Draft. Please do not circulate.

Kant on 'I' and the soul

Introduction. The problem.

According to Kant, rationalist doctrines of the soul rest on the illusion that it is possible to derive answers to the question: 'what am I?' from an analysis of the concept 'I' in 'I think.' Kant offers his criticism of such attempts in Chapter One of the Transcendental Dialectic, "The Paralogisms of Pure Reason."

Commentators of this chapter disagree on two central questions. First, what exactly does Kant intend to prove in the Paralogisms? Second, what does he succeed in proving? Answering these questions is made even trickier by the fact that the Paralogisms chapter is the only chapter in the Transcendental Dialectic that Kant almost completely rewrote for the second edition of the *Critique of Pure Reason*, so that one may wonder whether the answers one gives to the two questions for the first edition still hold for the second edition.

I will not consider disagreements in the current literature, although I'll be happy to discuss some of them in the Q&A. Let me instead go straight to my main claim. I think Kant does not defend any positive view of what the soul might be, or in our terms, what kind of entity we might be referring to when we use the concept 'I' in the proposition 'I think.' That is the whole point of his argument. Nothing whatsoever can be derived from the use of 'I' in 'I think' except that 'I' refers to an existing thinking being. Only a correct understanding of this negative aspect of Kant's argument allows us

to grasp the novelty of his analysis of the concept 'I', which is a primary source of its continuing interest to contemporary readers.

Within the limits of this paper I will focus only on the first Paralogism, the Paralogism of substantiality – purporting to derive from the thought 'I think' the claim that I am a thinking substance. I will proceed as follows.

1) Kant's original analysis of 'I think' is provided in the Transcendental Deduction of the Categories. I will first offer a few remarks on 'I think' in that chapter.

2) I will present Kant's argument in the first paralogism in the A edition of the *Critique of Pure Reason*, and I will try to answer the two questions listed above: what is he trying to prove, and does he succeed in proving it?

3) I will present Kant's argument in the second edition and try to answer the same two questions. Hopefully this will help answer the question: is the target in both editions the same. Is the success or failure of the argument, the same?

4) I will try to defend my claim, that only by grasping the strictly negative import of Kant's argument can we grasp the revolutionary nature of his analysis of 'I', and its import for contemporary discussions of the first person pronoun and its relation to thinking, introspection and agency.

1- **''I think' according to the Transcendental Deduction**

Kant's account of the meaning and role of the proposition 'I think' occurs in the Transcendental Deduction of the Categories, the central argument of the *Critique of Pure Reason*. Kant's goal in the transcendental deduction is to provide a justification of the claim that some fundamental concepts, or categories, are known to be true of all and only possible objects of experience, namely all and only objects of perceptual knowledge.

Especially prominent among those concepts are the concepts of the relation between a substance and its essential and accidental properties; and the concept of causal connection. Kant's method to justify his claim is to argue that objects of perceptual experience are available as the kinds of objects they are only if the perceptual contents derived from sensory inputs, or what he calls sensory "manifolds" are bound together, and if this binding occurs according to some fundamental binding structures, or "forms of synthesis" of sensory manifolds. Only if they are so bound, can sensory manifolds be associated according to empirical rules, compared, and eventually reflected under concepts, themselves bound in judgments.

Thus Kant maintains, in both the first and the second edition of the *Critique of Pure Reason*, that three cognitive components are necessary for any object at all to be presented to us in perceptual experience: a manifold of sensory perception, an ongoing activity of synthesis of that manifold; and an ongoing activity of unifying the synthesis and reflecting the resulting bundles under concepts. For instance, to represent the perceptual content resulting from a particular batch of sensory input as *a tree*, I need to be presented with a manifold of colors, odors, texture, and so on. I need to bind that manifold together and compare the bundle thus obtained with past instances of similar bundles; this allows me to come up with the recognitional concept: 'tree,' which I can now bind in judgment with other concepts for instance: 'all trees are plants,' 'some trees are large,' 'some trees have persistent foliage,' and so on. I can thus also bind these judgments according to inferential patterns; "All trees are plants, plants need watering, so trees need watering, so my Christmas tree needs watering," and so on.

Now suppose I stand in front of a window and look outside. I see a shape in the distance, and I say: 'this is a magnolia.' Someone might ask: 'are you sure?' To which I reply, reinforcing my initial statement: 'Yes, I think this is a magnolia!' 'I think' here expresses the fact that if pressed, I can give my reasons for identifying the object I am perceiving as a magnolia. Those reasons are made available to me by my own act of synthesizing the perceptual content of the sensory information available to me, comparing the bundles thus formed to other similarly bundled manifolds and recognizing the marks that allow me to identify the object as - a magnolia.

This is the process Kant has in mind when he writes, in a famous passage of the second edition Transcendental Deduction:

> The **I think** must be able to accompany all my representations; for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least would be nothing to me. (B131-132).

Note that according to the explanation I just gave, prefacing "this is a tree" with "I think this is a tree" is not turning my attention from the tree to myself. Quite the contrary: it is reinforcing my statement *about the tree*. The reinforcement may have more or less assertoric force, ranging from: 'It seems to me this is a tree' to 'I believe this is a tree' or 'I am certain this is a tree,' depending on the degree of epistemic probability yielded to my statement by the reasons I can muster for it on the basis of my activity of synthesizing the information I have and the concepts it provides me with.

The statement 'I think this is a tree' thus has the feature of transparency we associate with a familiar solution to Moore's paradox made prominent by Gareth Evans and more

recently, by Richard Moran. Why do statements of the type 'p, and I don't believe p' or 'I believe that p, but p is not true' strike us as impossible, even though they do not express a formal contradiction? This is because for the believer, the assertion of her belief *state* is "transparent to" the assertion of the *content* of her belief. Asserting the content just is, for the asserter and believer, asserting *her own attitude of belief* with respect to that content, or at least, and perhaps more properly, asserting the content is *being committed to* asserting her belief in the truth of the content.

Of course Kant says something more when he says, in the text cited above, that "the I think must be able to accompany all my representations" and then goes on to develop the point by saying: "For otherwise something would be represented in me that could not be thought at all, which is as much as to say that it would not be represented at all, or at least, would be nothing to me." What he is saying is that *a necessary condition for the fact that*, a representation *present me with an object* I can recognize *as* the object of that representation is that a synthesis and conceptualization should have occurred. This process of synthesis and conceptualization is what makes it possible for the transparency phenomenon set in: having a representation that 'p' just is also having available the thought 'I think p,' *not* by turning my attention to myself and seeking separate evidence for the fact that I think p. But just *in virtue of thinking* that p. If Kant is right, his theory of the role of synthesis and concept)yielding analysis is the explanation of the "transparency" phenomenon. For it is the explanation of the fact that *for the believer*, 'I believe that p' is transparent to 'p'. For 'I believe that p' is grounded in the very activity that makes available the evidence *for the fact that p*.

This means also that there is nothing I need to know about the referent of 'I' to meaningfully use it in this context. The proposition 'I think' is nothing but the conceptual expression of the consciousness – whether clear or obscure – , on the part of the synthesizer and conceptualizer, of being engaged in the activity of synthesizing and conceptualizing that is at work in yielding her assertion, expressing her belief, that 'p.' In other words, 'I' refers to whoever or whatever is currently thinking 'I think p,' and thinking it on the sole basis of thinking 'p'.

Let me now show how these elements play out in Kant's argument in the first Paralogism of Pure Reason.

2- The First Paralogism in A

Kant states the paralogism as follows:

> That whose representation is the **absolute subject** of our judgments and therefore cannot be used as determination of another thing is **substance**.
>
> I, as a thinking being, am the **absolute subject** of all my possible judgments, and this representation of myself [*von Mir selbst*] cannot be used as predicate of other things.
>
> So I, as a thinking thing (soul), am substance. (A348. Bold letters are Kant's emphasis)

[Comment on the major premise:

The major premise is one that rationalist metaphysicians raised in the Aristotelian tradition accept as definitional of substance. Substance just is that in which essential and accidental properties inhere. In a judgment, the relation between substance and accident

is properly represented by the relation between the subject-concept and the predicate-concept. Of course logical operations of conversion can always reverse the position of subject and predicate and place the concept of the property in the position of subject, the concept of the substance in the position of predicate, as in the example cited by Kant in the *Critique of Pure Reason*: "All bodies are divisible" expresses a relation between substance (bodies) and what we would call a dispositional property (divisible). But a simple rule of conversion may transform the judgment into: "Some divisible things are bodies" (cf. B128-29). Nevertheless, the relation in judgment that properly *maps* the ontological relation is that in which substance is represented by the subject concept, property (accidental or otherwise) is represented by the *predicate* concept in the proposition, since subject is that *of which* something is asserted and predicate is that which is asserted of something. So the proper representation of the ontological relation between "body" and "divisible" is that in which "body" is in the position of subject, "divisible" in the position of predicate.[1]

Because of this fundamental mapping of the logical concept of subject in the proposition on to that of substance, substance is also called subject. It is then a metaphysical subject, or substrate of properties and changing states. This creates an ambiguity in the use of the term "subject" that does not facilitate the understanding of Kant's argument, as we shall see.

Now something may be substance only in a relative sense: the piece of wax, in Descartes' second Meditation, counts as substance with respect to its changing shapes, odors, colors, sounds. But wax is really only a state of some more fundamental substrate

[1] Give reference to Aristote for subject/predicate, substance/accident./ And give just the definition of substance and accident in Baumgarten.

or metaphysical subject: a composite of fundamental particles. Any such composite is substance only in a relative sense, or as Baumgarten says; it is "substantivized phenomenon" (*phaenomenon substantiatum*]. A substance that is not itself the temporary state of a more fundamental entity or entities would be a substance "in the absolute sense." And in fact, only a substance "in the absolute sense" is substance properly speaking. This is what our major premise is saying. Kant, on this point, is in agreement with the whole metaphysical tradition. What the justification might be, to represent something as a substance in this sense, is a different matter, which we can set aside for now.

What we cannot set aside is the ambiguity in Kant's use of the term "of" in "that whose representation is the absolute subject *of* our judgments."[2] In my explanation I have assumed that by "subject of our judgment" Kant meant "subject *in* our judgment," namely subject-concept in the propositional content of our act of judging. This assumption is justified by the fact that he is talking about *the representation of* something, not that something itself. And he is talking about the *use* we make of that representation: assigning to it the place of subject not predicate in the propositional content of our judgment. But of course "subject of our judgment" might also mean the *metaphysical* subject, the entity that bears the act of judging as one of its properties or states or actions. It seems clear that Kant is not using the expression "subject of" in this second sense in the major premise, since he talks of the *representation* being the subject

[2] Note, however, that in German the word "of" is not present. The word 'of' translates the mere genitive form: 'Subject of our judgments' is in the original German 'Subjekt unserer Urteile.' But the ambiguity is the same.

of our judgment. But the ambiguity remains, and it is reinforced when we look at the minor premise.

Comment on the minor premise:

In the first half of the minor premise, it seems obvious that "subject of" should be interpreted in the second, ontological sense. "I, as thinking, am the absolute subject of all my judgments." How could we possibly understand this sentence otherwise than as: "I am the metaphysical subject or substrate of which the act of judging is an actualized power"? But the second half of the sentence talks of *representation*: "And this representation of myself cannot be used as predicate of other things." This takes us back to the first, logical interpretation: the *concept* 'I' can be used in all my possible judgments only as subject not as predicate of other things.

But why is that? Surely I am not the absolute subject *in* all my possible judgments? Tobias Rosefeldt suggests to interpret the sentence as meaning: "I am the absolute subject in all the possible judgments *in which 'I' is present.*" Or: " 'I' cannot be present in judgment otherwise than as subject, *not* as predicate of something else." This would certainly be true. But it does not account for the specification " I, *as thinking*." Kant has something more specific in mind. He has to mean: "In the judgment 'I think', 'I' cannot be present otherwise than as subject, not as predicate of something else."

And now it will help to return to the role of 'I think' in the argument of the Transcendental Deduction of the Categories. I have suggested above that according to that argument, thinking any proposition 'p' to be true of the world is at the same time being in a position to think 'I think p,' a conceptual representation of an activity I take to me my own, an activity of synthesis of "manifolds" presented to the senses, *and* analysis

of those synthesized manifolds into concepts, bound in judgments. But this means that any act of judging, yielding a judgment or proposition as the content of the act of judging, puts me in a position to use the concept 'I' as "the absolute subject" in the proposition 'I think' *and thereby* to think of myself as the absolute subject, in the metaphysical sense, of my judgment. In other words, the ambiguity in the use of the expression "subject of," [or Subjekt plus Genitiv, in German] that was present in the major premise is fully in play in the minor premise, but also, if we refer back to the Deduction, explained. The proposition 'I think' expresses my consciousness of being the agent of the act of thinking that yields all my possible judgments. Now in the proposition 'I think' the concept 'I' can occupy only the position of subject not predicate and in that sense is absolute (logical) subject of the proposition 'I think' that expresses the consciousness of an act of synthesis and analysis I ascribe to myself. But having and using that concept, I am thereby prone to think of myself as the absolute metaphysical subject of all my acts of judging. In saying "I am the absolute subject of all my judgments and that representation of myself etc…" Kant expresses both a view that is endorsed by the rationalist metaphysician as metaphysically true, and by himself as true at least in its second half. Now the second half suffices for the inference to go through:

> That whose representation is the **absolute subject** of our judgments and therefore cannot be used as determination of another thing is **substance**. **[def. of substance]**
>
> The representation of myself as a thinking being can be used only as the **absolute subject** of all my possible judgments [the absolute subject of the proposition 'I

> think' that must be able to accompany all my possible judgments], and therefore cannot be used as predicate of other things.
>
> So I, as a thinking thing (soul), am substance. (A348. Bold letters are Kant's emphasis)

Now this seems like a perfectly valid syllogism. Kant defines a paralogism as a "sophisma figurae dictionis" or fallacy due to form.[3] The fallacy due to form is generally a fallacy due to equivacation on the middle term, and this is indeed how Kant will describe it explicitly in the B edition. But where is the equivocation on the middle term here?

Here it is worth looking at Kant's comments on the first Paralogism (A348-49).

He starts by reminding us of the lesson gained from the Transcendental Analytic: pure categories of the understanding have no objective significance, namely no relation to a possible or actual object, unless a sensible intuition is subsumed under them. Absent such an intuition, they remain just what they are insofar as they belong merely to the understanding: functions of judging without any content. In the case of the category of substance, absent a sensible intuition, and the specific rule for synthesizing intuitions proper to the category of substance, all we are left with is the logical function and the corresponding form of categorical judgment, the relation between subject and predicate.

Now this is just what happens with our use of the concept 'I' in the proposition 'I think.' In the proposition 'I think,' the position of 'I' as only subject not predicate is not justified by its referent being presented in intuition according to the rule: "distinguish the permanent object from its changing states or accidents." It is justified only by the role of

---

[3] References!

'I' as referring, *for* the thinking being, in any instance of thinking, to the thinking being, conscious of being engaged in the activity of thinking whose outcome she assesses as valid or invalid, true or false.

But now we may ask: if the concept "what can be used only as subject not as predicate of something else" is just as purely intellectual in the major and in the minor premise, again, where is the paralogism?

At the end of Paralogisms A, Kant explains the fallacy in the following way: **[DO NOT READ THE TEXT. IT IS ON THE HANDOUT]**

> If one wants to give a logical title to the paralogism in the dialectical inferences of reason in the rationalist doctrine of the soul, ***insofar as they nevertheless have correct premises***: then it counts as a *sophisma figurae dictionis*, in which **[a]** the major makes a purely transcendental use of the category with respect to its condition, but **[b]** the minor and the conclusion, with respect to the soul, which is subsumed under this category, makes of this same category an empirical use. **[c]** So for instance the concept of substance in the paralogism of substantiality[4] is a purely intellectual concept, which without conditions of sensible intuition is of only transcendental use, that is to say, of no use at all. But in the minor premise **[d]** this same concept is applied to the object of all inner experience, but **[e]** without previously securing [*feszusetzen*] and offering as a ground [*zum Grunde zu legen*] the condition of the application of the concept *in concreto*, namely the

[4] In agreement with Adickes, I have replaced "simplicity" by "substantiality."

> permanence of the object [*die Beharrichkeit desselben*], **[f]** so that an empirical but illegitimate use is made of the category." (A402-403)[5]

In **[a]** and **[c],** Kant claims that in the major premise a "purely transcendental use" and thus "no use at all" is made of the category: it spins in the void. The category he is talking about is the purely intellectual category of substance: "that whose representation in our judgments can be only subject not predicate of something else." This is the category that appears as the subject in the major premise and the predicate in the minor premise. This should remind us of what Kant said in the Chapter on the Schematism of the Pure Concepts of the Understanding (A146-47/B185/87) and in the chapter on Phenomena and Noumena (A238/46-B297-303): absent the schemata, namely the rules for synthesizing sensible intuitions (and so here, the schema of permanence) the categories (here, the pure category as the purely intellectual concept of something that can only be subject not predicate of something else) have no application to objects at all and remain merely logical functions *for* forming concepts of objects.[6]

Now in **[b],** Kant says something more surprising about the minor premise: in the minor premise he says, an "empirical use" is made of the category (here again, as it

---

[5] I have translated "unzulässig" by "improper" rather than "unreliable," the term proposed in the Guyer/Wood translation. As I shall argue in the main text, Kant's meaning is not just that the use is unreliable. It is that it is illegitimate: absent an intuition, *no claim can be made at all* that any object is justifiably asserted to fall under the category of substance. The German original in brackets indicates other points at which I have altered the Guyer/Wood translation in an effort to make Kant's point clearer.

[6] Now from this it follows irrefutably that the pure concepts of the understanding can never be of transcendental, but only of empirical use, and that the principles of pure understanding can be related to objects of the senses only in relation to the general conditions of a possible experience, but never to things in general (without taking regard of the way in which we might intuit them. (A246/B303)

appears in the minor premise, the category is the pure concept of understanding, "that whose representation can be used only as subject not predicate of something else." This is surprising because in the texts of the Transcendental Analytic just mentioned, the "empirical use" of the categories, which was said to be the only possible use, was a use in which the category was applied to an object of experience through the mediation of a schema – a rule for the synthesis of a manifold of intuition – in the case of the category of substance, that schema is the rule of synthesis in virtue of which a distinction between permanent (the substance) and change (the alteration of the succession of its states) appears. But Kant claims that our use of the concept 'I' is supported by no such rule of synthesis. This is what is explained in **[e].** When we apply the concept of substance to the entity thought under the concept 'I', the concept is applied "without previously securing the condition of the application of the concept *in concreto*, namely the permanence of the object, and grounding the application of the concept on that condition." So how is the use "empirical", if it is not grounded on the schema of the concept of substance, permanence of the object while its states change?

[d] explains in what sense the use of the concept of substance, or more precisely, of the pure intellectual concept of absolute subject, is an empirical use when applied to what is represented by the concept 'I'**:** the concept (the purely intellectual concept of "absolute subject") is then applied to "the object of all inner experience." This in turn sends us back to the introductory sections of the chapter on the Paralogisms, where Kant said: "I, as thinking, am the object of inner sense and am called 'soul.' " He added: this object is distinct from the object of outer sense, called "body." A rational doctrine of the soul is a doctrine in which "nothing empirical is mixed" (A342/B400) except "the inner

experience" that grounds the proposition 'I think'. What Kant calls here "inner experience" should be more properly called "inner perception" or "self-perception," that is to say the affection of my inner sense by my act of thinking. In the B edition of the Paralogisms he will describe it as an "indeterminate empirical intuition, that is to say, a perception" (B423). It is in virtue of being grounded on such an affection that the proposition 'I think' contains, according to Kant, the proposition 'I exist.' 'I think' is not *purely* conceptual. Otherwise it could neither entail nor analytically contain an assertion of existence. Only "affection" of the sense (the inner sense in this case) can give us access to existence. In any instance of thinking, the thinker, in virtue of being affected by and thus aware of, her own act of thinking, is also immediately aware of her own existence.[7]

But why not say, then, that this does give us a permanent object that justifies the application of the category of substance? The reason is that all we have is an affection or impression of inner sense, present with any instance of thinking, and a fortiori in any instance of thinking 'I think.' The *affection* is empirical. But it gives us access to not object: as I argued earlier on the example of thinking 'this is a tree' to 'I think this is a tree,' our attention is still directed at the proposition the 'I think' accompanies, not at I, or myself, as thinking. Moreover, for *any* impression to give us access to a permanent, whether in an absolute or even just a relative sense, we would need the presence of the affection *at some point in time*, in however many repeated instances, to be complemented *by an intuition of space*. Only the intuition of an object in space allows us to experience the continued existence of that object while its states change and while our perceptions of

[7] See also *Prolegomena*.

it continuously change. For our concept 'I' in 'I think' no such spatial experience of permanence is available.[8]

In short: being affected by one's own act of thinking is having empirical access to one's own *existence.* It is "being affected" by it. But this is not sufficient to determine what kind of existence one is thereby given access to: that of a substance or the mere state of a substance or a temporary composite of substances, only the composition of which yields the thought "I think." It does not tell us whether in each case of thinking 'I think' and thereby thinking 'I exist' what we have access to is a particular state, different in each case, of some entity that is a subject "in the absolute sense" namely a substance; or a particular state of some entity that is not a subject "in the absolute sense" and so not a substance. Insofar as 'I' is used only as the logical subject (the subject-concept) of the proposition 'I think,' the existence the affection gives access to is an existence to which it is illegitimate to even try to apply our concepts of substance and accident *at all.* This is why Kant says, in **[f]**: in the minor premise "an empirical but illegitimate [*unzulässig*] use is made of the category."

---

[8] This importance of spatial intuition for the experience of permanence is made clearer in the second edition of the *Critique*. Thus in the "General Note on the System of Principles," added in B, immediately after having explained that absent the sensible conditions stated in their respective schemata, the categories remain "mere forms of thought," Kant adds: "It is even more remarkable, however, that in order to understand the possibility of things in accordance with the categories and thus to establish the objective reality of the latter, we do not merely need intuitions, but always outer intuitions. If we take, e.g., the pure concept of relation, we find that, 1) in order to give something that persists in intuition, corresponding to the concept of substance, (and thereby to establish the objective reality of this concept), we need an intuition ion space (matter), since space alone persistently determines, while time, however, and everything that is in inner sense, constantly flows" (B291). And he goes on to make similar points for the cases of causality and community (B291-93).

But again, then: where is the paralogism? The subject-concept in the major premise *is* the predicate concept in the minor premise. True, in the major premise "no use at all" is made of the concept, whereas in the minor premise it is applied to an existing entity. But as long as we are clear that in both cases the concept of "that whose representation can be only subject not predicate of our judgment" is merely intellectual or is a mere logical function of judgment for concepts, there is no paralogism.

One solution would be to say: the paralogism occurs in the conclusion. "Substance" is applied in a sense that is not that of the major premise. [see the handout]

Another solution, that would put the sophistical nature of *the inference* clearly in view, Kant would have needed to say two things: 1) The statement made in the major premise is verifiably true of an actual or possible object [rather than being the analytic development of the purely intellectual concept of substance] only if the following condition obtains: what can be thought only as subject not predicate of something else is also *experienced* as a permanent whose states change, and *on that basis* is represented as something whose representation can only be subject not predicate of something else. 2) That condition does not obtain in thinking an entity under the concept 'I' in 'I think.' 3) So the inference is invalid: there is an equivocation on the middle term, the conclusion that "I, as thinking, am substance," does not follow from the major and the minor premise.

This is exactly how Kant reformulates the Paralogism in B.

3: The First Paralogism in B:

In the B edition, Kant does not lay out explicitly the four (fallacious) syllogisms he laid out in A, that were supposed to ground the claims rationalist metaphysicians make about the soul: that it is substance, simple, has personhood, is known to exist independently of the body. He just lays out in four short paragraphs the fallacies that were presented in A in the form of four paralogisms of pure reason. Nevertheless, at the end of his exposition of the four fallacies Kant does give a syllogistic form to the rationalist fallacy. Unsurprisingly, the paralogism on which he focuses is the paralogism of substantiality. For the first question to arise is: what kind of entity am I (substance or state of a substance), and the other three paralogisms depend on the answer to this first question. Here's how Kant now lays out the paralogism and his refutation of it:

> What cannot be thought otherwise than as subject, does not exist otherwise than as subject, and is therefore substance.
>
> Now a thinking being, considered simply as such, cannot be thought otherwise than as subject.
>
> So it also exists only as subject, namely as substance. (B410-411)

Kant then comments (I am not going to read the comments, you have it on the hand out. I just want to point out the similarities and differences between his comments in A and B)

> **[a]** In the major premise one talks of a being that can be *thought from every perspective,* and therefore also as it may be given in intuition. **[b]** In the minor premise one talks of this same being insofar as *it considers itself* [my emphasis BL] as subject, *only with respect to thought and the unity of consciousness* [my emphasis BL], but not at the same time with respect to the intuition by which it is

given as object to thought. **[c]** Thus the conclusion is drawn by a *sophisma figurae dictionis*, namely by a fallacious reasoning.* (B411)

About the major premise, comparison of A and B: Whereas in A, Kant said that in the major premise, a "merely transcendental use, namely no use at all" was made of the category, in B Kant writes that in the major premise "one talks of a being that can be *thought from every perspective,* therefore also as it may be given in intuition." The reference to intuition is made necessary by the mention of existence in the statement of the premise, which was absent from A. For it to be true that "what cannot be thought otherwise than as subject, does not *exist* otherwise than as subject, and is therefore substance," a condition for its presentation in sensibility must be satisfied: that the object be experienced as something permanent whose states change. Only then does the concept have *any use at all* – as the Schematism chapter and the chapter on Phenomena and Noumena asserted. Of course Kant's point is that the rationalist metaphysician too assumes this condition to be satisfied, without having ever bothered to provide any justification for it.

About the minor premise and the conclusion, comparison of A and B:

Here the relation between A and B is more puzzling. Whereas in A Kant emphasized, in his analysis of the minor premise, the *empirical* although *improper* or *illegitimate* use of the concept of substance as "absolute subject," in B the term "empirical" has disappeared. The emphasis is on the impossibility of deriving from the position of 'I' in 'I think' any knowledge of the entity represented by the concept and word 'I.' But in addition, and more importantly, Kant's comment emphasizes the contrast between the first person standpoint of the minor premise, and the third person

standpoint of the major premise: "In the minor premise one talks of this same being insofar as *it considers itself* [my emphasis BL] as subject, *only with respect to thought and the unity of consciousness* [my emphasis BL], but not at the same time with respect to the intuition by which it is given as object to thought." In other words, thinking of oneself as that whose representation can only be subject not predicate of something else just is the standpoint any thinker has on her own activity of thinking, and expresses in the proposition "I think" that "must be able to accompany all my representations." Kant stresses the same point when he says that in the minor premise, "what is represented is only **"**the relation to oneself, as subject [*die Beziehung auf Sich, als Subjekt*] (as the form of thought) [*als Form des Denkens*]." The "form of thought" is the form the attribution of the predicate 'think' necessarily takes: necessarily 'think' is attributed in the first person, in judgments in which, necessarily, 'I' is with respect to 'think' can only be subject not predicate of the proposition. However, this tells us nothing at all with respect to the question: is the entity thereby represented a *real* subject, namely something that, *in its existence*, is an ultimate subject of determinations rather than the determination of something else? The *existence* of that entity, the referent of 'I', is not thereby put into question. But to the question, *what kind of entity* is that referent no answer can be offered on the mere basis of the nature of the representation 'I.'[9]

[9] Here my interpretation at odds with both Horstmann's and Melnick's interpretations, according to which what Kant calls 'the I' is an activity (cf. Horstmann (1998), Melnick (2009), Chapter 1). Horstmann's and Melnick's interpretation seems to me to ignore Kant's insistence on the claim that, despite its position as subject-concept to which 'think' is attributed, the representation 'I' cannot legitimately be said to refer to anything that is determinable either as substance or as accident of a substance. Our use of the concept 'I' gives no indication at all about *what kind of entity* 'I' might represent. This includes, presumably the supposition that the I is an activity. That 'I' does represent *some entity (Wesen*) is nevertheless as certain as the fact that 'I exist' is contained in 'I think.'

The upshot is that, despite the difference in emphasis between the two versions of the first paralogism, there is nothing in Kant's reconstruction of the rationalist fallacy in B that contradicts his reconstruction of it in A. One possible reason Kant left out the words "empirical" and "inner experience" in his B explanation is that he took them to be potentially misleading: neither of them has in this context the meaning they have when they are employed in connection with knowledge *of objects* given in intuition.

To sum up: what makes the concept 'I' unique is that it represents *for any entity making 'I' the subject-concept in the proposition 'I think,'* herself, the agent of the act of thinking 'I think', a proposition that must be able to accompany any representation if that representation is to be *something to me.* I thereby of course represent myself to myself as a thing that thinks, and I am justified in doing so by the mere fact that I think and that I am, in thinking, conscious of thinking. But I am not thereby justified in representing myself (the I that thinks, the referent of 'I' in 'I think') as a substance: as an absolute subject of inherence, a thinking substance.

There is thus a striking disconnect between the way in which necessarily, each thinker thinks about herself in virtue of *the fact that this way of thinking about oneself is indispensible to thinking anything at all*, on the one hand; and what she *knows* of herself on the other hand.

4- Concluding Remarks- Kant on 'I'

Let me now return to some of the points on which recent accounts of the first person pronoun in philosophy of mind and language present striking similarities with Kant's analysis of 'I'.

1) The first point concerns the so-called "transparency condition." The transparency condition is generally discussed as a statement about the transparency of belief: believing 'p' is being committed to believing 'I believe that p.' There is a lively discussion about the question: how can such a commitment hold ground, given that the evidence for the two beliefs would have to be quite different: one about a state of the world, another about a state of my own mind? The solution is to say that beliefs about my own beliefs are "transparent" to beliefs about the world: the very evidence that grounds my beliefs about the world is sufficient to also ground my beliefs about my own beliefs. But, one might charge, this is only renaming the mystery: how is it possible, that the very evidence that grounds my beliefs about the world should ground my beliefs about my own beliefs? Now Kant's statement: "It must be possible for the 'I think' to accompany all my representations" and his justification of the statement (which we discussed above) offers both a different formulation of the point and, I suggest, resources for a solution to the puzzle. When Kant writes "The 'I think' must be able to accompany all my representations," he is not just talking about the standing state of belief we are committed to assert when we assert (in language of thought) 'p'. He is talking about the mental process that leads eventually to being in a position to assert 'p.' It is this inseparability of the mental process relevant to being in a position to assert 'p' and the assertion that 'p' that makes it the case that asserting 'p' commits us to assert 'I think 'p.'[1]

2) The second point concerns the concept 'I' itself. The lesson of Kant's criticism of the Paralogisms is that a meaningful use of the concept 'I' in the thought 'I think' and of the word 'I' in the sentence 'I think' rests on no further justification than the fact that 'I'

refers to whoever is currently thinking 'I think' – namely, for each thinker, herself. Moreover, from Kant's analysis of the role of the thought 'I think,' it follows that any use of 'I,' namely the use of 'I' in any judgment, including those he calls, in his Lectures on Logic, use of 'I' in sensu latiori, presupposes the capacity to think 'I' in 'I think.' To have available the use of 'I' just is to have the capacity to think. To a contemporary reader, this is a striking ancestor of the contemporary view that 'I' is defined by its elementary reference rule: 'In any token use of 'I', 'I' refers to whoever is currently using 'I'.'[10] But here again Kant's explanation the *origin* of the use of 'I' offers resources to understand why the use of the first person pronoun in language and thought is fundamentally connected to agency – mental agency even more fundamentally than practical agency.

3) Kant's explanation of the unavoidable metaphysical illusions generated by the first person form of the activity of thinking accessible to consciousness, leaves entirely open the answer to the question: what kind of entity is the actual bearer, or support, or substrate of conscious thinking? Now given Kant's transcendental idealism, Kant's claim is that knowledge of that entity is inaccessible to us. But Kant's argument for transcendental idealism does not rest on his analysis of the first person. It rests on the metaphysical and transcendental exposition of space and time, and the metaphysical and transcendental deduction of the categories. It is thus a mistake to attribute to Kant what has recently been called "the illusion of transcendence." The "illusion of transcendence" is supposed to be the illusion, fostered by the peculiar features of the first person pronoun 'I,' that 'I' refers to an unknown and unknowable transcendental subject. Now it is true

[10] Refer to Peacocke.

that Kant thinks that ‘I’ refers to an unknowable transcendental subject. But again, his analysis of ‘I’ in ‘I think’ would not be sufficient to foster such a conclusion. It stands or falls independently of that conclusion. What is essential to Kant’s analysis is the striking disconnect between the way we cannot but represent ourselves in the first person, in thinking ‘I think’, and what we are justified in asserting to be true of the existing entities we are, accessible from an objective or third person standpoint.

The disconnect becomes, if anything, even more striking when we proceed from the first Paralogism to the second, third, and even fourth. But this would have to be a topic for many more papers.