y*, then *y* makes *x* possible, if *x* exists then *y* exists, *y* makes *x* thinkable, or some conjunction of these. But we should read (1) as another conclusion of the first refutation of idealism. Then perhaps, minimally, (1) asserts that my consciousness of my own existence over time presupposes (in some sense) my consciousness of what makes possible my location in time. In other words, I could not be conscious of myself as located in time (and perhaps then as persisting over time) unless it was also true of me that I am conscious of the

conditions for the possibility of my location in time. Now, if this is the correct interpretation, it is *prima facie* difficult to attach must plausibility to at least read as a general claim. This is because there seems little reason to suppose the truth of

(1) *x* is conscious of *y*

requires the truth of

(2) *x* is conscious of the conditions which make *y* possible.

It might well be that Kant is not relaying on an instance of some such general principle, but rather certain peculiarities of time and the self. But if we insist on reading (1) as a logical consequence of the first refutation of idealism, then we have specific recourse to the contrast theory of meaning which entails there could not be a use for 'I', 'me', and other first person singular forms, unless there were a use for correlative second or third person forms. Then the ‘possibility of this time determination' actually denotes something very specific: that (relatively permanent) portion of reality which I am not identical with, which is yet a possible object of my perception. That this 'permanent' in perception exists then meets the condition 'makes possible my (self) determination in time'. The existence of this permanent in perception is not sufficient for my consciousness of my existence in time, but it is necessary because of the logic of 'self'. If we read Kant this way, then force is given to 'necessarily' in (1). Then when Kant says my consciousness of my existence in time is necessarily bound up with the consciousness of the condition of the possibility of this time determination, he means that given my self determination in time, my consciousness of the conditions for this could not fail to obtain. So, the connection between my self location and my consciousness of its conditions is not a contingent one. This rules out (or at least reveals as inadequate) any causal interpretation of the conclusion of the refutation of idealism (given that if R is a causal relation, then R is a contingent relation) and any interpretation weaker than a conceptual one. The conceptual interpretation of the refutation of idealism and of (1) is strong enough to warrant Kant's use of 'necessarily' in (1). This is because if one concept depends on another for its possible use then it generates incoherence to employ such a concept in abstraction from its conceptual contrast. This does not yield quite such a strong notion of necessity as this: *p* is necessary if and only if the negation of *p* is a contradiction. But it does yield this: *p* is necessary if and only if the putative negation of *p* is senseless. It is this second notion of necessity which is at work in the refutation of idealism.

Why Kant should feel justified in using 'consciousness of' in its second occurrence in premise (1)? It is trivially true that I could not locate myself in time unless I were conscious of myself as existing in time. But it is more contentious to argue that this would not be possible unless a second condition was also met: that I be conscious of my location in time. The reason depends on the application of the

contrast theory of meaning to the indexical pronouns 'I' and 'me'. Let us accept that it is a necessary condition of my self-location in time that I have a use for 'I' and 'me' (or some sort of name that may individuate in first person singular ascriptions). Then not only must there exist some portion of reality to which these words are not applicable (in order for them to have a use), I must be acquainted with such a portion of reality. Otherwise these words could not have a use for me. For a person to have a use for some denoting concept ‘C’, which depends on some second concept C' for its meaning, that person not only must be acquainted with what 'C' denotes but be acquainted with (or at least have knowledge of) what C' denotes. If this is accepted then Kant is justified in placing 'consciousness of' in its second occurrence in (1).

The Second Premise (Ref 2)

Premise (2) is open to all the objections raised against the inference from

'There exists a relatively permanent element which is a possible object of perception.'

to

'At least one physical object exists'.

If we allow Kant the existence of an enduring objective reality, on the basis of the first refutation of idealism, and allow him to identify this with at least one physical object, or with the class of physical objects, then we may allow premise (2) also. Without these two assumptions, (2) cannot go through and not much substantiation is given to either assumption. Kant's account must be supplemented to show that an objective permanent element in perception is not in itself sufficient to do the work of making possible my self-determination in time (even given the other necessary conditions, such as my existence, the existence of time, my consciousness and so on).

A proof is needed that this permanent element must be, or consist in, physical objects. If I am to discriminate myself from the rest of reality, if I am to have a use for 'I' and the second and third persons, then I must conceptually and, perhaps perceptually, divide what is into two mutually exclusive and collectively exhaustive portions: the part that I am, and the part that I am not. But if reality were wholly uniform this process would be difficult. There must, it seems, exist some features that portion of the world which I am possesses which distinguish it from that part I am not. The ‘permanent in perception’ must possess particulars, and variations over time, given that it is temporal. This does not give us 'is a physical object' but physical objects are very good candidates for objective particulars which are relatively permanent (relative to the observer), are possible objects of perception, and admit of local variations over time. It is psychologically compelling that the particularwe we perceive are physical objects and events. It is those collectively that are permanent. The inference from Kant's description of the permanent to a description uniquely true

of physical objects can perhaps not be a deductive one, but, if not, it is still an inductive one of considerable strength because the most likely hypothesis which makes sense of the permanent as described by Kant is: Physical objects are the mind independent objects of perception.

The Conclusion (REf 2)

Kant's use of 'in other words' suggest that (3), the conclusion resing? on (1) and (2), should follow from (1) and (2) as a re-writing of them. It is not quite that however, because in (3) Kant has it that the consciousness of my existence is an immediate consciousness of things outside me. Now, there is a straightforward interpretation on which this is just false. If it is true that

(1) A is conscious of B

and

(2) A is conscious of C

and

(3) B is not identical to C

then

(4) It is false that A's consciousness of B is identical with A's consciousness of C.

But there is an equally straightforward interpretation on which it might be true. If it is true that

(1) A is conscious of B

and

(2) A is conscious of C

(3) B is not identical with C

then

(4) It is true that A is conscious of B, and A is conscious of C.

If we read Kant this way then it amounts to: if I am conscious of my existence, and if I am conscious of the existence of things outside me I am conscious of both. Kant says 'at the same time', but this would seem not to be needed. If we insist on 'at the same time' that would preclude the possibility that there might be times that I am conscious of myself but not objects outside me, and the possibility that there might be times that I am conscious of things outside me but not myself. It would be false to deny either of these possibilities and not to Kant's purpose, so 'at the same time' asserts to much and is best forgotten.

A less literal construal of 'is' is not 'is identical with' but 'is both a necessary and a sufficient condition for'. Then we have:

(1) If it is true that A is conscious of B then it is true that A is conscious of C

and

(2) If it is true that A is conscious of C then it is true that A is conscious of B

and

(3) B is not identical with C

So, if it is true that I am conscious of my own existence in time, it is also true that I am conscious of the existence of objects outside, me, and if it is true that I am conscious of objects outside me, then it is true that I am conscious of my own existence in time. Now, the first conjunct of this proposition does?? follow from the refutation of idealism, but this is not the case with second conjunct. Although Kant has shown that my self-determination in time does require my consciousness of something permanent in perception, he as not shown the reverse. It remains possible until further argument is supplied, that I may be conscious of the existence of objects outside me without thereby determining myself in time. To see this as plausible, consider that a being might be conscious of objects which are objective mind-independent particulars but not regard them as such.

The converse proof; that perception of the constant requires self-determination in time is not needed for the refutation of idealism but clearly one could be produced which made use of the fact that 'it' 'that' and so on have no uncontrasted meaning, and in particular that 'objective' and 'mind-independent' require 'subjective' and 'mind-dependent' to have some use, and this, in an obvious sense requires reference to oneself.

2. The Objects of Outer Sense

Kant deploys an argument against idealism which lends itself more to a causal than a merely conceptual reading. In understanding the steps in this argument it is worth

bearing in mind Kant's insistence that all the materials of inner sense are acquired by the exercise of outer sense:

Third Argument Against Problematic and Dogmatic Idealism

(1) 'The representation 'I am', which expresses the consciousness that can accompany all thought, immediately includes in itself the existence of a subject.'

Freilich ist die Vorstellung: Ich bin, die das Bewusstsein ausdrückt, welches alles Denken begleiten kann, das, was unmittelbar die Existenz eines Subjekts in sich schließt.

(2) 'But it does not so include any knowledge of that subject, and therefore also no empirical knowledge, that is, no experience of it.'

Aber [es ist] noch keine Erkenntnis desselben, mithin auch nicht empirische, d.i. Erfahrung.

(3) 'For this we require, in addition to the thought of something existing, also intuition, and in this case inner intuition, in respect that is, of time, the subject must be determined.'

Denn dazu gehört, außer dem Gedanken von etwas Existierendem, noch Anschauung und hier innere, in Ansehung deren, d.i. der Zeit, das Subjekt bestimmt werden muss.

(4) 'Outer experience is really immediate, and only be means of it is inner experience - not indeed the consciousness of my own existence but the determinism of it in time possible.

'Dass äußere Erfahrung eigentlich unmittelbar sei, dass nur vermittelst ihrer, zwar nicht das Bewusstsein unserer eigenen Existenz, aber doch die Bestimmung derselben in der Zeit [...] möglich sei.

(5) 'It therefore follows that inner experience is itself possible only immediately, and only though outer experience.' (CPR 245-6, B276-7)

So dass folglich innere Erfahrung selbst nur mittelbar und nur durch äußere möglich ist.

(C) 'In order so to determine it, (the subject of inner intuition), outer objects are quite indispensable.'

The first premise (1) contains straightforwardly the claim that the thought 'I am', is part of what the transcendental unity of apperception consists in. This is made clear by Kant's statement that 'I am' expresses a consciousness which may accompany all thought. We are already familiar from the arguments of the transcendental deduction with the idea that it is a condition of experience - a condition of mental processes being ordered episodes in a single mind - that the 'I' think' in principle be capable of being a prefix to any of my thoughts. If this were in principle impossible, we could not talk of 'owned' or 'unified' experience at all. Now Kant is saying that 'I' think' implies 'I am', or that 'I am' is already implicitly asserted in the sentence 'I think. there is nevertheless a certain ambiguity introduced into the 'I think', 'I am' relation here. This is, it is not clear whether 'I think' straight forwardly entails 'I am' or whether in some sense the utterance, thought or production of the sentence 'I think' presupposes the truth of 'I am'. For example, I claim with respect to Descartes' Second Meditation that a plausible account of the necessity of 'I exist' is that the sentence may only be produced on condition it is true. Similarly, in Kant's case, it is arguable that 'I think' may only be thought on condition it is true. Kant does not say whether he thinks 'I think' logically entails 'I am', or whether it presupposes it in some weaker sense. But it is clear that he thinks that if 'I think' is thought then in some sense (at least an empty formal sense) there thereby exists a thinker of that thought, even though the thinker is not Cartesian.

The second premise reminds us that we are no entitled to logically derive any empirical or metaphysical conclusions about the nature of the self from the sentences constituting the transcendental unity of apperception. It is perhaps misleading to put Kant's view this way: I know that I am by the truth of the transcendental unity of apperception, but not what I am, because it suggests the logical possibility of metaphysical discoveries about myself. In fact, Kant thinks everything I can find out about myself is either formal or empirical. Now, clearly I can on Kant's theory find out much about myself through the exercise of inner sense. It is this that makes my knowledge of my own mental states possible, and this in a broad sense is a kind of empirical knowledge. So we have to read premise (2) as consistent with the theory of inner sense, but as inconsistent with any thesis on which *a priori* psychology is possible. For example, it is the denial that informative propositions about the self may be derived from the formal conditions for experience. The transcendental unity of apperception is not any kind of experience, and so cannot yield any empirical knowledge, because empirical knowledge is precisely knowledge acquired through experience.

Premise (3) provides the necessary mention of inner sense. (2) has denied the possibility of *a priori* self knowledge, (3) asserts the possibility of *a posteriori* self knowledge. (3) is clearly fully congruent with Kant's general epistemology: the claim that thought and intuition are each singularly necessary for the possibility of knowledge, and so that knowledge cannot be derived from one of these sources in abstraction from the other. In the sense that Kant intends it, it is perhaps analytic that the specific type of intuition required for psychological self-knowledge is inner

intuition, because 'inner' means 'mental', and inner sense is the only faculty which provides intuitions of the mental.

(3) also introduces a claim about time. Time is the form of inner sense (and, indirectly thereby the form of outer sense also). When Kant says the subject must be determined in time in respect of inner intuition there are at least two reasons for this. firstly, if inner sense is exercised then its objects necessarily appear temporal: this is just because time is the form of inner sense. But secondly, that my intuitions of my own mental states are temporally ordered is a necessary condition for the coherence of my experience. Arguably, if my experiences per impossible then they would not exist, or would not count as experiences at all. But given that they are temporal they are temporally ordered. I mean by this, that for any experience, call it 'e', if 'e' occurs then e occurs before during or after some other experience 'f'. Also, if e occurs then e lasts some measurable period of time, and with regard to ones present state of mind e is either past, present or future. If this temporal ordering did not obtain, then it is implausible to think that experience would be possible. Or, at least this is true; ones putative experiences would be wholly unintelligible to oneself and it is not clear that what is wholly unintelligible can be coherently alleged to exist. So, I think Kant's view is that the temporality of experience, and the temporal ordering of experience are both necessary conditions for the existence of experience.

Kant now needs premise (4), that inner experience is dependent on outer experience, that is, that inner experience would not be possible unless outer experience were possible. This is a crucial step in the argument, and we need to be clear about the precise nature of the dependence of inner or outer sense. At this crucial juncture Kant is himself unfortunately vague, but the following possibilities seem open. Perhaps 'X exercises inner sense' logically implies 'X exercises or is capable of ex ercising outer sense'. This looks too strong to be remotely plausible without a great deal of ex tra argument. Another possibility is another application of the contrast theory of meaning, so that it does not make sense to talk of a being's exercise of inner sense, unless it makes sense to talk of a beings exercise of outer sense. This is more promising, but it odes not seem prima facie logically impossible, that there should be beings capable only of detecting their own psychological states. But then we should have to ask for the criteria on which they found 'their' mental state intelligible and this might require that they be capable of an 'inner'/'outer' distinction or one similar. Finally, and perhaps most promising is Kant' statement at CPR 34 fn Bx that it is from 'things outside us' that 'we derive the whole material of knowledge, even for our outer sense'. Clearly this is an empiricist thesis about the possibility of psychological self-knowledge. It is the theory that the content of inner sense, what is intuited in inner sense, is strongly empirical in origin, that is; acquired through sense experience. Kant, it seems to me, needs this empiricist thesis as an additional premise in the particular argument against idealism considered here. He does not supply independent reasons for the truth of the premise, but some might be provided. For example, if we consider what is intuited in inner sense is what the content of ones mental states is, then there are strong isomorphisms between that

and the putative objects of outer sense. It is additionally arguable on an empiricist theory of the imagination that what sense experiences are possible for a subject places a qualitative constraint on what may be the content of that person's inner experience. If these two considerations have any force then Kant's additional empiricist premise does posses independent support. It we allow this premise then we thereby allow Kant this: inner sense is causally dependent on outer sense for its content.

We can now attach sense to Kant's claim that outer sense is immediate but inner sense is (by implication) mediate. If A mediates B then B is possible only on condition that A is possible, but if A is immediate then A is not dependent on B as a condition of its possibility. Kant intends at least this in (4), but perhaps this suggestion is also being made: the perception of the objects of outer sense is direct, in the sense that the perceptions of them are not representations of them. But the perception of the objects of inner sense are in just this sense indirect they are representations (to greater or lesser degree) of the objects of outer sense. I think Kant is at least leaving this possibility open here. He cannot explicitly include it as a premise because it could be taken as making use of 'objects of outer sense' in a question begging way: ie. as mind-independent.

It is not the bare consciousness of one's existence that requires inner and so, outer sense, but the determination of one's existence in time. Now, there is more than one way of interpreting this but a highly plausible one is as follows. the temporal location of my mind and its states: requires outer sense because if I am to locate a specific mental state or set of mental states of mine mentally then it must in principle be possible for me to have thoughts of this form: experience *e* occurred before E occurred (where 'E' denotes some event detectable through inner sense). This is because a purely internal ordering of my mental states *viz a vis* one another falls short of their being ordered in time in the sense of time as a whole. I cannot be said to have ordered my mental events temporally in complete abstraction from the time series in which all events are ordered. Indeed, not much sense could be attached to any such private time ordering because the idea of a temporal ordering is conceptually dependent on one single time order. We could allow Kant to supply this as an additional premise at this point of the argument, if we accept this claim in the Transcendental Aesthetic that there exists only one time, and that putatively numerically distinct times are really parts of one and the same time. If that is correct, then it is not possible that there should exist discrete mental time orderings, and any such putative private time orderings will in fact turn out to be located which in the single unitary time ordering. Now, this makes inner sense and inner sense time ordering strongly or conceptually dependent on outer sense and outer sense time ordering. To put it in summary form: Unless there were public, objective, time there could not be private, subjective, time.

There is another sense in which we can interpret the claim that inner senses; determination of the subject in time depends on outer sense. This is partially the conjunction of the claims just made. If it is true that inner sense depends on outer

sense for its content then it follows that inner sense is dependent on outer sense for its time ordering also. This follows given the assumption that there cannot take place any time-ordering unless there exists some content to be ordered. So, straight forwardly, if inner sense requires outer sense for its content, then the subject requires outer sense for its 'self determination in time'. This interpretation of Kant is fully consistent with his insistence that the transcendental unity of apperception is purely formal and can provide no introspectible content. Clearly, Kant needs this premise (contained in (1)) here because if there were some non-empirical source of self knowledge then the possibility would seem to be opened up that in principal there could be some non-empirical source of the materials for time ordering. Clearly, this is something Kant cannot possibly allow.

If the premises as so far interpreted are true, then premise (5) follows soundly from them. It is already suggested by (4) that inner experience is mediate, and clear senses have been supplied in which it is reasonable to assume that inner experience is possible only on condition there is outer experience. So we may pass straight to the conclusion.

(C) is in fact crucially ambiguous. The problem lies in the meaning of 'outer objects'. If outer objects' just means 'the objects of outer sense', or the content of outer sense' then the conclusion follows from the premises as interpreted. But, such a conclusion is much to weak to damage the problematic or dogmatic idealist. This is because Descartes and Berkeley each accept a distinction which roughly corresponds to that between inner and outer sense. They each accept a distinction between psychological self-consciousness, and sense experience. What is left crucially undecided here is whether the objects of outer sense are 'mind-independent objects of outer sense' then the conclusion does not follow from the premises. All that even the strongest considerations about time have established is this: there exists an objective time order, within, which I may temporally order my own mental states. We do not yet have this: this object temporal order presupposes the existence of mind-independent physical objects, which are identical (at least partially) with the objects of outer sense. Kant has supplied no premises far to bridge this logical gap but really just one is required. this is: time presupposes physical objects, in particular, objective or mind-independent physical objects. If (for example) a quasi-Leibnizian, or relational, theory of time could be proven, and supplied as an additional premise for Kant's conclusion, then his refutation of idealism is sound.

A part of an argument from the objectivity of time to the objectivity of physical object is to be found at CPR 246, B277-8:

'We are unable to perceive any determination of time save through change in outer relations (motion) relatively to the permanent in space (for instance the motion of the sun relatively to objects on the earth').

Nicht allein, dass wir alle Zeitbestimmungen nur durch den Wechsel in äußeren Verhältnissen (die Bewegung) in Beziehung auf das Beharrliche im Raume (z.b. Sonnenbewegung in Ansehung der Gegenstände der Erde,) vornehmen können.

Although Kant elsewhere repudiates a Leibnizian theory of space, he here advocates one crucial aspect of it: *viz* that the perception of motion depends on the perception of (relative) permanent. So the sequence of dependence is this: there could not be time without change; there could not be change without motion and there could not be motion without (relative) permanence. This is a powerful supplement to Kant's argument because if he has established that time exists objectively ie, that there is a mind-independent time order- then it follows that there exist entities which are objectively time-ordered. Now Kant is supplying a description for these entities which they must confirm to if time is to exist. Arguably now, if time depends on change and so on, motion, and so on permanence, then objective time depends on objective change, objective motion and so objective permanence. So idealism, as the view that there does not or may not exist any mind-independent permanence is false. Supplemented in this quasi-Leibnizian way the power of the refutation of idealism is effectively increased. In fact it is both valid and sound.

Before leaving the Refutation of Idealism chapter we need to note a provision Kant makes about his refutation: From the fact that the existence of mind-independent objects is a necessary condition for psychological self knowledge it does not follow that any particular object of outer sense is in fact a genuine mind-independent object. Here is his statement of that view:

'All that we have here sought to prove is that inner experience in general is possible only through outer experience in general.'

Es hat hier nur bewiesen werden sollen, dass innere Erfahrung überhaupt, nur durch äußere Erfahrung überhaupt, möglich sei.

'From the fact that the existence of outer things is required for the possibility of a determinate consciousness of the self, it does not follow that every intuitive representation of outer things involves the existence of those thinkings.'

Daraus, dass die Existenz äußerer Gegenstände zur Möglichgkeit eines bestimmten Bewusstseins unserer selbst erfordert wird, folgt nicht, dass jede anschaulihce Vorstellung äußerer Dinge zugleich die Existenz derselben einschließe.
(CPR 247, B278-9)

Kant is asserting that the truth of his claim that inner sense is possible only on condition there is outer experience of mind-independent objects does not entail the mind-independence of every putative object of outer-sense. All he has proved is that if inner experience is possible then some outer experience of mind independent

object's must also be possible. Clearly it does not follow from that claim that all outer experience is of mind-independent objects. Examples of objects of outer sense which are not mind-independent include dreams and delusions. (CPR 247, B278) It is a moot point whether Kant should have called these objects of outer sense or inner sense. This does not much matter, or is at least a matter for stipulation. They appear, or may appear, to be mind-independent and so on those grounds have one of the features of the paradigmatic objects of outer sense, but they are not mind independent so possess one of the features of paradigmatic objects of inner sense. Their status is thus clear. When Kant says such delusions 'are merely the reproduction of previous outer perceptions' (CPR 247, B278) he means that these non-mind independent objects of outer sense could not exist unless there existed mind-dependent objects of outer sense could not exist unless there existed mind-dependent objects of outer sense of which they are reproductions. So, not only is the content of inner sense dependent on the content of outer sense but the illusory features of outer sense are dependent upon its ordinary and veridical use.

There is then some justification for Kant's claim that 'the game played by idealism has been played against itself'. (CPR 245, B276) If it is true that knowledge of ones own mental states depends upon acquaintance with mind-independent physical objects then scepticism about the latter depends upon its vacuousness.

The Fourth Paralogism in the Second Edition of the *Critique of Pure Reason*

Kant's argument against the validity of the Fourth Paralogism is

(1) 'That I distinguish my own existence as that of a thinking being, from other things outside me - among them my body – is an analytic proposition'

Ich unterscheide meine eigene Existenz, als eines denkenden Wesens, von anderen Dingen außer mir (wozu auch mein Körper gehgört), ist ebensowohl ein analytischer Satz

(2) 'For other things are such as I think to be distinct from myself'

Denn andere Dinge sind solche, die ich als von mir unterschieden denke.

(3) 'But I do not thereby learn whether this consciousness of myself would be even possible apart from things outside me through which representations are given to me.'

Aber ob dieses Bewusstsein meiner selbst ohne Dinge außer mir, dadurch mir Vorstellungen gegeben werden, gar möglich sei […] weiß ich dadurch gar nicht.

(C) '(I do not thereby learn) whether therefore I could exist merely as thinking being (ie without existing in human form).

[Ich weiß dadurch gar nicht] ob ich also bloß als denkend Wesen (ohne Mensch zu sein) existieren könne. (CPR 370, B409)

The first premise is only true as qualified by the second. The conjunction of 'I distinguish my own existence as that of a thinking being from other things outside me' and 'Other things are such as I think to be distinct from myself' is analytic but the first conjunct taken alone is not analytic. It is not true by definition, nor is it knowable to be true simply by conceptual analysis, that a person draws a distinction between himself as thinker and physical objects. To construe that claim as analytic would be to make the falsity of materialism a matter of logic, but it is clear that the sentence 'I am a physical object' is not self-contradictory. This is not to deny that the self-not self distinction might indeed be marked to coincide with a thinker-physical object distinction, but that demarcation would require an argument including at least some synthetic premises; for example the claim that it is not the body, nor any part of it, which thinks. That claim is synthetic because it is not self-contradictory to assert that it is the brain that thinks and the brain is part of the body.

Not only is the claim embedded in (1) taken alone is not analytic, but (2) taken alone is not analytic either. It is true by definition that other things are distinct from myself, 'other things' means 'things distinct from myself' but it does not follow straightforwardly from the fact that other things exist distinct from me that I think that they do, and that is what is asserted in (2). It might be, for example, that physical objects exist independently of my experience of them: that they exist whether I perceive them or not. But, it might be the case that I think idealisms is true: that is, I think physical objects are ex hausted by my experience of them. Then although it would be true that things exist distinct from me it would not be true that I believe it, so the conjunction of the claim that things exist independently of me and the claim that I think this true cannot be analytic. This is true even if as an idealist I still recognised the analyticity of the sentence other things exist independently of me. The analyticity of that claim is of course no threat to the idealist because for him it just ex plicates the concept of a physical object erroneously.

But there is a kind of self-not self distinction that even the most solipsistic of idealists must accept. Thus the straightforward claim that I am not the objects I perceive or think about but what perceives or thinks about them (self-consciousness ex cepted). This claim is analytic because it just amounts to; I am not anything other than myself. It is not analytic that anyone believes this, but it might be conceptually impossible to understand it without believing it. So the sentence 'I think I am not identical with any things other than myself' is perhaps necessary in the sense that it may be coherently thought only on condition it is true, even if it is not analytic.

(3) says that it is not established logically by (1) and (2) whether self-consciousness would still be possible even if the objects which I experience as

independent of myself did not exist. I think this is right it does not follow from the putative analyticity of a certain self-not self distinction, or from the analyticity of someone's making such a distinction that self-consciousness would or would not be possible if there existed no objects independently of oneself. But that there is some criterion for marking a self- not self distinction is a necessary condition for self consciousness. This is because we could not have a use for 'self-consciousness' at all unless we had a use for 'that which is not oneself' or some synonymous expression. But it is open to the rational psychologist, or to a mind-body dualist, to invoke alternative criteria for the individuation of selves: criteria which do not make the self-not self distinction rely on the existence of physical objects as causes of ones experiences. All Kant wishes to establish is that the dualist, to invoke alternative criteria for the individuation of selves: Criteria which do not make the self-not self distinction rely on the existence of physical objects as causes of ones experiences. All Kant wishes to establish is that the dualist conclusion does not follow from (1) and (2). He is not concerned at this point to repudiate every ground for dualism, just to deny the rational psychologist his conclusion of myself as a thinking being even if there were no physical objects.

The conclusion follows because it is true that it cannot be established by (1) and (2) whether I could be conscious of myself as a thinking being if there existed no physical objects as it does not follow either that I could exist as such a being. If it could be known that I could be conscious of myself as such a thinking being then it follows that I could be conscious of myself as such a being, because if it is possible for me to know that P is true then it is possible that P be true, and if 'conscious of myself as' entails 'conscious of the fact that I am' then it follows that I could be such a being, peculate if it is a fact that P then P is true, but the fact that I distinguish myself from what is not myself, even if I distinguish correctly, does not of itself imply that I could exist independently of the objects from which I distinguish myself. That possibility is not precluded either, but it is not obtained by the argument the rational psychologist adduces. (C) does not follow from (3) alone because, in general, from that fact that P's truth cannot be known it does not follow that not-P. If P's truth cannot in principle be known then one reason for that might be that P is false, but it is not logically impossible that there should be reasons why P could not be known to be true even if P is in fact true.

I conclude that Kant's objection to the validity of the Fourth Paralogism is conclusive: it is a paralogism because the conclusion is not logically implied by its premises. The criticisms of the four Paralogisms taken jointly are not sufficient for Kant's general conclusion that analysis of the way we think about ourselves yields no metaphysical knowledge of our nature, because it is not evident that Kant has considered all possible modes of thinking about ourselves. But that conclusion does go through if we accept his thesis that no metaphysical pretence whatsoever follows from any set of purely formal and *a priori* sentences. Then it will follow in particular that no metaphysical claim about the self could be deduced from any set of formal and a priori facts about the ways in which we think about ourselves.

V

The Paradoxes of Inner Sense

'The fact that man can have the idea 'I' raises him infinitely above all the other beings living on earth. By this he is a person'

Immanuel Kant *Anthropology* 10

As there is no single sustained characterisation of inner sense in the *Critique of Pure Reason* the theory has to be reconstructed from the many passages in which reference is made to it throughout the book. (1) The nearest Kant approaches to a definition is:

'"Inner sense", by means of which the mind intuits itself, or its inner state, yields indeed no intuition of the soul itself as an object; but there is nevertheless a determinate form (namely time) in which alone the intuition of inner states is possible and everything which belongs to inner determinations is therefore (re)presented in relations of time.' (CPR 67-8, A23, B37)

Der innere Sinn, vermittelst dessen das Gemüt sich selbst, oder seinen inneren Zustand anschauet, gibt zwar keine Anschauung von der Seele selbst, als einem Objekt; allein es ist doch eine bestimmte Form, unter der die Anschauung ihres inner Zustandes allein möglich ist, so dass alles, was zu den inner Bestimmungen gehört, in Verhältnissen der Zeit vorgestellt wird.

What does this mean? An intuition according to Kant's usage is 'that though which (knowledge) is in immediate relation to (objects)'. (CPR 65, A19, B34) So if a person intuits an object then that person experiences it in a way that entails being directly acquainted with it. In inner sense the mind intuits itself, so Kant thinks minds may experience themselves and so be directly acquainted with themselves.

Provisos need to be made about this reading. Firstly, the inaccessibility of things in themselves implies that minds are never directly acquainted with themselves just as they really are, only as they appear to themselves.

Secondly, Kant insists that a mind has a no experience of itself as a soul: as a non-physical substance. This is why Kant includes the sub-clause 'or its inner state'.

I take it a mind only intuits itself by intuiting one of its own states and has no acquaintance with itself independently of experience of one or more of its states.

'Inner sense' is then the name of a faculty or capacity for introspection which is explicitly contrasted with outer sense. If a person exercises outer sense then it follows that that person exercises one or more of the five senses: sight, hearing, taste, touch and smell. 'Outer sense' is a generic term for these faculties.

But Kant also uses the word 'sense' to characterise a mind's acquaintance with its own states, and this suggests a degree of similarity between the inner and outer faculties even though it is not obvious that there exists any sort of sense other than the traditional five. Before deciding the appropriateness of this extended use of 'sense' we should examine Kant's inner-outer distinction.

Inner and Outer

Paradigmatically, a person's mental states are 'inner' and physical objects and their physical states (including a person's body) are 'outer'. Kant marks this distinction by saying the objects of outer sense possess both spatial and temporal properties, or are located in both time and space, but the objects of inner sense possess only temporal properties or are located only in time and not also in space. So the inner-outer distinction is logically dependent on a spatio-temporal - merely temporal distinction.

Now, a number of difficulties arise for the view that mental states have no spatial properties. It is Kant's view that if we could not exercise outer sense we could not exercise inner sense either. This is because he subscribes to the empiricist epistemology according to which the content of introspection - what is introspected in introspection - always has its origin in the exercise of the five senses. At CPR 34, iii fn. for example, he speaks of 'things outside us (from which we derive the whole material of knowledge, even for our inner sense)'.

But, if that is right, then although it does not follow with logical certainty that that which has a spatial origin will itself possess spatial properties the onus is on the Kantian to show why this is not the case with the contents of inner sense.

Further it seems intuitive that people do at least tacitly ascribe to their mental states some spatial properties. For example the images of the visual imagination may appear to possess height and breadth, even if this height and breadth could never the measured. On top of that it is possible to imagine three dimension physical objects with extended and coloured surfaces, and even if it is not possible to touch them it is possible to imagine touching them. So, for any spatial characteristic apparent to outer sense there seems no a priori objection to the postulation of an inner sense correlate: the introspective awareness of an image of that characteristic.

A further difficulty is that the thesis that mental states possess at least some spatial properties is reinforced by Kant's use of 'sense' in 'inner sense'. Although there do not exist inner physical senses it is a part of ordinary language to speak of 'the mind's eye' and 'the mind's ear' {and there seems no a priori reason not to postulate further the mind's nose, the mind's tongue, or the mind's fingers). If it is true that

people partly think in spatial images then this gives a certain appropriateness to Kant's use of 'sense' here but also introduces a certain tension with the doctrine of the mere temporality of the mental: it implies incoherently that the mental both possesses and lacks spatial properties.

Next, it is not clear that anything could possess only temporal properties. Not much sense can be attached to a doctrine which has it that 'X exists' is true, but all that is further true of X is that X lasts, had an origin, will have an end, came after Y, before Z etc. And this is clearly at variance with any characterisation of the mental using other than merely temporal predicates: saying of one's emotions that they are intense etc.

Some of this tension may be dissipated by claiming that if X is an object of inner sense then X possesses the following two sorts of characteristic: X is mental and X is temporal, where 'X is mental' does not entail any sentence which ascribes a spatial property to X . But denying that entailment requires more argument than Kant supplies.

Finally, the terms 'inner' and 'outer' are themselves spatial predicates, so the sort of thinking that enables that distinction to be drawn is itself spatial, and it is partly a kind of thinking about oneself. The role of spatial metaphor in our self-conception is a matter for piecemeal investigation but I just note its existence here to point out a difficulty for making the inner-outer sense distinction rely on the temporal- spatio-temporal distinction.

Time and the Self

A less problematic aspect of the theory of inner sense is the thesis that if I experience one of my own mental states then that experience is necessarily temporal. Kant insists on this in particular but it is in fact a logical consequence of his general thesis that there is not any experience that is not temporal. When he says time is a 'determinate form in which alone the intuition of inner states is possible' he means at least that introspection would not be possible if time did not exist.

There is a sense in which it is difficult to imagine that this claim could be false. This is because if I am aware of one of my mental states - call it 'X ' - then it would seem to be a condition of my being aware of X *qua* one of my states that I be aware that X is a member of a series of mental states which make up a single mind capable of being called 'mine'.

It is not logically impossible that a person's experience of their own mental state should not be like this: just unusual. If we allow a distinction between a person being who he is and a person knowing who he is then it is not self-contradictory to maintain that a person may at one time t^1 be directly acquainted with one of his mental states 'X ' yet not at all be conscious that X was preceded by a series of earlier states, nor be conscious that X will be followed by a series of subsequent states. There are psychiatric cases which fit this description but it seems prevalent characteristic of such cases that the patient has only a radically diminished concept of

self. If having a concept of oneself as enduring over time is a condition of being a self over time and if being a self is a condition of being a person then it is not possible that a person's experience should be be of this punctual kind. In any case, it is clear from Kant's use of the plural 'relations of time' that Kant intends more than the analytic thesis that all a persons experiences are present experiences.

If we take Kant to be describing the minimal psychological conditions required for a normal mind's acquaintance with itself rather than to be making a psychological generalisation about any possible mind, then the following plausible view may be ex tracted. It is a condition for any experience of mine appearing as such to me that it be located by me as a member of a series of experience with which I could also in principle also be directly acquainted. This does not ascribe to Kant the rather implausible view that if I have an experience then I am automatically conscious of having it, just that if it is one of mine then it must be logically possible that I be conscious of having it.

So, although inner sense is not perpetually exercised, if it is exercised then its content is temporal in these ways: An object of inner sense, X , is experienced as after some number of earlier experiences A, B, C, and as preceding some number of subsequent experiences, Y, Z, and perhaps as simultaneous with some other experience W. In addition, X is experienced as possessing some duration, as having an origin and as ending.

Not all of these conditions need always be fulfilled perhaps, but that they are at least partly fulfilled is a condition of X being part of a self-conscious mind. This is why Kant says 'everything which belongs to inner determinations is [[...]] (re)presented in relations of time'.

It is now possible to interpret the following remarks:

'Through inner experience I am conscious of my existence in time (consequently also for its determinability in time), and this is more than to be conscious merely of my representation. It is identical with the empirical consciousness of my existence.' (CPR 35, Bx l) (Kant’s italics)

This is the connection between time and the self: That a mental state with which I am directly acquainted be thereby located in a temporally ordered series is partly constitutive of its being mine.

Kant draws a distinction here between 'my existence' and 'my (re)presentations', which is a distinction between myself and my mental states. It would, in principle, be possible for a person to be acquainted with a mental state but not thereby be aware of it as one of his own. The reason that does not happen is partly to be ex plained by that state's temporal relations to other states. It is that temporal ordering and the awareness of it which facilitates the transition from an awareness of a mental state that is in fact one of one's own to an awareness of a mental state that is one of ones own as one of one's own. Only when this second circumstance obtains is it legitimate

on Kant's theory to talk about my acquaintance with my existence. this whole psychological structure Kant calls 'the empirical consciousness of my existence'.

There is much plausibility in this account. It relies on the principle that if A is a part of B, and if a person is aware of A then that person is aware of B. So, if any mental state of mine is a part of my existence, a part of what my existence consists in, then if I am aware of one of my mental states then I am aware of my existence.

Kant also believes that if I am aware of one of my mental states qua mine then I am also thereby aware of my existence qua mine. By calling this awareness of one's own existence 'empirical' he is drawing attention to two features of it: Knowledge of one's own existence is through experience and not purely intellectual, not by ' intellectual intuition'.

But he also, intends this awareness to be something commonsensical. It would not be reasonable to make introspection, which is prevalent, depend upon a kind of philosophical consciousness of one's own existence, which is perhaps comparatively rare. Kant's meaning here can be captured by: A person is conscious of their own existence if and only if they would sincerely answer 'yes' to 'Do you exist?', where this is consistent with that person never having thought 'I exist'. Inner sense makes this answer possible, not by providing unique access to something inscrutable captured only by 'my existence' but by facilitating awareness of one's own occurrent mental states.

So, if a state I am aware of is occurrent then it exists, and if it is one of mine then I exist, so if I am aware of one of my mental states I am aware of my own existence. This is not the same as explicitly entertaining the though 'I exist', which is not required for inner sense to be exercised.

If it is true that the mental is necessarily temporal, it still does not follow that only the mental is temporal. But, on a rather idealist reading of Kant, this is a conclusion to which he subscribes:

'Time is nothing but the form of inner sense, that is, of the intuition of ourselves and of our inner state. It cannot be a determination of outer appearance.'

Die Zeit ist nichts anders, als die Form des inner Sinnes, d.i. des Anschauens unserer selbst und unsers innern Zustandes. Denn die Zeit kann keine Bestimmung äußerer Erscheinungen sein.
(CPR, 77, A33, B49)

This needs treating with care if Kant is to be read as saying something consistent. His view is that time is transcendentally ideal but empirically real, so time is a condition of all our experience -inner and outer - and both inner and outer objects do appear to us to be temporal. Because time is the form of inner sense - the way in which we experience - it is not possible for any of our experience not to appear temporal to us. So we have to read 'determination' above to mean 'non-empirical

determination' because clearly time is an empirical determination of outer sense. As he puts it:

'Nevertheless, in respect of all appearances, and therefore of all the things which can enter into our experience, it is necessarily objective.'

Nichts desto weniger ist sie in Ansehung aller Erscheinungen, mithin auch aller Dinge, die uns in der Erfahrung vorkommen können, notwendigerweise objektiv.
(CPR, 78, A35, B51).

But Kant does not rest content with the transcendental-empirical distinction. He deploys here a subjective - objective distinction. Read idealistically, Kant is saying time is subjective in that it pertains only to the psychology of the subject, and does not pertain to reality as it is independently of the subject's experience of it. Kant seems quite explicit on this point:

'Time is therefore a purely subjective condition of our (human) intuition [[...]] and in itself, apart from the subject, is nothing'

Die Zeit ist also lediglich eine subjektive Bedingung unserer (menschlichen) Anschauung […] und an sich, außer dem Subjekte, nichts.
(CPR, 77-8, A35, B51).

Although Kant can mount plausible arguments for the transcendental ideality and the empirical reality of time, I see little to substantiate the claim that time is subjective and not also objective in the senses defined above. Certainly, from the conjunction of the propositions that time is a condition of experience, that it is empirically real and that it is a property of mental states it does not follow that it does not exist independently of those mental states (that is, whether those mental states existed or not).

A more modest conclusion (which Kant occasionally shows signs of endorsing) is that form the fact that time is transcendentally ideal and empirically real, and does pertain to the psychology of the subject it does not follow that time is objective in the sense of existing independently of the psychology fo the subject. This would leave it open whether reality as it is in itself is temporal or not. This more modest claim does not need an argument - it is sufficient to point out that the objectivity claim cannot be derived from the sentences of the theory of time so far.

I conclude that Kant has not proved that time is only subjective, even though the sentences of his theory are logically consistent with that thesis.

Is Inner Sense Paradoxical?

Although we cannot allow Kant the 'nothing but' here, this passage draws attention to a paradox about self consciousness which Kant thinks every theory of the self succumbs to:

'(Time) can be nothing but the mode in which the mind is affected through its own activity namely through this positing of its representation), and so is affection by itself.' (CPR 87, B67-8)

[Zeit kann nichts anderes sein] als die Art, wie das Gemüt durch eigene Tätigkeit, nämlich dieses Setzen seiner Vorstellung, mithin durch sich selbst affiziert wird.

The difficulty is that in inner sense a mind is both active and passive, both subject and object, that is, both what experiences and what that experience is of. Clearly there is no formal contradiction in maintaining such a description applies to a self-conscious mind but Kant thinks some further ex planation of this possibility is called for. This is his argument for the difficulty:

(1) 'The conciousness of self (apperception) is the simple representation of the “I”.'

Das Bewusstsein seiner selbst (Apperzeption) ist die einfache Vorstellung des Ich.

(2) 'If all that is manifold in the subject were given by the activity of the self, the inner intuition would be intellectual.'

Wemm dadurch allein alles Mannigfaltige im Subjekt sesttätig gegeben ware, so würde die innere Anschauung intellektuell sein.

(3) 'In man this consciousness demands inner perception of the manifold which is antecedently given in the subject, and the mode in which this manifold is given in the mind must, as non-spontaneous, be entitled sensibility.'

Im Menschen erfordert dieses Bewusstsein innere Wahrnehmung von dem Mannigfaltigen, was im Subjekt vorher gegeben wird, und die Art, wie dieses ohne Spontaneität im Gemüte gegeben wird, muss, um dieses Unterschiedes willen, Sinnlichkeit heißen.

(4) 'If the faculty of coming to consciousness of oneself is to seek out (to apprehend) that which lies in the mind, it must affect the mind, and only in this way can it give rise to an intuition of itself.'

Wenn das Vermögen sich bewusst zu werden, das, was im Gemüte liegt, aufsuchen (apprehendieren) soll, so muss es dasselbe affizieren, und kann allein auf solche Art eine Anschauung seiner selbst hervorbringen.

(5) 'But the form of this intuition, which exists antecedently in the mind, determines in the representation of time, the mode in which the manifold is together in the mind.'

Deren Form aber, die vorher im Gemüte zum Grunde liegt, die Art, wie das Mannigfaltige im Gemüte beisammen ist, in der Vorstellung der Zeit bestimmt.

(6) 'It then intuits itself not as it would represent itself if immediately self active but as it is affected by itself, and therefore as it appears to itself and not as it is.'

Da es dann sich selbst anschauet, nicht wie es sich unmittelbar selbsttätig vorstellen würde, sondern nach der Art, wie es von innen affiziert wird, folglich wie es sich erscheint, nicht wie es ist.

(c) 'The whole difficulty is as to how a subject can inwardly intuit itself; and this is a difficulty common to every theory.' (CPR 88, B68-9).

Hierbei bereuht alle Schwierigkeit nur darauf, wie ein Subjekt sich selbst innerlich anschauen könne, allein diesde Schwierigkeit ist jeder Theorie gemein.

(1) is included as part of the claim that one sort of awareness - apperception - does not entail any paradox . Apperception is thinking about oneself, and this is to be distinguished from ex periencing oneself. I take it Kant is here talking about empirical apperception - the ordinary thinking about ourselves and our mental states - and not the transcendental unity of apperception which is a formal condition for experience.

Then we can understand (1) as a part definition of apperception; it is the thinking of the thought 'I think'. That any being capable of apperception in this sense be capable of thinking 'I think' is a necessary truth. This does not ex haust what empirical apperception is, because we may still ask what is thought. In other words, 'I think' gives the minimal grammatical structure of the sel-conscious thought that is apperception, but if that apperception is empirical then that thought must have some semantic content as well as that grammatical structure. The answer is that any thought that the thinker of 'I think' thinks may be embedded in a sentence as a clause governed by 'I think', and any thought that any self conscious being is capable of entertaining may be prefaced by 'I think'.

But this content to apperception is not what is essential to it: that is provided by 'the simple representation of the 'I'. Kant's thinking here is that if a being is self conscious then it has a use for 'I' (or some word with an analogous grammatical function) and, conversely, if a being has a use for 'I' then that being is self-conscious. This is the force of 'simple' in (1), which should not be taken to ex clude the fact, which Kant subscribes to, that empirical apperception has semantic content.

Kant then points out by (2) that if *per impossibile* all our self-consciousness took the form of empirical apperception, and none of it were derived from inner sense, then we would be endowed with what he calls 'intellectual intuition'. Apperception is a faculty for thinking, not for intuiting. That is, it is a part of the intellect or understanding (*Verstand*) and not a part of sensibility. It is that part of the understanding which is thinking with a special object: oneself Now, there are at least two ways of taking the claim that if we had purely apperceptual self-consciousness not derived from inner sense then our intellects would be intuitive. Firstly, this might just mean that I can think any self conscious thought I like - including any thought about any experience of mine-without my having previously detected that experience through introspection. This makes the intuitive intellect thesis the denial of a certain causal thesis. It is the denial of the empiricist view that I can think about one of my mental states only if I am or have been directly acquainted with one of my mental states through introspection. On the intuitive intellect account I may know what I am ex periencing without being thus caused to know by any direct awareness of that experience. The other way of taking Kant's claim, is that if a being exercises an intuitive intellect then there is in a sense no difference between that being's thinking what its mental state is and it's ex periencing what its mental state is. In particular, one way of marking that distinction is ex cluded: it is not logically possible for such a being to think mistakenly about one of its own experiences. This is because in the exercise of the intuitive intellect the manifold, or the unified experiences of the individual, are in fact generated by the activity of the intellect. Perhaps this intellectual faculty, which Kant clearly thinks at least logically possible even though human beings do not possess it, may be made more intelligible by this analogy. There is a distinction between imagining an emotion and ex periencing an emotion, imagining a pain and ex periencing none. But in the case of the intuitive intellect we have to think of what we experience as assimilated to a faculty rather like the imagination in this respect: Although the intuitive intellect is not the imagination, in both the subject is active in the production of its contents. But in inner sense the subject is passive, or simply subject to his experiences. I mention this analogy merely as a device to make more intelligible the idea of an intuitive intellect. The exercise of the intuitive intellect in our inner intuition would itself generate the mental states with which we are intellectually acquainted

(3) makes explicit Kant's insistence that human beings do not possess an intuitive intellect. 'This consciousness' refers back to 'apperception' or 'the consciousness of self' in (1), so (3) is partially the claim that empirical apperception would not be possible unless there were inner sense. In other words, empiricist psychology is true of humans. Unless we persons each had an ex periential acquaintance with our own mental states, then it would be impossible for us to think (correctly or incorrectly) about our own mental states either. I take it this is a causal thesis and not a logical or conceptual claim, because Kant does think that an intuitive intellect is a logical possibility even though it is not actual in the case of humans. If Kant were to make the empiricist thesis a necessary truth, then that would make the

doctrine of the intuitive intellect a logical impossibility. It is not, so the empiricist theory is not a necessary truth.

What is the nature of this dependence of intellect (or apperception) on experience (or inner intuition)? It is a special variant of Kant's general thesis about the relations between understanding and sensibility. In the case of we rational self-conscious beings, our understanding is constrained by our experience: the categories have only an empirical use. It follows that we can only think meaningfully about the content, or possible content, of our experience. If apperception is a version of understanding, and if inner sense is a version of sensibility then it follows that a person may meaningfully think about his own mental states only to the ex tent that he has or could have direct ex periential acquaintance with those states.

This sort of dependence is not only a causal one, it is properly called a transcendental one. It is an ex planation of how a certain kind of knowledge is possible, and knowledge of how knowledge is possible is called by Kant 'transcendental thesis'. However, a causal interpretation is almost certainly intended too.

If we ask what is thought if I think about my own mental states in empirical apperception, then that is an inquiry about the content of my self-conscious thinking. Or, to put it another way, it might be an inquiry about what my self-conscious thought is about. Now, on Kant's theory it is left rather obscure what the relation is between the content of thought in the sense of what is thought, or that which is thought, and on the other hand what thought is about, or what is thought about. But this much is clear, unless I could be acquainted with my mental states though inner sense - through an inner intuition of them - then I could not think about them - they could not be what my thought is about - and they could not provide the content of my thought either - what is thought, in the sense perhaps of names and predicates which designate those states.

On either construal, this is a causal thesis distinct from the transcendental thesis. The transcendental and causal readings each give a different meaning to Kant's 'demands', and his 'antecedently given'.

On the transcendental account 'demands' means, roughly. 'presupposes' so that if there is apperception then that presupposes there is or has been some exercise of inner sense. Also, that exercise is not necessarily chronologically antecedent to the exercise of apperception, but is antecedent in the sense of prior to transcendentally. I mean by this: A is transcendentally prior to B if and only if B is possible only if A is possible but it is not the case that A is possible only if B is possible. ('transcendentally prior to' is not a symmetrical relation).

But on the causal reading 'demands' means 'has to be caused by', so that if there is apperception then this is depended upon a certain causal condition: that there be or have been an exercise of inner sense. So 'antecedently' can then be interpreted as 'chronologically prior to' because on the causal thesis, if the exercise of inner sense is a causal condition for the exercise of apperception and if it is true that if an event C is a cause of some event E then E exists necessarily chronologically subsequent to C,

then the exercise of inner sense exists necessarily before the exercise of apperception it causes.

If we ask what sort of causal condition the exercise of inner sense is, then this is best understood as a necessary, and not as a sufficient condition nor as both a necessary and a sufficient condition. This is because although it is true that (for humans) there could not be apperception unless there were inner sense it is still true that it does not follow that there is apperception from the fact that there is inner sense. If the existence of some exercise of inner sense were either sufficient or else both necessary and sufficient for the exercise of apperception then it would follow that if there is inner sense then there is apperception. But the kind of necessary condition involved is not logical, it is either causal or transcendental, so there is room for doubting whether it follows from 'there exists apperception' that 'there exists inner sense'. It is not clear that Kant thinks that someone who affirmed the first of these but denied the latter would have contradicted himself, but clearly they would have done so if the second followed from the first with logical necessity.

The reason Kant includes (1), (2), and (3) is that intellectual intuition is unproblematic: It generates no paradox of subjectivity and objectivity and activity and passivity. This is because it is wholly spontaneous: There is no active-passive distinction in intellectual intuition because there is no difference between what it thinks and what is experiences.

But (3) draws a strict contrast between the case of the intuitive intellect and the case of human self-consciousness. We self-conscious rational persons are intellectually active (we share that much with the hypothetical being who exercises the intuitive intellect) but we are with regard to our experience passive. I take it this at least means that we do not choose what the course of our experience is or will be, (in the sense that for example we might be free to imagine different things). But there is another distinction here too. This is between what a person does and what happens to a person.

We can think of this as the distinction between being an agent and a subject, using these terms so that a person is an agent as performer of actions (including mental actions) but is a subject as subjected to experiences. Now we can say that a human being in Kant's view is an agent with respect to thinking, but a subject with respect to experience. A being endowed with intuitive intellect in contrast is an agent in both thought and experience. Its experiences, as well as its thoughts, feature amongst its performances.

This is why in (3), when describing the human case, Kant says our inner perception of the manifold is 'given' and is 'non spontaneous' and is a 'sensibility'. If X is given then X is not the performance of any agent even though X is in some sense undergone by that agent: It features amongst the experiences of the agent not qua agent but qua subject. By 'spontaneous' Kant means 'both active and free', so that if X is spontaneous then X is a performance of some agent, and X, logically, might not have been performed, if the agent had not chosen to perform X .

Finally, if X is not spontaneous then Kant has said X is a sensibility. This means that if X is not free and actively performed then X is an experience (not a thought or action), and experiences are passive. Kant uses (as we have seen) the terms 'sensibility' and 'intuition' to mean the capacity for having experiences in general (including for example the exercise of inner sense) and does not restrict it to sense perception (outer sense) in particular. (3) implied that if X is not spontaneous then X belongs to sensibility. Elsewhere he calls sensibility a sort of receptivity.

I take it that if X is a mental event (say either a thought or an experience) then it follows that X is either spontaneous or receptive but not both: that is, as predicates characterising mental states 'spontaneous' and 'receptive' are mutually exclusive and collectively 3exhaustive. There is no paradox involved in the description of intellectual intuition because it is only spontaneous, a mind which exercise both aperception and inner sense is by virtue of the former spontaneous, but by virtue of the latter receptive, so it stands in need of ex planation how one and the same mind may be both spontaneous (active) and receptive (passive).

(4) is a description of the respect in which the mind is active with regard to itself in self-consciousness. There is unclarity in 'coming to consciousness of oneself' in (4). This expression could be just another characterisation of empirical apperception, or alternatively it might include within its scope both empirical apperception and inner sense.

Suppose we take the first view. Then Kant is saying that empirical apperception, by its very exercise, affects its object: the manifold of perception as object of inner awareness. Only if that happens, on this reading, is it possible for the mind to have 'an intuition of itself'.

The trouble with this is that it makes empirical apperception a condition for inner sense. Now, although in the case of persons our empirical self-consciousness requires both inner sense and apperception it does not follow from the fact that inner sense exists that empirical apperception exists. There could presumably be (non-rational) beings who possessed inner sense but not empirical apperception.

So, if we take the second reading and allow 'consciousness of oneself' to include both apperception and inner sense, then we can take Kant to be saying that if a mind intuits itself - has an intuition of one of its own mental states - then that mind affects itself. I take it that if A affects B then A (partly) changes what B is, or A alters B, or A partly causes B to be what it is. So, here if the mind intuits itself then the mind alters itself, changes what it itself is, or partially causes itself to be what it is.

Is there anything paradox ical about that ? I think not. There is a different sort of causal paradox which involves this reflective notion: 'X causes itself to be'. It is perhaps logically incoherent to suppose that something might bring itself into existence, because that would seem to require a time when something both did and did not exist. But no such claim is made here. If the mind intuits itself it allegedly alters its nature in some respect, and this does not require it causing itself to be only to be what it is.

Now, it is not obvious that perceiving something alters what it is (except in the trivial sense that it makes it be perceived) but let us accept Kant's point here without argument. Then we have the theory that in inner sense the mind perceives itself and thereby partially alters its own nature. He says 'only in this way can it give rise to an intuition of itself', meaning that it is a necessary condition of the mind's having any intuition of itself that it affect itself.

Now, there is no paradox about this if we assume that the respect in which the mind is active is not exactly the same respect as the respect in which it is passive; and we are not committed by the theory so far to the inconsistent view that these respects are in fact numerically identical. But it is here in Kant's view that the paradox largely lies, or this premise (4) on which (c) mainly depends.

(5) is Kant's explication of the manner in which the mind alters itself by the exercise of inner sense. 'Intuition' is unfortunately ambiguous here. Although an intuition is an experience, if it is true that in inner sense a mind intuits itself then in any particular act of inner sense there are two possible sorts of candidate for the title 'intuition'.

If I intuit myself I have experience of myself. But if I intuit one of my experiences, as I might in inner sense, then I intuit one of my intuitions; because they are what my experiences are.

So, in inner sense there are putatively two sorts of intuition: the intuiting intuitions and the intuited intuitions. Kant is deeply committed to this duplicationist model by making inner sense rather like (a non-spatial) outer sense.

(5)'s 'the form of this intuition' is time, whether 'intuition' is taken in the first or the second sense. He now says that this form of intuition-time-is what in fact alters the mind in inner sense. It is in some way the temporality of the act or content of inner sense that alters partially what the mind is. There are two possibilities here, one strong and one weak.

The weak one is that in inner sense the mind alters the temporal ordering of its own mental states in this sense: inner sensing makes their temporal relations to one another different from what they would have been had there not occurred that exercise of inner sense. Or, perhaps, it makes their temporal ordering different form what it was before the exercise of inner sense and different from what it will be after the exercise of inner sense. That is the weak reading, which is consistent with his view that inner intuitions alter what the mind is: It alters its nature by altering the temporal ordering of its contents.

The strong interpretation is that the exercise of inner sense is what makes the mind temporal tout court. This is consistent with time being only the form of inner sense. Then we read 'mode in which the manifold is together in the mind' as 'temporal mode'. In other words, the mind appears to itself to be temporal, or at least, it is Kant's view that the mind as it really is in itself is not temporal. Temporal predicates cannot be coherently used to characterise it as it is in itself so far as we know.

This strong reading is congruent with his 'exists antecedently in the mind' so long as 'the mind' here means 'the mind which possesses the faculty of inner sense', then this is just an elliptical reiteration of the doctrine that time is the form of inner sense, transcendentally ideal, and subjective.

The weak and strong interpretations of (5) are in fact mutually consistent because it is logically possible that a mind should render itself temporal by the exercise of inner sense, or be temporal just in virtue of the exercise of inner sense, and it might in addition be true of that mind that the exercise of inner sense alters or in some way determines what the temporal ordering of its contents is.

If we say: If the mind makes itself temporal by its intuition of itself then there is no room for its temporal ordering during any inner sensing to be different from any temporal ordering before or after inner sense because ex hypothesi there is no temporal ordering outside inner sensing, then a quasi-Kantian reply could be that transcendentally that is true, but empirically that is false: It is at least a condition of the experience of a self-conscious rational being that it assume its mental states are temporally ordered whether it is introspectively aware of them or not even if, noumenally speaking, this claim is false.

If that is right, then the strong and weak interpretations are mutually consistent just so long as transcendental idealism and empirical realism are mutually consistent.

Either interpretation or both allows (6) to make sense. In inner sense the mind affects itself and so presents itself to itself not as it is in itself but as it appears to itself. This is either because it makes itself temporal and so makes itself appear to itself temporal when in itself it is not temporal, or it is because it temporally re-orders itself and so makes itself appear to itself to possess a different temporal order than it would have if it did not appear to itself, or both.

The other half of (6) reiterates Kant's view that there is nothing paradox ical about an intuitive intellect. Presumably, if the intuitions of the intuitive intellect are spontaneous then the intuitive intellect does not affect itself: It always freely thinks itself just as it freely experiences. This is what he means by its being 'immediately self-active'.

I have tried to dissipate the paradox by remarking on Kant's premises, but if we accept Kant's view that there is a problem about how a mind may be both active and passive in an introspective act, then (c) follows, although it is not at all clear that 'every theory is faced with the difficulty'.

In a sense the paradox is a pseudo-problem because we just have to stipulate that the mind is not active in the respect in which it is passive, nor vice versa. Indeed we must adopt this view if we accept Kant's view that there is both spontaneity and receptivity in self-consciousness.

There are two other versions of the paradox of inner sense, which Kant presents over CPR 165-168, B152-B157. The first is at CPR 165-6, B152-3; where Kant speaks of

'the paradox which must have been obvious to everyone in our ex position of the form of inner sense: namely, that that this sense represents to consciousness even our own selves only as we appear to ourselves, not as we are in ourselves'

Hier ist nun der Ort, das Paradoxe, was jedermann bei der Exposition der Form des inneren Sinnes auffallen musste, verständlich zu machen: nämlich wie dieser auch so gar uns selbst, nur wie wir uns erscheinen, nicht wie wir an uns selbst sind, dem Bewusstsein darstelle.

and the second is at CPR 167, B155 where he asks:

'How the 'I' that thinks can be distinct from the 'I' that intuits itself, and yet, as being the same subject, can be identical with the latter'

Wie aber das Ich, der ich denke, von dem Ich, das sich selbst anschauet, unterschieden und doch mit diesem letzeren als dasselbe Subjekt einerlei sei.

The first is putatively a paradox about appearance and reality, and the second a paradox about identity. I will treat them in that order.

Appearance and Reality in Inner Sense.

Kant makes the appearance and reality paradox depend upon the activity - passivity paradox . This is clear from his remarks at CPR 166, B153, where he says

'We (re)present ourselves to ourselves only as we appear to ourselves not as we are in ourselves. For (my italics) we intuit ourselves only as we are inwardly affected (Kant's italics), and this would seem to be contradictory, since we should then have to be in a passive relation (of active affection) to ourselves.'

Nämlich wie dieser auch so gar uns selbst, nur wie wir uns erscheinen, nicht wie wir an uns selbst sind, dem Bewusstsein darstelle, weil wir nämlich uns nur aschauen wie wir innerlich affiziert werden, welches widersprechend zu sein scheint, indem wir uns gegen uns selbst als leidend verhalten müssten.

As we have seen, Kant is wrong in thinking that something cannot affect itself, or that the descriptions of something affecting itself have to contain contradictions.

What is interesting here is that he thinks the appearance-paradox is generated by the active-passive paradox . He does not make explicit the connection here, but a plausible account for him to give would be this: The mind alters itself in self-consciousness, so that it appears to itself other than how it would be if it were not conscious of itself. In one sense this claim is just analytic, because a self-conscious mind can only appear to itself as a mind conscious of itself. This interpretation is

quite consistent with the one given above: that through inner sense the mind appears to itself to be temporal, but as it really is in itself it is non-temporal. These Kantian doctrines provide the necessary step from 'the mind alters itself in self consciousness' to 'the mind is conscious of itself only as it appears to itself in self consciousness, not as it really is in itself'.

I shall now examine Kant's method of escaping what he takes to be the paradox of appearance and reality in inner sense. The argument designed to dissipate the paradox involves distinguishing inner sense and empirical apperception, and ex plaining the relationship between the two:

(1) '(Apperception), as the source of all combination, applies to the manifold of intuition in general, and in the guise of the categories, prior to all sensible intuition, to objects in general.'

[Apperzeption geht] als Quell aller Verbindungen, auf das Mannigfaltige der Anschauungen überhaupt unter dem Namen der Kategorie, vor aller sinnlichen Anschauung auf Objekte überhaupt.

(2) 'Inner sense on the other hand, contains the mere form of intuition, but without combination of the manifold in it, and therefore so far contains no determinate intuition.'

Dagegen [geht] der innere Sinn die bloße Form der Anschauung, aber ohne Verbindung des Mannigfaltigen in derselben, mithin noch gar keine bestimmte Anschauung enthält.

(3) 'Apperception and its synthetic unity is, indeed, very far from being identical with inner sense.'

Die Apperzeption und deren synthetische Einheit ist mit dem inneren Sinne so gar nicht einerlei.

(4) '(Determinate intuition) is possible only through the consciousness of the determination of the manifold by the transcendental act of imagination (synthetic influence of the understanding upon inner sense), which I have titled 'figurative synthesis'

[Bestimmte Anschauung] welche nur durch das Bewusstsein der Bestimmung desselben durch die transzendentale Handlung der Einbildungskraft (synthetischer Einfluss des Verstandes auf den inneren Sinn) welche ich die figürliche Synthesis genannt habe, möglich ist.

(5) 'Synthesis [[...]] if the synthesis be viewed by itself alone, is nothing but the unity

of the act of which, as an act, it is conscious to itself.'

So ist seine Synthesis, wenn er für sich allein betrachtet wird, nichts anders, als die Einheit der Handlung deren er sich, als einer solchen, auch ohne Sinnlichkeit innerlich in Ansehung des Mannigfaltigen, was der Form ihrer Anschauung nach ihm gegeben werden mag, zu bestimmen vermögend ist.

(c) 'Thus the understanding under the title of a transcendental synthesis of imagination performs this act upon the passive subject, whose faculty it is.'

Es also übt, unter der Benennung einer transzendentalen Synthesis der Einbildungskraft, diejenige Handlungs aufs passive Subjekt , dessen Vermögen er ist. (CPR 166, B153-4)

The first premise characterises apperception in a way that could not be used to describe inner sense and the second premise characterises inner sense in a way that could not be used to describe apperception. This yields, by Leibniz's law; $(\forall x)$, $(\forall y)$ (x =y) $\leftrightarrow$ ($Fx \leftrightarrow Fy$), the conclusion (3) that apperception and inner sense are not identical, because by (1) and (2) we have (Fx & $-Fy$), and so $-(x = y)$.

Although the inference from (1) and (2) to (3) is formally valid we need to inspect the semantics of those sentences further, to understand the non-parax odical theory of self consciousness.

(1) is in fact a short definition of 'apperception'. We need to bear in mind the distinction Kant makes between empirical apperception and the transcendental unity of apperception to interpret (1).

Empirical apperception is commonsensical acquaintance with ones own mental states, but the transcendental unity of apperception is a formal condition for there existing ordered self-conscious experience. It is the doctrine that each of my thoughts must in principle be capable of being prefix ed by 'I think [...]'.

Now, clearly empirical apperception and the transcendental unity of apperception are logically related, because the exercise of empirical apperception consists in thinking, and if any of those thoughts is to count as an episode in a unified self-conscious mind then it must be capable of being couched in the first person singular psychological mode. So, (1) is really a definition of pure apperception, and only derivatively or secondarily a claim about empirical apperception. This is because it is the formal fact about self-consciousness, rather than any empirical fact, which is the 'source of all combination'.

The rest of (1) is about synthesis: the role of the understanding (empirical apperception in self-consciousness) in unifying intuitions into intelligible objects of experience. Kant also reminds us here that pure apperception is a condition of knowledge in general, not just of self-knowledge: We have to read apperception as a special case of understanding: the case in which we understand our own mental states.

Use of 'mere' in (2) is intended, I think, to make it clear that none of the facts about apperception expressed in (1) applies to inner sense, and in any case this is made explicity by including 'without combination of the manifold'. Inner sense provides only what is passively experienced in self-consciousness. This has to be organised and made intelligible by synthesis in order for self-knowledge to be possible.

When Kant says inner sense contains no determinate intuition he does not mean that inner sense does not provide intuition, because it does, he means these intuitions are not intelligible to the self-conscious subject without the contribution of the understanding, in this case; apperception. Indeed, there could not be unified self-conscious experience of any sort without the transcendental unity of apperception.

Once (1 and (2) are interpreted, (3) can be seen to clearly follow from them. If it is the case that inner sense generates intuitions but apperception generates no intuitions, and if it is further the case that apperception provides the synthetic unity of the manifold, but inner sense provides no synthetic unity of the manifold, then it follows that apperception and inner sense are distinct.

There is perhaps *prima facie* room for disagreement here. Suppose it is true that inner sense is exercised always and only when apperception is exercised, and suppose further we call this joint exercise of apperception and inner sense 'self-consciousness'. What criterion might then be invoked to say we are still dealing with two psychological faculties and not one - called 'self-consciousness'- which may be subsumed under two different descriptions ? As sensible it may be called 'inner sense' and as intellectual it may be called 'empirical apperception'.

The disagreement is only apparent. Kant would have little to quarrel with in the above account. It is in fact a particular reiteration of his general thesis that the intellectual and the sensory are mutually dependent for their operation, but that we should not therefore confuse the contribution of the one with the other. So, then we can read Kant's inner sense-apperception distinction as precisely separating out the espective contributions of sensibility and the intellect in self-consciousness.

Kant thinks that some psychologists have assimilated faculties rather like inner sense and apperception one to another, in order to avoid what he thinks of as the active-passive and appearance-reality paradoxes:

'It is to avoid this contradiction that in systems of psychology inner sense, which we have carefully distinguished from the faculty of apperception, is commonly regarded as being identical with it.' (CPR 166, B153)

[Um diesen Widerspruch zu vermeiden] daher man auch lieber den innern Sinn mit dem Vermögen der Apperzeption (welchen wir sorgfältig unterscheiden) in den Systemen der Psychologie für einerlei auszugeben pflegt.

It is reasonably clear why assimilating inner sense to apperception or apperception to inner sense should dissipate the appearance of paradox . Kant thinks it stands in need

of ex planation how one and the same mind can be in both an active and a passive relation to itself. But this description only appears in his theory because apperception (as part of the understanding) is active (a 'spontaneity') and inner sense (as part of sensibility) is passive (a 'receptivity'). If a psychologist were to suggest either that in self-consciousness a mind is entirely passive: it, say, just intuits itself, or that in self-consciousness a mind is entirely active; it, say, just thinks itself, then clearly in neither case would a mind be postulated as both active and passive and so what Kant sees as a paradox would not arise.

Similarly, suppose a psychologist thought we either intuit ourselves just as we really are and not only as we appear to ourselves, or else he thought we think of ourselves just as we really are and not just as we believe ourselves to be. Then, clearly there would arise no appearance-reality distinction in the case of self-consciousness, so what Kant thinks of as the appearance-reality paradox could not arise either.

Now, obviously, none of these strategies is open to Kant. He refuses to assimilate inner sense to apperception, nor apperception to inner sense and he has already gone to some pains to say that humans have no intuitive intellect, so there can be no assimilation of the two faculties on that model either.

The only strategy open to him is to describe the relation between apperception and inner sense in a way that is not paradoxical. I should reiterate here that this is a pseudo problem, because there is in fact nothing paradox ical about something appearing to itself other than as it really is, nor indeed something's being in both an active and a passive relation to itself. But Kant thinks there is so we should follow the remainder of his argument.

(4) means that a mind may have experiences of itself which are intelligible to itself only if the faculty he calls 'transcendental imagination' is active in self-consciousness in thinking the intuitions of inner sense in quasi-spatial ways. 'Figurative synthesis' is our propensity to think of the non-spatial - the abstract or the mental-in terms of spatial images and metaphors. Kant thinks this faculty is employed in our understanding of our own experiences in introspection.

There is some merit in this suggestion because it is arguably the ability to think of one's mental states in spatial terms which allows them to be intelligible as memories, imaginings or experiences of, or derived from, the external world (or the world of 'outer (spatial) sense').

The relationship between figurative synthesis, the transcendental imagination, the understanding, and inner sense is as follows.

Figurative synthesis is one capacity which is exercised though the transcendental imagination: the capacity, roughly, for thinking in spatial images, and perhaps spatial predicates and nouns. So figurative synthesis could not exist unless the transcendental imagination existed.

The transcendental imagination is distinguished from the empirical imagination by being a condition for self-conscious experience. This is the force of 'transcendental' here. It is itself a part of the understanding. This is clear from the

fact that the transcendental imagination exercises synthesis (here 'figurative synthesis'), and synthesis is a function of the imagination. This is confirmed by Kant's putting in brackets after 'transcendental act of imagination' 'synthetic influence of the understanding'.

Transcendental acts of the imagination are acts of transcendental synthesis of a special kind, *viz* imaginative ones, and here Kant is concerned with a special application of them, *viz* their role in making self-consciousness possible.

It is not too misleading to think of the relations between these three faculties as part and whole. Figurative synthesis is part of the transcendental imagination and the transcendental imagination is part of the understanding.

Now, inners sense, is part of none of these. The intuitions of inner sense are subject to the activity of these three, and it is that activity on the intuitions of inner sense that gives rise to 'determinate intuition': a mind's experience of itself that is intelligible to itself.

(5) ex plains further what synthesis is. By viewing synthesis alone, Kant means considering synthesis in abstraction from its content: that which is synthesised. So, if we think of what synthesis is, without considering the intuitions that are subject to it, then we should conclude that synthesis has the following features: It is an act, it is a unity, and it is self conscious.

The claim that synthesis is an act, is consistent with Kant's view that synthesis is an activity of the understanding: something the understanding performs.

When he says it is a unity, he at least means it is a unifying act, but he may also mean that the act of unifying is itself a unity. I take it unity also at least means 'organised' here, or 'forming a coherent whole'.

Finally, when he says the unified act of synthesis is 'conscious to itself' he makes 'synthesis' itself the grammatical subject of 'conscious'. This implies that the acts of synthesis are, so to speak, transparent to themselves, at least in self-consciousness. If a mind is engaged in acts of synthesis (in making inelligible to itself intuitions of inner sense) then that mind thereby knows it is engaged in synthesis, or at least, is consciously aware of that process that Kant calls 'synthesis of intuitions'.

There is a difficulty about Kant claiming that synthesis is 'conscious to itself', because synthesis has a transcendental dimension-that it occurs is a condition for knowledge. But in general, if X is transcendental then it is impossible that X be an object of empirical knowledge. *A fortiori* it is impossible that X be an object of self consciousness either.

One way out of the difficult is to insist that it is the act of synthesis which is a self-conscious one. Then the whole experience which results from the synthesis of the intuitions of inner sense may be regarded as a self-conscious act of introspection. Then we can interpret the claim as partially a denial that there is ever any unknowing introspection: that is, if a person introspects then they know they do. The trouble with this option is that Kant has already stipulated that we should ignore or abstract from the content of synthesis, but the 'self-conscious act' reading requires our not doing so.

Finally, (c) is a characterisation of the self-conscious mind that Kant thinks is not paradox ical. It is indeed not paradox ical; no more paradox ical than the earlier characterisations he did think paradox ical. What Kant has done is stipulate a respect in which the mind is active, and a respect in which it is passive. In so far as it exercises the transcendental imagination through synthesis, a mind is active. In so far as a mind (qua possessor of the intuitions of inner sense) is subject to this synthesis, it is passive. Clearly, there would only have obtained a contradiction if we said a mind is passive in just and only the respect it is active, where 'active' and 'passive' are mutually ex cluding predicates. But even if one can only be applied where the other cannot, they can still be applied to different aspects of the same subject.

It is indeed one and the same mind that is the object of transcendental synthesis and which exercises transcendental synthesis. This is partly what makes this part of Kant's account a description of self consciousness. 'Whose faculty it is' refers back to 'subject' by 'whose' and to 'understanding under the title of transcendental synthesis of imagination' through 'faculty'.

That completes Kant's non-paradox ical construal of the understanding-intuition relations in self-consciousness; a solution to a paradox that never really existed.

Identity and Inner Sense

The paradox about the identity of the self in self- consciousness is easily stated:

'How the 'I' that thinks can be distinct from the 'I' that intuits itself [[...]] and yet as being the same subject can be identical with the latter.' (CPR 167, B155)

Wie aber das Ich, der ich denke, von dem Ich, das sich selbst anschauet, unterschieden und doch mit diesem letzteren als dasselbe Subjekt einerlei sei.

and its solution is decidable *a priori*: the I cannot differ from itself because it is logically impossible that anything not be what it is.

The most coherent way of viewing this as a paradox at all is to ask: how is it possible for a theory of the self to appear to generate a logical impossibility ? The answer is: by involving a conceptual muddle. Kant's argument that there is a paradox about the identity of the self if:

(1) 'I as intelligence and thinking subject know myself as an object that is thought.'

Ich, als Intelligenz und denkend Subjekt, erkenne mich selbst als gedachtes Objekt.

(2) 'I am given to myself (as something other or) beyond that (I) which is (given to myself) in intuition.'

So fern ich mir noch über das in der Anschauung gegeben bin.

(3) '[I] know myself, like other phenomena, only as I appear to myself, not as I am to the understanding.'

[Ich erkenne mich] nur, gleich andern Phänomenen, nicht wie ich vor dem Verstande bin, sondern wie ich mir erscheine.

(4) ‘The I that thinks can be distinct from the ‘I’ that intuits itself.’ (CPR 167, B155) [This is Kemp Smith and corresponds to nothing in the German.]

(1) is the thesis that a mind is the object of its own understanding. This means I can come to have knowledge of myself by thinking about myself, but 'thought' here also carries the implication that I partly constitute myself in synthesis. This is the force of the distinction between thinking myself and thinking about myself.

If (1) is true it follows that because if I know myself by thought, and thought is not intuition, then I know myself in some way other than just intuition. But in (2) there is a slide from; I know myself as an object of thinking, to I know myself as a self ('that I') as an object of thinking. This second formulation, which is present in (2), leaves open the putative possibility that there exist two numerically distinct I's. This duplication of the self is further suggested by Kant's inclusion of 'as something other'. (3) reverts comfortably to my knowing myself, but this knowledge is only of myself as I appear to myself, not as I appear to myself when I think about myself.

This appearance reality distinction is not the same as the phenomenon/thing- in-itself distinction. it is a distinction between how I am presented to myself in intuition, and how I am presented to myself in thinking. Kant's view is that although if I intuit myself in inner sense the understanding is active in the synthesis of intuitionS that are then given, iT still remains possible that I should think about myself without, at that time, intuiting myself in inner sense.

Now, he is saying there is a difference between how I appear to myself in (synthesised) inner sense, and how I appear to myself when I just think about myself. It is not strictly accurate to talk of oneself appearing to oneself when one must think about oneself because it is only though sensibility that I may appear to myself. That is why Kant says '[[...]] as I am to the understanding'. This 'am' should not be taken to imply that I may use the understanding to think of myself as I (noumenally) really am in myself. I can only know myself phenomenally - as I appear to myself - through the use of inner sense and understanding. The difference is between two sorts of phenomenal appearance: myself as I intuit myself, and myself as I think (represent) myself.

Now, there need not be any paradox, still less contradiction, in the idea that there may be differences between how a person appears to himself in introspection and how a person represents himself in thought. And, even if there is some incompatibility between these two modes of presentation that need no commit us to

the incoherent view that there are two modes of presentation that need no commit us to the incoherent view that there are two selves: just to the view that one and the same self may have two -sometimes conflicting- perspectives on itself, or two different means of acquiring knowledge about itself.

If the two perspectives are compete, if we believe sentences acquired through one mode that are logically inconsistent with sentences acquired though the other mode, then it follows that one of the two is false, and we have to give one of them up. That is because the mutual inconsistency of any two propositions is a sufficient condition for the false hood of one of them.

If we are clear on these points then (6) does not follow from its putative premises. There is a muddle involved in 'distinct from' in (c). This is a conceptual muddle about identity or sameness.

Kant, by (3), is suggesting the possibility of two numerically distinct I's, but all the epistemology of the self so far outlined commits us to is two different modes of access to one and the same self. Far from presupposing two numerically distinct selves, this presupposes one and the same self.

Kant fails to appreciate this distinction so thinks his paradox es cannot be resolved. As he puts it:

'These are questions that raise no greater nor less a difficulty than how I can be an object to myself at all.' (CPR 167, B155)

[Dies] hat nicht weniger Schwierigkeitn bei sich, asl wie ich mir selbst überhaupt ein Objekt […] sein könne.

Ironically, the key to their solution lies implicit in Kant's oft reiterated distinction between appearances and things in themselves:

'As regards inner sense [...] by means of it we intuit ourselves only as we are inwardly affected by ourselves; in other words [...] so far as inner intuition is concerned, we know our own subject only as appearance, not as it is in itself' (CPR 168, B146).

Dass wir dadurch uns selbst nur so anschauen, wie wir innerlich von uns selbst affiziert werden, d.i. was die innere Anschauung betrifft, unser eigenes Subjekt nur als Erscheinung, nicht aber nach dem, was es an sich selbst ist, erkennen.

There is no more need to postulate two numerically distinct selves in the case of the 'I' that is intuited and the 'I' that is thought, than there is in the case of myself as I appear to myself and myself as I am in myself. Indeed, Kant insists on the identity of the 'I' in self consciousness as a condition for there being such a thing as self-consciousness at all. In the observation on the antithesis of the second antimony he says:

'Self consciousness is of such a nature that since the subject which thinks is at the same time its own object, it cannot divide itself.'

Es bringt also nur das Selbstbewusstsein es so mit sich, dass, weil das Subjekt, welches denkt, zugleich sein eigenes Objekt ist, es sich selber nicht teilen kann. (CPR 408, A445, B463)

This is the only principle Kant needs to resolve the paradoxes of inner sense.

V. The Existence of the Soul

In Section IV of the Critique of Practical Reason: 'The Immortality of the Soul as a Postulate of Pure Practical Reason' IV. ([122] Die Unsterblichkeit der Seele, als ein Postulat der reinen praktischen Vernunft) Kant offers this argument:

(1) The realization of the *summum bonum* in the world is the necessary object of a will determinable by the moral law.
Die Bewirkung des höchsten Guts in der Welt ist das nothwendige Object eines durchs moralische Gesetz bestimmbaren Willens.

(2) But in this will the perfect accordance of the mind with the moral law is the supreme condition of the *summum bonum*.

In diesem 5 aber ist die völlige Angemessenheit der Gesinnungen zum moralischen Gesetze die oberste Bedingung des höchsten Guts.

(3) This then must be possible, as well as its object, since it is contained in the command to promote the latter.

Sie muß also eben sowohl möglich sein als ihr Object, weil sie in demselben Gebote dieses zu220 befördern enthalten ist.

(4) Now, the perfect accordance of the will with the moral law is holiness, a perfection of which no rational being of the sensible world is capable at any moment of his existence.

Die völlige Angemessenheit des Willens aber zum moralischen Gesetze ist Heiligkeit, eine Vollkommenheit, deren kein vernünftiges 10 Wesen der Sinnenwelt in keinem Zeitpunkte seines Daseins fähig ist.

(5) Since, nevertheless, it is required as practically necessary, it can only be found in a progress *in infinitum* towards that perfect accordance, and on the principles of pure practical reason

(6) it is necessary to assume such a practical progress as the real object of our will.

Da sie indessen gleichwohl als praktisch nothwendig gefordert wird, so kann sie nur in einem ins Unendliche gehenden Progressus zu jener völligen Angemessenheit angetroffen werden, und es ist nach Principien der reinen praktischen Vernunft nothwendig, eine solche praktische 15 Fortschreitung als das reale Object unseres Willens anzunehmen.

(7) Now, this endless progress is only possible on the supposition of an endless duration of the existence and personality of the same rational being (which is called the immortality of the soul).

Dieser unendliche Progressus ist aber nur unter Voraussetzung einer ins Unendliche fortdaurenden Existenz und Persönlichkeit desselben vernünftigen Wesens (welche man die Unsterblichkeit der Seele nennt) möglich.

(8) The *summum bonum*, then, practically is only possible on the supposition of the immortality of the soul;

Also ist das höchste Gut praktisch nur unter der Voraussetzung 20 der Unsterblichkeit der Seele möglich,

BOOK2|CHAPTER2 ^paragraph 35

(9) consequently this immortality, being inseparably connected with the moral law, is a postulate of pure practical reason.

mithin diese, als unzertrennlich mit dem moralischen Gesetz verbunden, ein Postulat der reinen praktischen Vernunft

Kant clarifies 'postulate of pure practical reason':

'by which I mean a theoretical proposition, not demonstrable as such, but which is an inseparable result of an unconditional a priori practical law.'

(worunter ich einen theoretischen, als solchen aber nicht erweislichen Satz verstehe, so fern er einem a priori unbedingt geltenden praktischen Gesetze unzertrennlich anhängt). 25

'This principle of the moral destination of our nature, namely, that it is only in an endless progress that we can attain perfect accordance with the moral law, is of the greatest use, not merely for the present purpose of supplementing the impotence of speculative reason, but also with respect to religion.'

Der Satz von der moralischen Bestimmung unserer Natur, nur allein in einem ins Unendliche gehenden Fortschritte zur völligen Angemessenheit mit dem Sittengesetze gelangen zu können, ist von dem größten Nutzen,221 nicht blos in Rücksicht auf die gegenwärtige Ergänzung des Unvermögens der speculativen Vernunft, sondern auch in Ansehung der Religion.

'In default of it, either the moral law is quite degraded from its holiness, being made out to be indulgent and conformable to our convenience, or else men strain their notions of their vocation and their expectation to an unattainable goal, hoping to acquire complete holiness of will, and so they lose themselves in fanatical theosophic dreams, which wholly contradict self-knowledge.'

In Ermangelung 30 desselben wird entweder das moralische Gesetz von seiner Heiligkeit gänzlich abgewürdigt, indem man es sich als nachsichtlich (indulgent) und so unserer Behaglichkeit angemessen verkünstelt, oder auch seinen Beruf und zugleich Erwartung zu einer unerreichbaren Bestimmung, nämlich einem verhofften völligen Erwerb der Heiligkeit des Willens, spannt 35 [123]und sich in schwärmende, dem Selbsterkenntniß ganz widersprechende theosophische Träume verliert,

'In both cases the unceasing effort to obey punctually and thoroughly a strict and inflexible command of reason, which yet is not ideal but real, is only hindered.

durch welches beides das unaufhörliche Streben zur pünktlichen und durchgängigen Befolgung eines strengen, unnachsichtlichen, dennoch aber nicht idealischen, sondern wahren Vernunftgebots nur verhindert wird.

For a rational but finite being, the only thing possible is an endless progress from the lower to higher degrees of moral perfection.

Einem vernünftigen, aber endlichen Wesen ist 5 nur der Progressus ins Unendliche von niederen zu den höheren Stufen der moralischen Vollkommenheit möglich.

The Infinite Being, to whom the condition of time is nothing, sees in this to us endless succession a whole of accordance with the moral law; and the holiness which his command inexorably requires, in order to be true to his justice in the share which

He assigns to each in the summum bonum, is to be found in a single intellectual intuition of the whole existence of rational beings.

Der Unendliche, dem die Zeitbedingung Nichts ist, sieht in dieser für uns endlosen Reihe das Ganze der Angemessenheit mit dem moralischen Gesetze, und die Heiligkeit, die sein Gebot unnachlaßlich fordert, um seiner Gerechtigkeit in dem Antheil, den 10 er jedem am höchsten Gute bestimmt, gemäß zu sein, ist in einer einzigen intellectuellen Anschauung des Daseins vernünftiger Wesen ganz anzutreffen.222

All that can be expected of the creature in respect of the hope of this participation would be the consciousness of his tried character, by which from the progress he has hitherto made from the worse to the morally better, and the immutability of purpose which has thus become known to him, he may hope for a further unbroken continuance of the same, however long his existence may last, even beyond this life, * and thus he may hope, not indeed here, nor in any imaginable point of his future existence, but only in the endlessness of his duration (which God alone can survey) to be perfectly adequate to his will (without indulgence or excuse, which do not harmonize with justice).

Was dem Geschöpfe allein in Ansehung der Hoffnung dieses Antheils zukommen kann, wäre das Bewußtsein seiner erprüften Gesinnung, um aus seinem bisherigen Fortschritte vom Schlechteren zum moralisch 15 Besseren und dem dadurch ihm bekannt gewordenen unwandelbaren Vorsatze eine fernere ununterbrochene Fortsetzung desselben, wie weit seine Existenz auch immer reichen mag, selbst über dieses Leben hinaus zu hoffen [12] und so zwar niemals hier, oder in irgend einem absehlichen künftigen 223 Zeitpunkte seines Daseins, sondern nur in der (Gott allein übersehbaren) 20 [124]Unendlichkeit seiner Fortdauer dem Willen desselben (ohne Nachsicht oder Erlassung, welche sich mit der Gerechtigkeit nicht zusammenreimt) völlig adäquat zu sein.

Footnote 12:

* It seems, nevertheless, impossible for a creature to have the conviction of his unwavering firmness of mind in the progress towards goodness.

[12] Die *Überzeugung* von der Unwandelbarkeit seiner Gesinnung im Fortschritte zum Guten scheint gleichwohl auch einem Geschöpfe für sich unmöglich zu sein.

On this account the Christian religion makes it come only from the same Spirit that works sanctification, that is, this firm purpose, and with it the consciousness of steadfastness in the moral progress.

Um deswillen läßt die christliche Religionslehre sie auch von demselben Geiste, der die Heiligung, d. i. diesen festen Vorsatz und mit ihm das Bewußtsein der Beharrlichkeit im moralischen Progressus, wirkt, allein abstammen.

But naturally one who is conscious that he has persevered through a long portion of his life up to the end in the progress to the better, and this genuine moral motives, may well have the comforting hope, though not the certainty, that even in an existence prolonged beyond this life he will continue in these principles; and although he is never justified here in his own eyes, nor can ever hope to be so in the increased perfection of his nature, to which he looks forward, together with an increase of duties, nevertheless in this progress which, though it is directed to a goal infinitely remote, yet is in God's sight regarded as equivalent to possession, he may have a prospect of a blessed future;

Aber auch natürlicher 25 Weise darf derjenige, der sich bewußt ist, einen langen Theil seines Lebens bis zu Ende desselben im Fortschritte zum Bessern, und zwar aus ächten moralischen Bewegungsgründen, angehalten zu haben, sich wohl die tröstende Hoffnung, wenn gleich nicht Gewißheit, machen, daß er auch in einer über dieses Leben hinaus fortgesetzten Existenz bei diesen Grundsätzen beharren werde, und wiewohl er in seinen eigenen 30 Augen hier nie gerechtfertigt ist, noch bei dem verhofften künftigen Anwachs seiner Naturvollkommenheit, mit ihr aber auch seiner Pflichten es jemals hoffen darf, dennoch in diesem Fortschritte, der, ob er zwar ein ins Unendliche hinausgerücktes Ziel betrifft, dennoch für Gott als Besitz gilt, eine Aussicht in eine *selige* Zukunft haben;

for this is the word that reason employs to designate perfect well-being independent of all contingent causes of the world, and which, like holiness, is an idea that can be contained only in an endless progress and its totality, and consequently is never fully attained by a creature.
[BOOK2|CHAPTER2 ^paragraph 40]

denn dieses ist der Ausdruck, dessen sich die Vernunft bedient, um ein von allen zufälligen 35 Ursachen der Welt unabhängiges vollständiges *Wohl* zu bezeichnen, welches eben so wie *Heiligkeit* eine Idee ist, welche nur in einem unendlichen

Progressus und dessen Totalität enthalten sein kann, mithin vom Geschöpfe niemals völlig erreicht wird.

Notes

I. *Kant's Attack on Rational Psychology* 4

p. 4 (1) *Immanuel Kant's Critique of Pure Reason* trans. Norman Kemp Smith (Macmillan, London, 1978). Page references to that edition are prefaced by 'CPR'. References to the first and second edition pages are prefaced by 'A' and 'B' respectively.

II. *Simple Souls and the Unity of Consciousness* 20

p.86 (1) Immanuel Kant's *Critique of Pure Reason* trans. Norman Kemp Smith (Macmillan, London, 1978), Immanuel Kant *Critique of Practical Reason* trans. Lewis White Beck (Indianapolis, 1956). For a historical review of Kant on immortality see Karl Ameriks *Kant's Theory of Mind: An Analysis of the Paralogisms of Pure Reason* (Ox ford, Ox ford University Press, 1982) pp. 177-182.

p. 88 (2) I cannot accept C.D. Broad's claim that in order to confirm or refute the immortality of the soul it is necessary 'to state, in terms of the ordinary common sense view of the matter, what are the main facts at the back of the proposition that Mr. Jones's soul now animates Mr. Jones's body' C.D. Broad *Kant: An Introduction* (Cambridge, Cambridge University Press 1978). How psycho-physical causal relations are possible, if they are, is logically independent of the soundness of proofs of the immortality of the soul. For example, inferring the soul's immortality from its simplicity does not depend upon taking a view about psycho-physical causation or how the soul 'animates' the body.

p. 93 (3) Irrespective of the soundness of Kant's ethical argument, if the soul is immortal and not spatio-temporal the question arises of the coherence of the supposition that there is life after death through the immortality of the soul. Kant examines this question in 'The End of All Things' (in Kant *On History* edited by Lewis White Beck,Indianapolis, Bobbs-Merrill, 1963). Kant draws a distinction between infinite duration and non-temporal existence. Although Kant thinks we may form no positive concept of the latter he speculates that this kind of immortality is in a way alluring and in a way appalling. (op. cit. p.69) Bernard Williams argues that immortality is undesirable in 'The Makropulos Case: reflections on the tedium of immortality' in his *Problems of the Self* (Cambridge, Cambridge University Press,

1976) pp. 82-100. Williams has in mind infinite duration rather than atemporal existence.

III. *The Problem of Personal Identity* 44

p. 44 (1) As Bennett says, 'The Third Paralogism remains unexplained.' (Bennett, 1974: 93) He accuses Kant of reporting 'a routine case of the more general triviality that *anything* is numerically identical with *itself* throughout the time of *its* existence.' (Bennett, 1974: 94) But Kant intends us to make the background assumption that generates the problem of personal identity by tacitly adding: '*despite change*'. Bennett only credits Kant with

$(\forall x)\ (x$ lasts over $t^1[...].t^2) \rightarrow (x\ t^1 = x\ t^2)$

but a charitable reading of Kant would ascibe to him:

$(\forall x)\ (x$ lasts over $t^1[...].t^2) \rightarrow (x\ t^1 = x\ t^2)\ \&\ (Fx\ t^1\ \&\ \text{-}Fx\ t^2)\ v\ (\text{-}Fx\ t^1\ \&\ Fx\ t^2)$

which is not trivial because it invites the question of how numerical identity might be preserved, given that qualitiative identity is broken. Kant expects us to take Leibniz's Law for granted:

$(\forall x)\ (\forall y)\ (x = y)\ (Fx \leftrightarrow Fy)$

and, as is often pointed out but rarely spelled out, the problem of personal identity is generated by a *prima facie* violation of Leibniz's Law because by substitution,

$(\exists x\ t^1)\ (\exists y t^2)\ Fx\ t^1\ \&\ \text{-}Fy t^2\ \&\ (x = y)$

we have

$(\exists x)\ (\exists y)\ Fx\ \&\ \text{-}Fy\ \&\ (x = y)$

but by Leibniz's Law:

$(\exists x)\ (\exists y)\ (Fx\ \&\ \text{-}Fy) \rightarrow\ \text{-}(x = y)$

so

$\text{-}(x = y)$

p. 58 (2) Bennett rightly points out that Kant says the conclusion of the Third Paralogism is 'analytic' (at 408) but Bennett then wonders how it can be interestingly analytic. I suggest Kant thinks '$(\forall x)$ (x lasts over t^1[...].t^2) then (x t^1 = x t^2)' is analytic but the interest lies in the problem of the numerical identity of anything over time *given its gaining and losing properties*.

p. 67 (3) Kant does not have an explanation of the asymmetry between one's own existence and that of others which makes possible the non-inferential self-knowledge some comentators construe him as drawing our attention to (at A 362). Bennett says

'In the Cartesian basis I cannot know that someone was *F* and wonder whether it was I: I recollect my past states *as mine*, and so judgments of the form "It was I who was *F*" cannot be "inferred" [[...]].' (Bennett, 1974: 95)

and Strawson says:

'It would make no sense to think or say: "*This* inner experience is occurring, but is it occurring to *me*?"' (Strawson, 1966: 165)

It is an unsolved philosophical problem why there should obtain asymmetries between the undergone and the observed and the private and the public. We would need to know much more about how inalienability is possible in order to solve this problem.

p. 76 (4) It is possible to doubt Peter Strawson's claim '[[...]] it would make no sense to think or say: "I distinctly remember that inner experience occurring, but did it ocurr to me?" (Strawson, 1966: 165). Suppose I have discovered that J. J. C. Smart (or whoever) had wired me up for a while so that the neurology necessary and sufficient for my *believing that I am in phenomenological state 'S'* intermittently obtained. (Suppose there is neurology with this property.) I know that others have been similarly wired up. Suppose further that I now distinctly remember phenomenological state 'S' occurring. It would then make sense to ask 'Did it occurr to me?'. Memory is not truth entailing here because (with certain exceptions such as 'I have at least one belief') belief is not truth entailing. The failure of the incorrigibility of the mental extends to past tense first person singular psychological ascriptions.

These considerations also cast doubt on Bennett's claim

'In remembering something, I have to remember it "from the inside" or "from the outside" – so my memory of a given event comes with the identity of the subect built right into it.' (Bennett, 1974: 97)

but do not threaten Shoemaker's position as reported by Bennett:

‘[a quasi-memory] does not necessarily involve [a] past state’s having been a state of the very same person who subsequently has the [quasi-memory]’
(Shoemaker, 1970: 271, Bennett, 1974: 98)

nor, therefore, Bennett’s

‘I have construed Kant as saying that when I make a memory-based judgment of the form “I was *F* at *t*”, I may be wrong about *who* was *F* at *t*. He is absolutely right: there is nothing incoherent in the idea that something which I take to be a memory is really only a quasi-memory.’ (Bennett, 1974: 100)

p. 76 (5) Bennett rightly observes that any putative substratum is not empirical:

‘The notion of a substratum corresponds to the idea of an impossibly absolutely continuous observation – the idea of knowing for sure that x really did last from t^1 through to t^2, and not merely that things like x were to be found at various times throught that interval. The substatum analysis, aiming to get right down to real, jumpless, across time identity, descends to a level where it loses all empirical content.’ (Bennett, 1974: 105)

but then invalidly concludes that

‘[[...]] the substratum theorist puts [the identity question] beyond the reach of any findable answer.’ (Bennett, 1974: 105)

This inference only goes through if the only solution to the problem of personal identity is an empirical answer. Despite the ingenuity of the last half century, empiricist solutions have been a failure whether invoking a memory criterion or the spatio-temporal continuity of the body. It is time to bite the bullet and realise that only a metaphysical solution will do. The soul is the unchanging. Anything less than the unchanging will not do justice to personal *identity* but, at the most optimistic, account for continuity or intermittent existence.

It could be that Bennett misconstrues the problem of personal identity as epistemological and empirical. (His methodological assumptions seem to lie somewhere between Logical Positivism and the later Wittgenstein.) But the problem of personal identity is ontological and metaphysical. It is not:

How can we know that *x* at t^1 is *x* at t^2?

still less:

What kinds of observation are necessary and sufficient for our knowing that x at t^1 is x at t^2?

It is:

What are the necessary and sufficient conditions for *x* at t^1 being *x* at t^2?

Why has Bennett make this mistake? An empiricist is likely to misconstrue problems as epistemological problems because any empiricist ontology is essentially derived from observation and so existence seems exhausted by perceptible existence. Kant himself tends to assimilate What does my identity consist in? and What is it for me to know what my identity consists in?

Bennett claims to identify '[[...]] a common mistake – namely that of confusing (a) "thing in itself" with (b) "substratum".' (Bennett, 1974: 106) and says Kant commits it: 'Kant muddled (a) with (b).' (Bennett, 1974: 107) It is implausible that Kant makes this assimilation. In the Paralogisms chapter, Kant is at pains to repudiate the doctrine that the self is a substratum: an immaterial Cartesian substance which bears the properties it gains and sheds over time. Yet, for example in the Third Antinomy, he insists that a person exists as a thing-in-itself. Kant is in fact anxious to draw sharply the very distinction which Bennett accuses him of collapsing. Whatever it means to say that a person exists as thing-in-tself it cannot mean a person exists as the soul putatively proven by Rational Psychology.

p. 79 (6) Although in his reference to 'ancient schools' Kant no doubt paradigmatically has in mind Heraclitean doctrines, he fails to distinguish between strong and weak flux doctrines. By a weak flux doctrine I mean each thing is always changing. By a strong flux doctrine I mean there are only changes, so it is ultimately wrong to speak of things that change. The *locus classicus* of this distinction is Plato's *Theaetetus* (152 d-e). See for example John McDowell's translation (Oxford, Oxford University Press, 1977). Nevertheless, Kant is only concerned with flux doctrines strong enough to be inconsistent with the existence of Cartesian substance.

IV. *The Refutation of Idealism* 93

p. 98 (1) 'determination' has badly misled the major commentators. I choose Guyer as the most recent example. Despite ingenious and heroic attempts Guyer is unable to make sense of the Refutation of Idealism. He says:

1. '[…] nothing in the published text of the translation explains *how* its conclusion is supposed to be reached. It offers no argument at all for the premise that something permanent is needed to make temporal determinations' (Guyer, 1987: 280)

Guyer's 'translation' is right here. 'Bestimmt' means 'located' in the Refutation of Idealism. We need to drop 'temporal determinations'.

2. '[…] the published text simply assumes that something permanent is required for the determination of time and proffers only the most perfunctory argument that such an object cannot be the self or anything in it.' (Guyer, 1987: 283)

Here we need to drop 'determination of time' and replace it with 'location in time'.

3. '[…] even if it is taken for granted that time-determination requires something which endures, it is not obvious why this cannot be an enduring self.' (Guyer, 1987: 283)

If we drop 'time determination' and replace it with 'location in time' it is evident why postulation of an enduring self will not solve the problem. It is the location of a (relatively) enduring self in time that stands in need of explanation.

'Why should an enduring self not be an adequate substratum for the permanence of time […]?' (Guyer, 1987: 283)

The permanent cannot be the self because the self is the explanandum so cannot feature in the explanans without gross circularity. Because the location of the self in time is to be explained, whatever explains this is what the self is located in, not the self or a part of the self.

3. 'But unless it is explained why the impermanence of representations should also imply the nonendurance of the empirical self itself – and after all, Kant's reference to "matter" can mean only that the impermanence of the states of an enduring *physical* substance do not themselves imply the nonendurance of that object – it will still not be apparent why we require knowledge of anything external to the self for subjective time-determination.' (Guyer, 1987: 285)

If we replace 'time determination' by 'location in time' only something external to me will provide the explanation.

4. '[…] if we look to the argument for the connection between permanence and time-determination on which Kant obviously means to rely – that is, of course, the first analogy of experience – we will find nothing which makes it self evident why determinate judgements about the temporal structure of self-consciousness require claims to knowledge of objects which are external to the self *either* as merely phenomenologically spatial or "in the transcendental sense", that is, numerically distinct from the self and independent of its representations for their own existence.' (Guyer, 1987: 283)

Guyer is right about the Analogies chapter. We need to drop 'time determination'.

5. Guyer misses the self/not-self distinction presupposed by having a concept of oneself as located in time:

'But it remains unclear why anything more than mere *acquaintance* with representations which in fact succeed one another in otherwise uninterpreted experience, or anything other than the mere *occurrence* of such representations, should be necessary for one to *judge that* there has been such a succession.' (Guyer, 1987: 285-6)

The italics need shifting back from 'judge that' to 'one.

6. (B 291) 'This assertion is utterly opaque.' (Guyer, 1987: 286)

7. (B 292) 'But again these claims are puzzling.' (Guyer, 1987: 287)

8. This is what Guyer thinks the Refutation of Idealism is about:

'[…] we shall consider Kant's argument that subjective successions can be made determinate only if they are regarded as caused by objects conceived to exist independently of the self and its states, which are represented as spatial precisely because that is how we can represent the independence we ascribe to them.' (Guyer, 1987: 297)

We need to drop 'made determinate'.

9. '[…]in spite of the stress Kant places on the contrast between a "thing outside me" and a "mere representation", it is not obvious what this contrast *means*. Thus exactly *what* thesis the refutation is supposed to prove is unclear.' (Guyer, 1987: 280)

'thing outside me' refers to physical object. 'mere representation' means 'presentation' or 'sensory content' (intuition means this). *How* the self-not self distinction should be drawn is an unsolved philosophical problem: Where do I end and where does the 'external world' begin? *That* there is a self-not self distinction is a necessary condition for the soundness of any refutation of idealism.

10. 'what Kant argues in 1787 is that for purposes of even subjective time-determination we must employ the intuition of space to represent objects which we conceive as existing independently of ourselves'(Guyer, 1987: 282)

Here ‘subjective time-determination’ is wrong. Kant means ‘location of myself in time’. ‘conceive’ is much too weak here. Guyer needs ‘exist’.

[NB: Move to Kant’s Attack on Rational Psychology. That there is free will is part of Rational Psychology (endorsed by Plato, Descartes, Christian philosophy, a priori metaphysics) but not repudiated by Kant (and possibly not recognized by him, as such, to be part of Rational Psychology). Is Kant’s compatibilism logically defensible?]
[NB: Freedom is *spared* in Kant’s attack on Rational Psychology. Is he entitled to spare freedom? His arguments for freedom should be the model for arguing about the soul.]
[NB: Move to Kant’s Attack on Rational Psychology when replacable.
4th Paralogism is part of Rational Psychology. Refutation of Idealism is a continuation of Kant’s attack on Rational Psychcology.]

Bibliographies

Kant

Allison, H. (1983) *Kant's Transcendental Idealism: An Interpretation and Defense.* New Haven, CN: Yale University Press.

Allison, H.E. (1996) *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy*, Cambridge: Cambridge University Press.

Ameriks, K. 2000. *Kant's Theory of Mind: An Analysis of the Paralogisms of Pure Reason,* 2nd edition.Ox ford: Ox ford University Press.

Beck, L.W. (1965) *Studies in the Philosophy of Kant*, Indianapolis, IN: Bobbs-Merrill Company.

Beck, L.W. (1969) *Early German Philosophy: Kant and his Predecessors*, Cambridge, MA: Harvard University Press.

Beck, L.W. (1978) *Essays on Kant and Hume*, New Haven and London: Yale University Press.

Beiser, F.C. (1987) *The Fate of Reason: German Philosophy from Kant to Fichte*, Cambridge, MA: Harvard University Press.

Bennett, J. (1974) *Kant's Dialectic*. Cambridge: Cambridge University Press.

Brook, A. (1993) "Kant's *A Priori* Methods for Recognizing Necessary Truths." In *Return of the A Priori,* Philip Hanson and Bruce Hunter, eds. *Canadian Journal of Philosophy,* Supplementary Volume 18, pp. 215-52.

Brook, A. (1994) *Kant and the Mind.* Cambridge and New York: Cambridge University Press.

Brook, A. (1998) "Critical Notice of L. Falkenstein, *Kant's Intuitionism: A Commentary on the Transcendental Aesthetic*." *Canadian Journal of Philosophy* 29, pp. 247-68.

Brook, A. (2001) "Kant on self-reference and self-awareness." In A. Brook and R. DeVidi, eds. 2001.

Brook, A. (2004) "Kant, cognitive science, and contemporary neo-Kantianism." In D. Zahavi, ed. *Journal of Consciousness Studies,* special number.

Caygill, H. (1995) *A Kant Dictionary*, Ox ford: Blackwell.

Chadwick, R. (ed.) (1992) *Immanuel Kant: Critical Assessments*, London: Routledge, 4 vols.

Dryer, D.P. (1966) *Kant's Solution for Verification in Metaphysics*, London: George Allen & Unwin.

Falkenstein, L. (1995) *Kant's Intuitionism: A Commentary on the Transcendental Aesthetic.* Toronto: University of Toronto Press.

Förster, E. (ed.) (1989) *Kant's Transcendental Deductions: The Three 'Critiques' and the 'Opus postumum'*, Stanford: Stanford University Press.

Gardner, S. (1999) *Kant and the Critique of Pure Reason* , London: Routledge.

Grier, M. (2001) *Kant's Doctrine of Transcendental Illusion*, Cambridge: Cambridge University Press.

Guyer, P. (1987) *Kant and the Claims of Knowledge.* Cambridge and New York: Cambridge University Press.

Guyer, P. (ed.) (1992) *The Cambridge Companion to Kant*, Cambridge: Cambridge University Press.

Henrich, D. (1976) *Identität und Objektivität*. Heidelberg: Carl Winter Universitäts-Verlag.

Henrich, D. (1994) *The Unity of Reason: Essays on Kant's Philosophy*, ed. R. Velkley, Cambridge, MA: Harvard University Press.

Höffe, O. (1994) *Immanuel Kant*, trans. M. Farrier, Albany: State University of New York Press.

Howell, R. (1992) *Kant's Transcendental Deduction: An Analysis of Main Themes in his Critical Philosophy*, Dordrecht and Boston: Kluwer.

Kemp Smith, N. (1958) *A Commentary to Kant's Critique of Pure Reason*. London: Macmillan.

Kitcher, P. (1990) *Kant's Transcendental Psychology*. New York: Ox ford University Press.

Klemme, H.F. and Kuehn, M (eds) (1999) *Immanuel Kant*, Ashgate: Dartmouth, 2 vols.

Korner, S. (1955) *Kant*. Harmondsworth: Penguin Books.

Meerbote, R. 1989. "Kant's functionalism." In: J. C. Smith, ed. *Historical Foundations of Cognitive Science*. Dordrecht, Holland: Reidel.

Paton, H.J. (1936) *Kant's Metaphysics of Experience: A Commentary on the First Half of the Kritik der reinen Vernunft*, London: George Allen & Unwin, 2 vols.

Pippin, R. (1987) "Kant on the spontaneity of mind." *Canadian Journal of Philosophy* 17, pp. 449-476.

Prauss, G. (1974) *Kant und das Problem der Dinge an sich (Kant and the Problem of the Ding an sich)* , Bonn: Bouvier.

Sassen, B. (2000) *Kant's Early Critics.* Cambridge and New York: Cambridge University Press.

Sellars, W. (1970) "… this I or he or it (the thing) which thinks … ". *Proceedings of the American Philosophical Association* 44, pp. 5-31.

Strawson, P. F. (1966) *The Bounds of Sense.* London: Methuen.

Chris W. Surprenant 'Kant's Postulate of the Immortality of the Soul' International Philosophical Quarterly March 2008: IPQ 48(1):85-98

Vaihinger, H. (1881-92) *Commentar zu Kants Kritik der reinen Vernunft (Commentary on Kant's Critique of Pure Reason)* , Stuttgart: W. Spemann and Union Deutsche Verlagsgesellschaft, 2 vols.

Van Cleve, J. (1999) *Problems from Kant*, New York: Oxford University Press.

De Vleeschauwer, H.J. (1934-7) *La Déduction Transcendentale dans l'œvre de Kant (The Transcendental Deduction in the Work of Kant)*, Antwerp, Paris, and the Hague: De Sikkel, Champion, and Martinus Nijhoff, 3 vols.

De Vleeschauwer, H.J. (1962) *The Development of Kantian Thought: The History of a Doctrine*, trans. A.R.C. Duncan, London: Thomas Nelson.

Walker, R. C. S. (1978) *Kant.* London: Routledge, Kegan Paul.

Ward, K. (1972) *The Development of Kant's Ethics*. Ox ford: Blackwell.

Wax man, W. (1991) *Kant's Model of the Mind.* New York: Ox ford University Press.

Wolff, R.P. (ed.) (1967) *Kant: A Collection of Critical Essays*, Garden City: Doubleday Anchor.

Wolff, R.P. (1963) *Kant's Theory of Mental Activity: A Commentary on the Transcendental Analytic of the Critique of Pure Reason*, Cambridge, MA: Harvard University Press.

Personal Identity

Ayer, A. J., 1936, *Language, Truth, and Logic*, London: Gollancz.

Ayers, M., 1990, *Locke*, vol. 2, London: Routledge.

Baker, L. R., 2000, *Persons and Bodies: A Constitution View*, Cambridge University Press

Behan, D., 1979, 'Locke on persons and personal identity', *Canadian Journal of Philosophy* 9: 53–75

Carter, W. R., 1989, 'How to Change Your Mind', *Canadian Journal of Philosophy* 19: 1–14

Chisholm, R., 1976, *Person and Object*, La Salle, IL: Open Court

Collins, S., 1982, *Selfless Persons: Imagery and Thought in Theravada Buddhism*, Cambridge University Press

Garrett, B., 1998, *Personal Identity and Self-Consciousness*, London: Routledge

Heller, M., 1990, *The Ontology of Physical Objects: Four-Dimensional Hunks of Matter*, Cambridge University Press

Hirsch, E., 1982, *The Concept of Identity*, Ox ford University Press

Hudson, H., 2001, *A Materialist Metaphysics of the Human Person*, Ithaca: Cornell

Hume, D., 1978, *Treatise of Human Nature*, Ox ford: Clarendon Press (original work 1739); partly reprinted in Perry 1975

Johnston, M., 1987, 'Human Beings', *Journal of Philosophy* 84: 59–83

Lewis, D., 1976, 'Survival and Identity', in *The Identities of Persons*, A. Rorty (ed.), Berkeley: California, and reprinted in his *Philosophical Papers* vol. I, Ox ford University Press, 1983

Locke, J., 1975, *An Essay Concerning Human Understanding*, ed. P. Nidditch, Ox ford: Clarendon Press (original work, 2nd ed., first published 1694); partly reprinted in Perry 1975

Lowe, E. J., 1996, *Subjects of Experience*, Cambridge University Press

Ludwig, A. M., 1997, *How Do We Know Who We Are?*, Ox ford University Press

Mackie, D., 1999, 'Personal Identity and Dead People', *Philosophical Studies* 95: 219–242

Martin, R., 1998, *Self Concern*, Cambridge University Press

Martin, R. and J. Barresi (eds.), 2003, *Personal Identity*, Ox ford: Blackwell.

McDowell, J., 1997, ‘Reductionism and the First Person’, in *Reading Parfit*, J. Dancy (ed.), Ox ford: Blackwell

Merricks, T, 1998, ‘There Are No Criteria of Identity Over Time’, *Noûs* 32: 106–124

Nagel, T. 1971, ‘Brain Bisection and the Unity of Consciousness’, *Synthèse* 22: 396–413, and reprinted in Perry 1975 and in Nagel, *Mortal Questions*, Cambridge University Press 1979

——, 1986, *The View from Nowhere*, New York: Ox ford University Press

Noonan, H., 1998, ‘Animalism Versus Lockeanism: A Current Controversy’, *Philosophical Quarterly* 48: 302–318

——, 2003, *Personal Identity*, Second edition, London: Routledge

Nozick, R, 1981, *Philosophical Ex planations*, Cambridge: Harvard University Press

Olson, E. 1997. *The Human Animal: Personal Identity Without Psychology*, New York: Ox ford University Press

——, 2002a, ‘Thinking Animals and the Reference of "I"’, *Philosophical Topics* 30: 189–208

——, 2002b, ‘What does Functionalism Tell Us about Personal Identity?’, *Noûs* 36: 682–98

——, 2003a, ‘An Argument for Animalism’, in Martin and Barresi 2003

——, 2003b, ‘Was Jekyll Hyde?’, *Philosophy and Phenomenological Research* 66: 328–48

Parfit, D, 1971, ‘Personal Identity’, *Philosophical Review* 80: 3–27, and reprinted in Perry 1975

——, 1976, ‘Lewis, Perry, and What Matters’, in *The Identities of Persons*, A. Rorty (ed.), Berkeley: University of California Press

——, 1984, *Reasons and Persons*. Ox ford: Ox ford University Press

——, 1995, ‘The Unimportance of Identity’, in *Identity*, H. Harris (ed.), Ox ford: Ox ford University Press. Reprinted in Martin and Barresi 2003.

Penelhum, T., 1970, *Survival and Disembodied Existence*, London: Routledge.

Perry, J., 1972, ‘Can the Self Divide?’ *Journal of Philosophy* 69: 463–488

----- (ed.), 1975, *Personal Identity*, Berkeley: University of California Press

Puccetti, R., 1973, ‘Brain Bisection and Personal Identity’, *British Journal for the Philosophy of Science* 24: 339–355

Quinton, A., 1962, ‘The Soul’, *Journal of Philosophy* 59: 393–403, and reprinted in Perry, ed., 1975

Rea, M., ed., 1997, *Material Constitution: A Reader*, Lanham, MD: Rowman & Littlefield

Rigterink, R., 1980, ‘Puccetti and Brain Bisection: An Attempt at Mental Division’, *Canadian Journal of Philosophy* 10: 429–452

Rovane, C., 1998, *The Bounds of Agency*, Princeton University Press

Russell, B., 1918, ‘The Philosophy of Logical Atomism’. *Monist* 28: 495–527 and 29: 32–63, 190–222, 345–380; reprinted in R. Marsh, ed., *Logic and Knowledge* (London: Allen & Unwin, 1956), and in D. Pears, ed., *The Philosophy of Logical Atomism* (La Salle, IL: Open Court, 1985) [page numbers from the latter]

Schechtman, M., 1996, *The Constitution of Selves*, Ithaca: Cornell University Press

Shoemaker, S., 1963, *Self-Knowledge and Self-Identity*, Ithaca: Cornell University Press

——, 1970, ‘Persons and Their Pasts’, *American Philosophical Quarterly* 7: 269–285

——, 1984, ‘Personal Identity: A Materialist's Account’, in Shoemaker and Swinburne, *Personal Identity*, Ox ford: Blackwell

——, 1997, ‘Self and Substance’, in *Philosophical Perspectives* 11, J. Tomberlin (ed.): 283–319

——, 1999, ‘Self, Body, and Coincidence’, *Proceedings of the Aristotelian Society*, Supplementary Volume 73: 287–306

——, 2004, ‘Functionalism and Personal Identity--A Reply’, *Noûs* 38: 525-33

Sider, T., 2001, *Four Dimensionalism*, Ox ford University Press

Snowdon, P., 1990, ‘Persons, Animals, and Ourselves’, in *The Person and the Human Mind*, C. Gill. (ed.), Ox ford: Clarendon Press

——,1996, ‘Persons and Personal Identity’, in *Essays for David Wiggins: Identity, Truth and Value*, S. Lovibond and S. G. Williams (ed.), Ox ford: Blackwell

Swinburne, R., 1984, ‘Personal Identity: The Dualist Theory’, in Shoemaker and Swinburne, *Personal Identity*, Ox ford: Blackwell

Thomson, J. J., 1997, ‘People and Their Bodies’, in *Reading Parfit*, J. Dancy (ed.), Ox ford: Blackwell

Unger, P., 1979, ‘I do not Exist’, in *Perception and Identity*, G. F. MacDonald (ed.), London: Macmillan, and reprinted in Rea 1997

——, 1990, *Identity, Consciousness, and Value*, New York: Ox ford University Press

——, 2000, ‘The Survival of the Sentient’, in *Philosophical Perspectives* 11, J. Tomberlin (ed.), Malden, MA: Blackwell

van Inwagen, P., 1980, ‘Philosophers and the Words "Human Body"’, in *Time and Cause*, P. van Inwagen (ed.), Dordrecht: Reidel, and reprinted in his *Ontology, Identity, and Modality* (Cambridge University Press, 2001)

——, 1985, ‘Plantinga on Trans-World Identity, in *Alvin Plantinga*, J. Tomberlin and P. van Inwagen (ed.), Dordrecht: Reidel, and reprinted in his *Ontology, Identity, and Modality* (Cambridge University Press, 2001)

——, 1990, *Material Beings*, Ithaca: Cornell University Press

Wiggins, D., 1967, *Identity and Spatio-Temporal Continuity*, Ox ford: Blackwell

Wiggins, D., 1980, *Sameness and Substance*, Ox ford: Blackwell

Wilkes, K, 1988, *Real People*, Ox ford: Clarendon Press

Williams, B, 1956–7, ‘Personal Identity and Individuation’, *Proceedings of the Aristotelian Society* 57, and reprinted in his *Problems of the Self* (Cambridge University Press, 1973)

——, 1970, ‘The Self and the Future’, *Philosophical Review* 59, and reprinted in his *Problems of the Self* (Cambridge University Press, 1973)

Wittgenstein, L., 1922, *Tractatus Logico-Philosophicus*, London: Routledge

Wollheim, R., 1984, *The Thread of Life*, Cambridge University Press

Zimmerman, D., 1998, ‘Criteria of Identity and the "Identity Mystics"’, *Erkenntnis* 48, 281–301

Idealism

Berkeley, G. (1710) *A Treatise Concerning the Principles of Human Knowledge*, in *The Works of George Berkeley*, vol. 2, ed. T.E. Jessop, London: Thomas Nelson & Sons Ltd, 1949.

Berkeley, G. (1713) *Three Dialogues between Hylas and Philonous*, in *The Works of George Berkeley*, vol. 2, ed. T.E. Jessop, London: Thomas Nelson & Sons Ltd, 1949.

Bradley, F.H. (1897) *Appearance and Reality*, Ox ford: Ox ford University Press, 1930. (This work represents the high point of British idealism in the nineteenth century.)

Ferrier, J.F. (1854) *The Institutes of Metaphysics*, Edinburgh.

Fichte, J.G. (1794, 1797) *The Science of Knowledge*, trans. P. Heath and J. Lachs, Cambridge: Cambridge University Press, 1982.

Findlay, J.N. (1970) *Ascent to the Absolute*, London: George Allen & Unwin.

Foster, J. (1982) *The Case for Idealism*, London: Routledge & Kegan Paul.

Green, T.H. (1883) *Prolegomena to Ethics*, Ox ford: Ox ford University Press.

Hartshorne, C. (1962) *The Logic of Perfection*, Peru, IL: Open Court.

Hegel, G.W.F. (1807) *Phenomenology of Spirit*, trans. A.V. Miller, Ox ford: Ox ford University Press, 1977.

Howison, G.H. (1901) *The Limits of Evolution, and Other Essays Illustrating the Metaphysical Theory of Personal Idealism*, New York: Macmillan.

Husserl, E. (1913, 1982) *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*, trans. F. Kersten, Dordrecht: Kluwer.

McTaggart, J.M.E. (1921, 1927) *The Nature of Existence*, Cambridge: Cambridge University Press, 2 vols.

Priest S. M. (ed.) (1987) *Hegel’s Critique of Kant*. Oxford: Oxford University Press.

Priest S. M. (1990) *Theories of the Mind*. London: Penguin Books.

Priest S. M. (2007) *The British Empiricists*. (2nd edition) London: Routledge.

Schopenhauer, A. (1818) *The World as Will and Representation*, trans. E.F.J. Payne, New York: Dover Publications, 1969, 2 vols.

Royce, J. (1885) *The Religious Aspect of Philosophy*, Gloucester, MA: Peter Smith, 1965.

Royce, J. (1919, 1964) *Lectures on Modern Idealism*, New Haven, CT and London: Yale University Press.

Schelling, F.W.J. (1797) *System of Transcendental Idealism*, trans. P. Heath, Charlottesville, VA: University Press of Virginia, 1978.

Sprigge, T.L.S. (1983) *The Vindication of Absolute Idealism*, Edinburgh: Edinburgh University Press.

Sturt, H. (ed.) (1902) *Personal Idealism: Philosophical Essays by Eight Members of the University of Ox ford*, London: Macmillan.

Whitehead, A.N. (1929) *Process and Reality*, New York: The Free Press, 1978.

[NB: Move to Kant's Attack on Rational Psychology when replacable.
4th Paralogism is part of Rational Psychology. Refutation of Idealism is a continuation of Kant's attack on Rational Psychcology.]

(2) m is false. If that is right, then the strong and weak interpretations are mutually consistent. Read idealistically, Kant is saying time is subjective in that it pertains only to the psychology of the subject.

$\Diamond(\forall x\ )\ (\forall p)\ \mathrm{K}x\ p \rightarrow \blacksquare p$
$(\forall x\ )\ (\forall p)\ \mathrm{K}x\ p(t^1) \rightarrow p(t^2)$
$(\exists \mathrm{x}\ )\ (\exists)$ - & v $\blacksquare \rightarrow \leftrightarrow$
$\Diamond(\forall x\ )\ (\forall p)\ \mathrm{K}x\ p \rightarrow \blacksquare p$
$(\forall x\ )\ (\forall p)\ \mathrm{K}x\ p(t^1) \rightarrow p(t^2)$
$(\exists \mathrm{x}\ )\ (\exists)$ - & v $\blacksquare \rightarrow \leftrightarrow$

Stephen Priest is a member of the Faculty of Philosophy in the University of Oxford. He is Senior Research Fellow of Blackfriars Hall, Oxford and a member of Wolfson College, Oxford and Hughes Hall, Cambridge. He is the author of *The British Empiricists*, *Theories of the Mind*, *Merleau-Ponty*, and *The Subject in Question*, editor of *Hegel's Critique of Kant*, *Jean-Paul Sartre: Basic Writings* and co-editor (with Antony Flew) of *A Dictionary of Philosophy*. Stephen Priest has lectured widely in the United States and Europe and his writing has been translated into Dutch, Spanish, Russian, Japanese, Korean and Macedonian.