

**Mathilde Byskov Jakobsen**

**‘Kant’s Principle of Transcendental Apperception’**

**MPhil April 2002  
University College London**

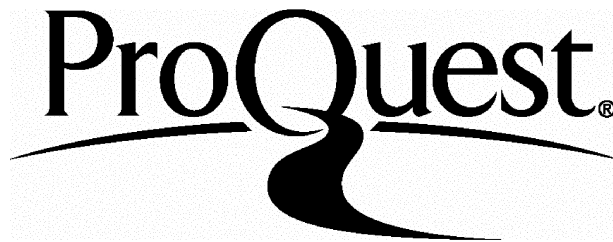
ProQuest Number: U642497

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest U642497

Published by ProQuest LLC(2015). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code.  
Microform Edition © ProQuest LLC.

ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

## **Abstract**

In this dissertation I defend the thesis that Kant's claims about the principle of transcendental apperception can form the basis of a Kantian theory of mind.

In order to defend this thesis I offer, in part one of the dissertation, an interpretation of transcendental apperception as it appears in the 'Transcendental Deduction'. I argue that transcendental apperception is primarily a principle of the unity of consciousness from which a non-substantial identity claim follows. I also examine the relation of the unity of consciousness to the 'I think', spontaneity, apperceptive self-awareness, and to the deduction of the categories.

In the second part I give an analysis of the 'Paralogisms' and of central issues that arises from this analysis, such as the noumenal ignorance thesis, and the status of transcendental psychology and its relation to cognitive science. I argue that Kant's claims about transcendental apperception do amount to a theory of mind, but only to a theory of the mind considered transcendently. I argue that this restriction means that transcendental apperception gives us information about the functions and capacities that we must necessarily represent the mind as having, when we consider it in transcendental reflection. However, it does not give us any information about how, or even whether, these functions and capacities are realised.

**Acknowledgements**

I would like to thank the AHRB for their financial support for this project. I would also like to thank Sebastian Gardner, Alan Thomas, and Joel Smith for their help and inspiration.

I draw on all the works from the bibliography, whether cited or not.

0. Introduction.....	5
1. The principle of Transcendental Apperception in the ‘Transcendental Deduction’.....	5
1.a. The ‘Transcendental Deduction’ as a Regressive Argument.....	5
1.b. Kant’s Conception of Objective Experience.....	8
1.c. Transcendental Apperception.....	10
1.d: The ‘I think’ .....	10
1.e. The Unity of Consciousness.....	14
1.f. Identity of Apperception.....	17
1.g. Transcendental Apperception and Spontaneity.....	19
1.h. Apperceptive self-Awareness .....	26
1.i. Transcendental Apperception and the Deduction of the Categories.....	28
1.j. Conclusion of Part 1 .....	33
2. A Theory of Mind? .....	35
2.a. Introduction .....	35
2.b. The ‘Paralogisms’ .....	35
2.c. The First Paralogism – The Self as Substance .....	37
2.d. The Second Paralogism – Of Simplicity.....	40
2.e. The Third Paralogism – Of Personality .....	42
2.f. Conclusion on the ‘Paralogisms’ .....	45
2.g. Noumenal vs. Transcendental Considerations of the Mind.....	45
2.h. ‘The Imaginary Subject of Transcendental Psychology’ .....	50
2.i. Transcendental Apperception and Cognitive Science .....	51
2.j. Conclusion of Part 2 .....	53
3. Conclusion .....	54
4. Bibliography: .....	57

## **0. Introduction**

My aim in this dissertation is to give an analysis and interpretation of the principle of transcendental apperception in Kant's *Critique of Pure Reason*. I will give a detailed analysis of the second edition 'Transcendental Deduction' as it is in this part of the *Critique of Pure Reason* that we find Kant's main treatment of the principle of transcendental apperception. My focus is not primarily on the deduction of the categories, but I will include an analysis of this argument in order to contextualise my analysis of the principle of transcendental apperception in the 'Transcendental Deduction'. The second aim of my dissertation is to determine the status of Kant's claims about the principle of transcendental apperception. More specifically, I will examine whether or not these claims can be interpreted as providing a basis, either partial or complete, for a Kantian theory of mind. By a theory of mind, I mean a theory that gives us information and makes claims about the metaphysical or ontological properties of the mind. This should be contrasted with a theory which merely considers the mind with respect to its relation to the question of the possibility of objective knowledge. The latter kind of theory could assign metaphysical properties to the mind, but only in relation to the question of how and whether claims to objective knowledge are justified. The former kind of theory can assign metaphysical properties to the mind apart from, or in abstraction from, the epistemological question of objective knowledge.

In order to answer this question I will analyse Kant's critique of rational psychology in the first edition 'Paralogisms', and deal with the most important issues that arise from this analysis.

## **1. The principle of Transcendental Apperception in the 'Transcendental Deduction'**

### **1.a. The 'Transcendental Deduction' as a Regressive Argument**

There are two preliminary questions that need to be settled for one to begin to give an interpretation of the principle of transcendental apperception in the 'Transcendental Deduction'. These two questions are: what is Kant's aim in the chapter and what conception of experience is his starting point?

The interpretation of the principle of transcendental apperception that I will defend in this dissertation depends on my answer to these two questions. Therefore, I will start with an account of what I think Kant's aim is in the 'Transcendental Deduction' and what conception of experience he uses in it. I think the 'Transcendental Deduction' contains a regressive argument. I mean by this that it takes objective experience for granted rather than supplying a proof for it. I also think that Kant, in the 'Transcendental Deduction' starts from a maximal rather than minimal conception of experience. It is part of the conception of experience used in the 'Transcendental Deduction' that experience is objective and contains knowledge. Walker and Ameriks both present arguments for the view that Kant presents only a regressive argument in the 'Transcendental Deduction'.<sup>1</sup> This implies that Kant presupposes objective experience and that the aim of the argument of the 'Transcendental Deduction' is to work out the conditions of this kind of experience. Ameriks, for instance, claims that the main argument of the 'Transcendental Deduction' is a regressive argument because it shows only the preconditions of empirical knowledge. That is, it shows only the necessary conditions of knowledge of objects. It does not prove that we actually have knowledge of objects (because it does not show its sufficient conditions as well). Ameriks claims that the 'Transcendental Deduction' is not alone in containing a regressive argument, but that the Aesthetic also consists of regressive arguments.<sup>2</sup>

We can contrast the regressive readings' answers to the two preliminary questions with Strawson's progressive readings' answers to the same two questions. Strawson holds, first, that Kant's aim in the 'Transcendental Deduction' is to prove that experience necessarily contains knowledge of objects.<sup>3</sup> Second, he holds that Kant starts from a minimal conception of experience.<sup>4</sup> Experience in its minimal conception, requires, according to Strawson, simply that the representations it consist of are combined or united in some way. This combination of answers means that it is very difficult to reconstruct a successful deduction in Strawson's terms. However, if this were possible the success of such

---

<sup>1</sup> Ameriks (1999) pp. 274-275, Walker (1978) p. 76.

<sup>2</sup> Ameriks (1999) pp. 274-275.

<sup>3</sup> Strawson (1995) p. 88.

<sup>4</sup> Strawson (1995) p. 25-26.

an argument would be a foundation on which Kant could build his philosophical project. In contrast, seeing the 'Transcendental Deduction' as containing only a regressive argument means that this argument is more likely to be successful. However, it does not offer a *foundation* for Kant's project. Its importance lies, instead, in the insight into the preconditions of objective experience that Kant elucidates.

The main problem with Strawson's reading is that there are very few sections of the *Critique of Pure Reason* which support it. In fact 'The Refutation of Idealism' is probably the only chapter in which Kant actually seems to be clearly and explicitly attempting the kind of analysis that Strawson identifies. In most of the other parts of the *Critique of Pure Reason*, including the 'Transcendental Deduction', Kant is much more descriptive and relies mainly on regressive arguments.<sup>5</sup>

Our understanding of the principle of transcendental apperception is very likely to differ, depending on which of these two readings of the aim of the 'Transcendental Deduction' we adopt. In many of the progressive reconstructions of the central argument of the 'Transcendental Deduction', the principle of transcendental apperception plays the role of an independent step. It is that necessary postulate for which the categories are, in turn, a necessary condition. Strawson's reconstruction is a clear example of this strategy. He argues that the ability to self-ascribe our representations is dependent on our ability to make judgements about objects. It is part of the minimal conception of experience, which Strawson ascribes to Kant, that the representations it consists of must be connected in the unity of consciousness, in order to be called an experience. Strawson further argues that the ability to become aware of my diverse representations as one and all mine is a necessary condition for claiming that they are part of a unity of consciousness and therefore for claiming that they are part of an experience. Kant has thus argued from the minimal conception of experience as unified to the maximal conception of experience as objectified.<sup>6</sup>

The role of transcendental apperception in Henrich and Guyer's reconstructions of the proof of objective experience is similar. In Henrich's reconstruction of the

---

<sup>5</sup> Gardner (1999) pp. 141-143.

<sup>6</sup> Strawson (1995) pp. 94, 98.



proof, he argues that objective experience (experience conceptualised after a priori rules) is a necessary condition for it to be possible for the subject always to be aware of its own identity in relation to all its representations.<sup>7</sup> Henrich's and Strawson's strategies are similar in that the principle of transcendental apperception is a principle which is established independently of the objectivity of experience. It can therefore function as a basis of a proof of the objectivity of experience.

Guyer's reconstruction starts with the empirical fact of self-consciousness and argues that this fact is dependent on our ability to make judgements about outer objects.<sup>8</sup> In all three reconstructions transcendental apperception is a principle that is established independently of the argument for objective experience. For this reason the proof of objective experience can be based on it. The reference to the necessity of the 'I think' at the beginning of §16 is interpreted as referring to the 'fact' of self-ascription or self-reference, and this fact stands independently of the demands of an objective experience.

We can contrast this reading of the role of transcendental apperception with the role this principle plays in a reconstruction of the central argument as a regressive argument. If we think of Kant as giving only a regressive argument, the principle of transcendental apperception can be thought of not as an independent step in a proof, but as itself a precondition for objective experience. If we grant that the 'Transcendental Deduction' should be read as giving a regressive argument only, then we should read the 'Transcendental Deduction' including the principle of transcendental apperception as describing the preconditions for objective experience. Kant's analysis should thus tell us what other things we can deduce about experience and about our selves if we consider experience to be objective.

### **1.b. Kant's Conception of Objective Experience**

According to Kant, in order for experience with objective content to be possible two faculties are necessary: sensibility and understanding supplying respectively intuitions and concepts. In the 'Aesthetic' Kant describes how the faculty of

---

<sup>7</sup> Henrich (1994) pp. 186-197.

<sup>8</sup> Guyer (1987) p. 152.

sensibility must have its own a priori forms by which it represents the given, if knowledge is to be possible, and he also argues that space and time are these a priori forms of intuitions for human beings. However, intuitions do not by themselves give us experience with objective content, rather a process of conceptualisation (the deduction of the categories aims to prove that this process of conceptualisation must happen after a priori rules) is also necessary for the representation of objects. This process of conceptualisation takes the form of judgements.<sup>9</sup> As judgements have the capacity to be either true or false, representations with objective content are normative.

This account of objective experience means that objective experience is necessarily also synthesised experience (as it requires both synthesis in intuitions and synthesis of the manifold of intuitions in concepts). Kant argues for his claim that synthesis is something that any subject of discursive experience<sup>10</sup> must be able to perform in §15, which is the starting point of the Deduction. In this paragraph Kant argues that any experience must be synthesised experience and that synthesis is an act of spontaneity. §15 ends by stating that a principle of unity is necessary to make synthesis possible:

But the concept of combination includes, besides the concept of the manifold and of its synthesis, also the concept of the unity of the manifold. Combination is representation of the synthetic unity of the manifold. The representation of this unity cannot, therefore, arise out of the combination. On the contrary, it is what, by adding itself to the representation of the manifold, first makes possible the concept of combination. (...) We must therefore look yet higher for this unity, namely in that which itself contains the ground of the unity of diverse concepts in judgement (...). (B131)

Kant supplies this principle of unity in §16 with the unity of consciousness.

It is therefore part of Kant's conception of objective experience that it is synthesised, conceptualised, and takes the form of objective judgements.

---

<sup>9</sup> According to Allison every act of conceptualisation is a judgement and every judgement is an act of conceptualisation. (Allison (1983) p. 69).

<sup>10</sup> The fact that our experience is discursive means that we can experience objects only by bringing into play the two faculties of sensibility and understanding. In order to experience objects we must be both affected by the object and bring it under concepts. In contrast for an intuitive intellect there would be no distinction between the object and the representations of the object, as thinking an object would be to create the object. (Allison (1983) p. 65, Gardner (1999) p. 66-70).

### **1.c. Transcendental Apperception**

Kant formulates the principle of transcendental apperception in many different ways in §16: He begins his account of the principle of transcendental apperception with the claim that it must be possible for the ‘I think’ to accompany all the representations, which are not nothing to me. Secondly, he relates this aspect of the principle of transcendental apperception to the unity of consciousness, which he also formulates as a principle of identity of self-consciousness. Finally he claims that transcendental apperception is an act of spontaneity. For interpretative reasons I will treat each of these aspects of the principle of transcendental apperception separately, but I will also examine the connection and relation between these different aspects of transcendental apperception, so as to show that the principle of transcendental apperception is a unified and coherent principle despite the many different aspects of this principle discussed by Kant.

### **1.d: The ‘I think’**

Kant introduces the principle of transcendental apperception in §16 as follows:

It must be possible for the ‘I think’ to accompany all my representations; for otherwise something would be represented in me which could not be thought at all, and that is equivalent to saying that the representation would be impossible, or at least would be nothing to me. (B131-132).

There are two important things to note at this stage.

First, I take it that Kant is establishing a necessary link between the ‘I think’ and *conscious* mental states, but that he leaves open the possibility that there are non-conscious mental states that are mine even though they are nothing to me. Whether or not he thinks that all my representations must have at least a *transcendentally* potential relation to the ‘I think’ is not clear from the quote, but I do not think there is anything wrong with allowing for the possibility that there are some representations that, for empirical reasons, cannot be accompanied by the ‘I think’. So, though the quote can be seen as pointing in this direction, I don’t think Kant means to claim that an actual relation to apperception is what establishes ownership of a representation, nor do I think that ownership is determined by a possible relation to apperception. There is nothing incoherent in

claiming that we have representations that can never be accompanied by the ‘I think’ though they are still ours and can still influence our behaviour.<sup>11</sup> His claim is thus restricted to those of our representations that are conscious.

Secondly, I therefore take it that Kant describes a necessary relation between the ‘I think’ and intentional consciousness, the content of which is conceptualised so that it can be thought. The fact that Kant stresses that a representation that does not have the necessary relation to the ‘I think’ *cannot be thought and is therefore* nothing to me, supports this claim as he explicitly links consciousness to the possibility for thought and so to the conceptualisation of content.

How strong the demand for reflectivity expressed in B131-132 is meant to be is a matter of controversy. The reflectivity thesis takes its strongest form in what Ameriks calls ‘the strong apperceptive thesis’ (SAT).<sup>12</sup> According to the strong apperceptive thesis the claim that it must be possible for the ‘I think’ to accompany all my representations’ means that it is necessary that we self-ascribe all of our representations and so demands that *all* my representations are represented in a higher order self-ascriptive thought. In effect this rules out the possibility of first order thoughts that are not represented in higher order thoughts. The strong apperceptive thesis is clearly too demanding. The strong apperceptive thesis conflates consciousness and self-consciousness by claiming that in order for a representation to be something to me (for it to be conscious) I must self-ascribe that representation. This suggests that it is the act of self-ascription that confers consciousness on the first order representation.<sup>13</sup> This leads the position into an infinite regress of higher order thoughts, as according to its own claim, a representations is only something to me if it is itself represented in a higher order, self-ascriptive thought.<sup>14</sup> To avoid the infinite regress, someone who believes in

---

<sup>11</sup> Keller (1998) is an example of someone who interprets Kant’s argument as resting on the assumption that only representations which I can self-ascribe are mine at all. (Keller (1998) p. 66-67).

<sup>12</sup> Ameriks (1997) p. 59.

<sup>13</sup> Though Kant often expresses the unity of consciousness via a possible unity of self-consciousness, he is capable of separating the two phenomena and does so on several occasions. (Brook (1994) pp. 58-59).

<sup>14</sup> Contemporary higher order thought theorists of consciousness, such as Rosenthal, avoid the infinite regress by stating that the higher order thought is not itself a conscious thought. However, I do not think that this option is open to Kant as it is clear that he considers the I-thought itself to be something to me.

the strong apperceptive thesis may claim that higher order thoughts do not need to be self-ascribed to be conscious. Kant explicitly says in B132 that the ‘I think’ is a representation which cannot itself be accompanied by any further representations. However, though this stops the regress, it does not make the strong apperceptive thesis any more plausible. In its new version the strong apperceptive thesis suffers from the problem of being unable to explain why higher order thoughts, whose content is an I-thought, are conscious in themselves while first order thoughts need to be self-ascribed to be conscious. The difference between the two kinds of thoughts is presumably to be found in their contents alone and it is not clear why a difference in content can explain why one kind of thought is conscious and the other is not.

In any case, there is strong textual evidence for the claim that Kant did not hold this problematic view.<sup>15</sup> First, Kant speaks of the *possibility* of self-ascription in the first claim of §16, and secondly, he says (later in §16): “As *my* representations (even if I am not conscious of them as such) they must conform to the condition under which alone they *can* stand together in one universal self-consciousness” (B132) This speaks against the strong apperceptive thesis, as it seems to mean that it is not necessary that I in fact self-ascribe all of my representations for them to be conscious.

Allison takes apperception to indicate only ‘the necessity of a possibility’, thereby rejecting the strong apperceptive thesis. According to Allison it is necessary that the subject should be able to self-ascribe all its representations for these to be something to it, but it is not necessary that the subject in fact self-ascribe all its representations.<sup>16</sup> What Allison’s suggestion in effect does is to identify transcendental apperception with a capacity – the capacity for self-ascription, rather than with actual instances of self-ascription. Kant’s claim that it must be possible for the ‘I think’ to accompany all my representations is thus

---

<sup>15</sup> Ameriks claims that Henrich ascribes something like the strong apperceptive thesis to Kant, but I don’t think this is a correct ascription. (In Ameriks (1994)).

<sup>16</sup> Allison (1983) p. 137.

interpreted as referring to the necessary capacity for self-ascription rather than to the necessity of having thoughts about one's first order thoughts.<sup>17</sup>

So, although the quotation above from B131-132 can seem to invite the charge, I do not think that Kant actually conflates consciousness and self-consciousness, though his claim can point in that direction. The fact that he refers only to a capacity for self-ascription, rather than to actual instances of self-ascription, speaks against the claim that he conflates consciousness and self-consciousness. Instead I think we should interpret the connection, which Kant establishes between the 'I think' and conscious representations as a connection between consciousness and the possibility of self-ascription as Allison suggests. I think Kant should be read as saying that the fact that we can self-ascribe a representation in an I-thought is a sign or a consequence of the fact that this representation is not nothing to us – that it is conscious.

Kant does not conflate consciousness and self-consciousness, but he also does not separate them as sharply as some of us may want to. However, even if we think that creatures can have conscious representations even though they lack the ability to self-ascribe these representations, we can still agree with Kant that to be able to self-ascribe a representation in an I-thought is *a* (sufficient) sign (even if not a necessary sign) of the fact that that representation is not nothing to us.

The connection between possible self-ascription and consciousness is thus one of a sufficient rather than necessary condition: if we are able to accompany one of our representations with the 'I think' then this means that the representation in question is not nothing to us and can be thought.

---

<sup>17</sup>Alan Thomas (1997) has suggested that Kant holds an adverbial theory of consciousness. This means that 'conscious' rather than attaching to a kind of mental state, attaches to a *mode* of being in an intentional state. The mental states is thus the same (it has the same content) whether it is conscious or not, but the mode in which we are in that intentional mental state changes when it goes from being non-conscious to being conscious. The adverbial modification means that intentional states which we are in consciously, have a potentiality for being apperceptively self-ascribed. This suggestion means that the distinction between non-conscious and conscious mental states is not a distinction in kind but a distinction in the mode in which we are in them. Thomas thus also holds that the 'I think' should be interpreted as the necessity of a possibility which attaches to the representations which are not nothing for us (which we are in consciously).

### 1.e. The Unity of Consciousness

Kant's concept of the unity of consciousness, as employed in the 'Transcendental Deduction', expresses the fact that in order for a complex thought to be conscious it is not enough that all its parts are conscious, but they must be part of a unified consciousness. This fact can be illustrated by imagining a number of people each thinking of one word of a sentence each. In this scenario no one will be conscious of the whole sentence. The point is that no matter how hard each of them thinks of her own word there will be no consciousness of the whole sentence. In order for the whole sentence to be conscious it is necessary that one person think of the entire sentence. This shows that the consciousness of each of the parts of a complex thought or representation must be unified with the consciousness of the other parts of the representation in order for the complex representation to be conscious.

I think the claim that *the* 'I think' can accompany all my conscious intentional states should be seen as *expressing* the unity of consciousness. The fact that it is possible for me to ascribe all my conscious intentional states to one 'I think' is an expression of the fact that all my representations are part of the same unity of consciousness. In the claim that *the* (self-same) 'I think' can accompany all my conscious states lies the claim that the conscious states of my experience compose a unity. As it stands this is a purely tautological claim, which expresses merely that *my* conscious states belong together in a unity, and that this unity can be expressed by the claim that it is possible for me to accompany each of these states by *the same* 'I think'. This is in coherence with the fact that Kant calls the unity of consciousness an analytic unity. (B133, B135)

However, Kant does not intend that the unity of consciousness be an empty principle based simply on the claim that all my *conscious* states belong to the unity, which is made up by *my* conscious states. The claim that I can attach the self same 'I think' to all of my conscious states needs to be grounded in something outside itself, if it is to express anything but a trivial truth. This cannot be the 'I think' itself, because simply claiming that there is a unity between my I-thoughts takes us no further than saying there is a unity between my conscious thoughts.

Hurley (1994) shows how that the unity of consciousness cannot be established simply from the fact of the subject making judgements about its own identity,<sup>18</sup> and she expands this argument into an argument, which shows that the unity of consciousness cannot be grounded in anything that has its place within the representational content. She argues that I-thoughts are ‘just more content’ and thus are no more able to establish the unity of consciousness than any other thoughts.<sup>19</sup> Instead she claims we need to ground the unity of consciousness in something outside it in order to show that the conscious states, which I call mine, in fact form such a unity.

Kant also claims that the analytic unity is based on a synthetic unity (B133, B135). The synthetic unity which Kant claims that the unity of consciousness is dependent on, is, in my view, the synthetic unity of the manifold of representations, which can be synthesised together with each other into more complex representations. The fact that different representations of which I am conscious can be synthesised into new and more complex representations shows that there is a unity of consciousness between these representations. Synthesis between representations does not constitute the unity of consciousness but it shows it. Necessarily, if representations can be synthesised to form more complex representations, then they belong to a unity of consciousness, and vice versa. The fact that two representations can be synthesised into a more complex representation shows that the two original representations were part of the same unity of consciousness. Note that as §16 deals with experience in abstraction from the human epistemic conditions such as space and time, the synthesis it examines can only be atemporal synthesis, that is, synthesis between different representations existing along side each other into more complex representations. §16 therefore also deduces only atemporal unity of consciousness, and not diachronic unity of consciousness. As any thinker in general (or any maker of judgement) must be capable of synthesis, it must have synchronic unity of consciousness, whether or not it also has diachronic unity of consciousness will depend on its special features, and cannot be decided in the Deduction as it is not a necessary condition for experience in general.

---

<sup>18</sup> even on our judgements of our identity as active beings. (Hurley (1994) pp. 138-141, see also part III pp. 144-164).

<sup>19</sup> Hurley (1994) pp. 142-144.



Allison deduces the necessity of the unity of consciousness by means of an argument, which starts from the fact of complex thoughts – leaving it an open question how these complex thoughts have come about. If we presuppose that experience involves complex thoughts, then we can conclude that transcendental apperception must contain a principle of unity. The argument for this goes as follows. If we imagine a complex thought X which consists of elements A, B, and C, then it is necessarily the case that there is a unity between the subject that thinks A, and the subject which thinks B, and the subject which thinks C, as otherwise the complex thought X would not be thought. Unity of consciousness therefore follows from the possibility of complex thoughts. Turning this argument around we can say that the possibility of synthesising a number of representations into more complex representations is a criterion, which we can use for claiming that these representations belong to a unity of consciousness. The unity of consciousness is thus deduced as a condition for synthesis and, because of the connection between synthesis and knowledge, thereby for objective experience.<sup>20</sup>

We therefore should not look for the condition for saying that two representations are part of a unity of consciousness in any identification of their content, first personal or otherwise, as Hurley shows successfully that this is not possible. However, we do not need to look for any new criteria which has to be met for two representations to belong to a unified consciousness, because this criteria is already available: If two representations can be synthesised into one new representation then they are part of a unity of consciousness. There is thus a double dependence between the unity of consciousness and synthesis, because the unity of consciousness is what makes the synthesis of two representations possible and the synthesis of two representations is the condition for claiming that they are part of a unified consciousness.

So the possibility of I-thoughts (or self-ascriptive thoughts) is not what establishes the unity of consciousness. Rather, the unity of consciousness is what makes it true to say that the self-same ‘I think’ can accompany all my conscious representations.

---

<sup>20</sup> Allison (1983) pp. 138-140.

### 1.f. Identity of Apperception

As said at the outset of this analysis, Kant employs the concept of 'identity' in addition to that of 'unity' in the context at hand. The question therefore arises of how this concept of identity is to be understood, and what its relation to the concept of unity is.

Identity of transcendental apperception is introduced in §16 as an analytic principle, which follows from and is dependent on 'a synthetic unity'. The term 'identity of apperception' refers to the view that the 'I think' which accompanies one representation is identical to the 'I think' which accompanies all other representations – *within one unified experience*. This is introduced as an analytical principle, and indeed it follows from the fact that all the representations in question are part of the same experience. As the I of apperception is nothing but the subject of experience,<sup>21</sup> the I of conscious states must be identical to each other - *within the same experience*. The identity claim of §16 is a formal identity claim made on the basis of the subjects of consciousness within one unified experience. This principle is dependent on a synthetic unity because it is only transcendently relevant to claim that the 'I think's of individual conscious states are identical if these individual representations are synthesised into new representations with a new embracing 'I think'. Ameriks compares empirical apperception with the unity of transcendental apperception in order to illustrate this point.

(E): ('I think that x', 'I think that y', 'I think that z')

(T): (I think that 'I think that x, I think that y, I think that z')

One of the differences between (E) and (T) is that the 'I's (the subjects) of (T) are presumed co-referential, while no such presumption is made in (E).<sup>22</sup> Identity of apperception thus follows as a consequence of the unity of transcendental apperception (the unity of experience/consciousness), but this is a formal identity claim, which follows directly from the unity of consciousness. It is therefore not a substantial identity claim of the kind that Kant criticises in the 'Paralogisms'. It follows from this, on the face of it, that the subject of consciousness (or apperception) is conceptually different from both the noumenal self and the

---

<sup>21</sup> As Kant says in the B-Paralogisms: "Now in inner intuition there is nothing permanent, for the 'I' is merely the consciousness of my thought." (B413).

<sup>22</sup> Ameriks (1997) p. 58.

empirical self, and an identity claim on behalf of the subject of consciousness thus make no claims either about our noumenal identity or about our identity as temporal and spatial persons. It should be noted that §16 cannot even establish a formal identity claim on behalf of the subjects of consciousness over time, as the first part of the argument in the ‘Transcendental Deduction’ abstracts from space and time, and therefore deduces only unity of consciousness in abstraction from conditions of intuition. It thus expresses only the formal identity of the subjects of consciousness in one experience.

The formal identity claim is also a necessary (but not sufficient) condition for the possibility of the subject making an empirical identity claim. The possibility of making of an actual empirical identity claim is dependent on an empirical synthesis of representations. An identity claim made on behalf of the apperceptions of a manifold presupposes a synthesis and is possible only through the consciousness of this synthesis. An actual identity claim requires that the subject becomes aware of the fact that two representations have been synthesised, because only on the basis of this awareness does it make sense to claim that the two subjects in the two distinct representations are identical. If two I-thoughts: ‘I think that F’ and ‘I think that G’ are synthesised into the new thought: ‘I think that F and G’, then this can only give rise to an identity claim, if there is an awareness of the fact that the new thought is the result of a synthesis of the two old thoughts. So in order for synthesis to lead to an actual identity claim formulated by the subject, the subject must be aware of the fact that two thoughts have been synthesised to create a new thought.<sup>23</sup>

The importance of the identity claim is, in my view, not the subject’s ability to make claims about its own identity, but rather the fact, which an identity claim establishes about the subjects of consciousness within the unity of consciousness. In my opinion, then, the concepts of unity and identity are related in the following ways: unity is primary in the sense that it is a principle we can establish from the fact of the possibility of synthesis between, and self-ascription of, the representations which belong to the unity of consciousness. Having established

---

<sup>23</sup> Note that the full conditions for an actual empirical identity claim for human beings may include an awareness of ourselves as embodied beings in space and time. However, Kant is not concerned with describing the necessary and sufficient conditions for empirical identity claims for human beings, but only with describing the necessary conditions for objective experience.

the unity of consciousness we have also, thereby, established a weak identity claim. Knowing that our conscious representations belong to a unity of consciousness, tells us that a weak identity claim holds between the subjects of consciousness within this unified experience. The fact that this is a weak identity claim means that it is established solely between the subjects of the conscious representations which belong to the unity of consciousness, and that we cannot deduce personal identity from it.

### **1.g. Transcendental Apperception and Spontaneity**

There are two connections in which Kant claims that spontaneity belongs to the conditions for objective experience in the *Critique of Pure Reason*. Firstly, he claims that the faculty of the understanding is a faculty of spontaneity, and secondly he claims that the 'I think' is an act of spontaneity and that we are aware of our own spontaneity in apperception.

In §15 of the B-Deduction, Kant says that:

But the combination (conjunctio) of a manifold in general can never come to us through the senses, and cannot, therefore, be already contained in the pure form of sensible intuition. For it is an act of spontaneity of the faculty of representation (...). Being an act of the self-activity of the subject, it cannot be executed save by the subject itself. (B130)

The synthesis of a manifold in general is therefore a spontaneous act performed by the subject.

In §16 he links spontaneity to apperception by saying:

All the manifold of intuition has, therefore, a necessary relation to the 'I think' in the same subject in which this manifold is found. But this representation is an act of spontaneity, that is, it cannot be regarded as belonging to sensibility. (B132).

In §25 in B158-159 and in a footnote to B158, Kant claims that we are conscious of our own spontaneity in apperception, for instance: "I exist as an intelligence which is conscious solely of its power of combination".

The claim that the synthesis of a manifold by the faculty of the understanding is an act of spontaneity forms a starting point for an argument for the claim that

spontaneity is a necessary component of Kant's account of the conditions of objective experience. It is inherent to transcendental idealism<sup>24</sup> that the form of the phenomenal world cannot be explained in terms of that which is given to us, but rather is a form which the world has only in so far as it is described under our epistemic conditions. The form of the phenomenal world is thus dependent on us. Kant's reference to spontaneity can therefore be understood as referring to the fact that the world as it is to us is given its form by us. This can be characterised as an act of spontaneity because there is nothing in that which is given to us that determines this form. On this account the spontaneity of synthesis is not meant to describe the voluntary powers of a noumenal agent but rather to describe the dependence of the phenomenal world on the conceptualisation of it in the faculty of the understanding.

Ameriks, for instance, gives such an interpretation of spontaneity in his book *Kant's Theory of Mind*, "the 'spontaneity' of the mind that Kant speaks of refers basically to the fact that there are structures of experience that precede objects in that they cannot be understood as mere consequences of individual experiences in a typical empiricist fashion."<sup>25</sup> As is clear from this quote, an important consequence of this interpretation of spontaneity is that spontaneity in the sense which is at issue here in the 'Transcendental Deduction', does not refer to the 'freedom' of an individual agent, but rather to the fact that the world as it is for us is dependent on the epistemic conditions of a human discursive experience. So saying that synthesis is an act of spontaneity does not indicate that it is a matter of voluntary choice for the (noumenal) agent how to do this, but just that the conceptualisation of the content is not determined by what is given in the content of experience.

Kant's claim that the 'I think' is an act of spontaneity can also be interpreted along these lines: the 'I think' is an act of spontaneity because it is not part of the first order representation, but added to it 'spontaneously'. Apperception is spontaneous because the representation, which is self-ascribed, does not already contain a representation of the subject. The self-ascription Kant refers to in the principle of transcendental apperception is thus based on a capacity to accompany

---

<sup>24</sup> 'Transcendental idealism' should be taken here as a purely epistemic doctrine, and so not in a sense that implies a commitment to transcendental psychology.

<sup>25</sup> Ameriks (2000) p. 289.

representations with the ‘I think’ even though there is nothing in the first-order representations themselves that compels this self-ascription.<sup>26</sup>

It may be noted that this interpretation of spontaneity does not in itself conflict with the project of naturalising the mind. Sellars has argued that the interpretation of spontaneity is compatible with seeing the mind as a causal system,<sup>27</sup> which has to be set in motion by external forces<sup>28</sup> and which will then give form to the given content in accordance with its own internal rules. The form it gives to the given content will be ‘spontaneous’ simply because, and in the sense that, it is undetermined by the content of what is given. Nonetheless it will still be determined by a causal system - albeit the mind’s own causal system. The kind of spontaneity ascribed to the mind by this argument is thus only relative spontaneity.<sup>29</sup>

This raises a further issue. While this interpretation does seem to make good sense of Kant’s claims about spontaneity in the ‘Transcendental Deduction’, it may be argued that the project of naturalising the mind is not part of Kant’s project, indeed, that Kant intends to refute the possibility of naturalising the mind, and that spontaneity plays a key role, or the key role, in this intended refutation. Consequently, to many commentators it has seemed problematic to interpret Kant as ascribing only relative spontaneity to the mind. Pippin, in his article ‘Kant on the Spontaneity of Mind’ from 1987, gives both a criticism of views that ascribe only relative spontaneity to Kant, and an argument for absolute spontaneity. Pippin bases his argument for absolute spontaneity on the view that Kant in the ‘Transcendental Deduction’ is concerned with establishing the necessary conditions of objective experience, which includes judgements and knowledge claims. Spontaneity is thus seen as being a necessary condition for *knowledge*.<sup>30</sup> And the ascription of spontaneity to the subject is relative to the subject’s status as a knower.<sup>31</sup>

---

<sup>26</sup> Kant’s conception of apperceptive self-ascription therefore has some similarities with contemporary judgemental theories of introspection. Shoemaker for instance argue that the basis of self-ascription cannot be an observation of the self or its states, as it is not possible to observe the self in inner sense. (Shoemaker (1996) pp. 3-25).

<sup>27</sup> Sellars (1971) p. 22-24.

<sup>28</sup> Sellars (1971) p. 23.

<sup>29</sup> Sellars (1971) p. 24.

<sup>30</sup> Pippin (1987) pp. 451,462,468,469.

<sup>31</sup> Pippin (187) p. 469.

The starting point of Pippin's argument is the claim that apperception is an act of spontaneity. Pippin sees apperception as an adverbial qualification on consciously perceiving, imagining, remembering etc. The modification of content characterised by this adverbial qualification is that such activities are inherently reflexive: "It is reflexive because, according to Kant, whenever I am conscious of anything I also 'apperceive' that it is *I* who am *thusly* conscious."<sup>32</sup> The reflexivity of apperceptive consciousness is interpreted as an adverbial qualification, because it must be seen as "an inseparable component of what it is *consciously* to perceive, imagine, remember etc."<sup>33</sup> Pippin also points out that this account of consciousness is restricted to intentional consciousness, the content of which is directed towards truth or falsity. Having established this interpretation of apperception, Pippin then argues that spontaneity is a necessary condition for an objective experience, which includes knowledge claims and judgements, and not simply mental events in a causal system. He argues, first, that for something to be a representation of an object, it is necessary that it should be a representation of the object *for the subject*. This condition cannot be fulfilled but by an active spontaneous involvement by the subject. A mental event in a causal system cannot be (under another consideration<sup>34</sup>) a representation of an object for the subject unless the subject has taken it to be so. The representation of objects thus requires spontaneity. Secondly, and related, Pippin also argues that a judgement or a belief would not count as a judgement or a belief if it was only a mental event in a causal system. For instance he says that "if it turns out we really are causal systems on the noumenal level, then the states, beliefs and judgements produced by such systems would not be epistemic claims, even if the beliefs can be said to correspond both to phenomenal and noumenal reality."<sup>35</sup> Pippin thus relates the necessity of spontaneity to the possibility of us gaining knowledge. The claim is that we must view the mental events of inner sense *also* as acts of spontaneity in order to see them as representations of objects, and thus as judgements directed at truth and falsity. So in order to see the subject of experience also as a subject of knowledge it is necessary that we think of the subject and its apperceptively

---

<sup>32</sup> Pippin (1987) p. 459 and p. 463.

<sup>33</sup> Pippin (1987) p. 460.

<sup>34</sup> Note the similarity between this view and Davidson's.

<sup>35</sup> Pippin (1987) p. 472.

conscious states not only as mental events in a causal system (in inner sense) but also as spontaneous (in apperception).

Pippin thus argues that it is a condition for seeing mental states as representations of objects (and as judgements and knowledge claims) that we consider them as being spontaneous activities, as well as being events in a causal system (when considered via inner sense).

Pippin also stresses the special form of self-awareness, which is involved in apperception. The reflexivity, which the adverbial qualification ‘apperceptively’ refers to, is not to be understood as demanding explicit self-ascription of the representation in question, nor should it be understood as involving explicit self-consciousness. Apperception is not an event in inner sense, it is not a awareness of one-self, which consists in a higher order thought, and it does not involve an inference about my own identity: “apperceptive thinking cannot be said to be an experience at all, but a necessary component of any possible experience of objects.”<sup>36</sup>

An almost identical argument is given by Allison (1996).<sup>37</sup> Allison argues that we must presuppose spontaneity to explain the normative character of the act of ‘taking as’ involved in the conceptualisation Kant is concerned with. This normative character is linked to apperceptive acts and it, unlike the data processing function involved in relative spontaneity, requires absolute spontaneity.<sup>38</sup> ‘Taking as’ is, according to Allison, an act that requires that one is *conscious* of the act of recognition, so that one takes oneself to take x as F.<sup>39</sup> Allison thus takes Kant to argue that conceptual recognition is a normative act, which requires that one is conscious of one’s act of taking x to be F, and that one is conscious of one’s reasons for this judgement. Allison thereby links synthesis to the concept of always consciously applying a rule. It is not sufficient that one conceptualises in accordance with rules, but rather one must consciously apply these rules in order to recognise something in a concept. Allison thus goes further than Pippin who explicitly denies that normativity requires that we always

---

<sup>36</sup> Pippin (1987) p. 463.

<sup>37</sup> This argument can be found in several places in Allison’s writings. I discuss the argument as it appears in ‘On Naturalizing Kant’s Transcendental Psychology’ and ‘Kant’s Refutation of Materialism’ both found in Allison (1996).

<sup>38</sup> Allison (1996) p. 63.

<sup>39</sup> This claim seems to be in danger of turning into a regress. Does one have to take oneself to take oneself to take x as F?



consciously apply a rule. Instead he holds that it requires only that we consciously follow a rule.<sup>40</sup> The act of conceptualisation thus requires, according to Allison, both apperceptive self-consciousness, that is, consciousness of one's own activity, and absolute spontaneity (the ability to question and reason over one's own activities and of choosing to apply rules).<sup>41</sup> Note that the self-consciousness required is not to be understood as involving a self-conception, so it is not consciousness of the self as an object, rather what is required is consciousness of one's own activity. This therefore explains Kant's claim that in apperception we are aware of our own spontaneity.

Allison's and Pippin's arguments are meant to show that relative spontaneity is not enough if we want to think of the subject of experience as a maker of judgements and a knower. If we want to think of mental states as representations of objects, then we have to think of them in terms of absolutely spontaneous acts.

I think Pippin's and Allison's ascription of absolute spontaneity to the subject is problematic because it exaggerates the activity of the subject in producing normative representations. The production of representations with objective content does involve an activity according to Kant, but this is true *only in the sense that* it involves a process of conceptualisation and synthesis. However, I think Allison especially, exaggerates this activity when he claims that the subject must self-consciously and deliberately 'take' something to be a representation in order for it to be normative, for two reasons. First, this does not seem to be a correct account of the origin of most of our normative representations, because most of our normative representations are not voluntary or deliberate and they do not seem to require self-consciousness. Second, this view leads Allison (and also Pippin, though to a lesser extent) into difficulties regarding the status of the subject or agent that does the 'taking'. The subject whose activity secures normativity in Allison's account cannot be the subject empirically considered, because when I look around me normative thoughts presents themselves to me without any act of mine (other than the act of looking at the world). We also cannot claim that it is the subject considered noumenally that is the agent of this activity because that would be to go beyond the boundaries of our knowledge of

---

<sup>40</sup> Pippin (1987) p. 460.

<sup>41</sup> Allison (1996) pp. 63-64 and p. 95.

the subject. It thus seems that Allison is forced to postulate agency without an agent to ascribe this agency to. This conflicts with the footnote to B158 (§25) which illustrates the fact that though I am conscious of my spontaneity I cannot determine myself as an active being. Kant says:

Now since I do not have another self-intuition which gives the *determining* in me (I am conscious only of the spontaneity of it) prior to the act of *determination*, as time does in the case of the determinable, I cannot determine my existence as that of a self-active being; all that I can do is to represent to myself the spontaneity of my thought, that is, of the determination; and my existence is still only determinable sensibly, that is, as the existence as an appearance. But it is owing to this spontaneity that I entitle myself an *intelligence*. (Footnote to B158).

Though I think Allison and Pippin fail to show that absolute spontaneity (of the kind he describes) is a necessary part of Kant's account of transcendental apperception, I also think that their arguments can help us develop the first argument, i.e. Ameriks' argument, I presented for spontaneity as a condition of objective knowledge. By relating spontaneity to normativity, Allison and Pippin show how transcendental apperception and the faculty of the understanding as described by Kant can only be realised by a system that can produce normative representations. However, in his argument for the normativity thesis Allison and Pippin depend on an interpretation of the *activity* involved in Kant's principle of transcendental apperception, which is too strong to be textually or philosophically defensible.<sup>42</sup> However, I do not think that our conception of normativity is dependent on this strong interpretation of activity. The demand for normativity therefore in itself puts limits on the possible realisations of the mind as described by transcendental apperception. However, as Sellars argues, the demand for normativity and relative spontaneity does not mean that the mind could not be realised by a causal system as long as it is a causal system which can produce normative representations

---

<sup>42</sup> Note that Kant does not identify theoretical reason with practical reason. While his account of practical reason does have a strong emphasis on voluntarism, I do not think that the same emphasis can be found in his view of theoretical reason.

### 1.h. Apperceptive self-Awareness

As we have seen, Kant claims several times and in different contexts that transcendental apperception involves a special kind of self-awareness. For instance he claims that transcendental apperception contains a reference to the capacity of becoming reflexively aware of one's representations as one's own ("It must be possible for the 'I think' to accompany all my representations..."). He also claims that by uniting a manifold of representations in one consciousness, I can represent to myself the identity of that consciousness. (B134). Finally, he claims that I am conscious of my spontaneity in apperception. (B158).

Brook (1994) gives an insightful description of the kind self-awareness involved in transcendental apperception with his discussion of apperceptive self-awareness. Apperceptive self-awareness is the "awareness of oneself as the common subject of one's representations".<sup>43</sup> It must therefore be distinguished from what Brook calls empirical self-awareness, which is awareness of psychological states such as desires. The first difference between them lies in the different status of that which they are consciousness of. Empirical self-awareness is, according to Brook, awareness of psychological states such as desires, that is, it is consciousness of the self as an object of a representation. Empirical self-awareness thus refers to the kind of self-awareness we have in inner sense. Apperceptive self-awareness on the other hand is 'awareness of one self as the common subject of ones representations.'<sup>44</sup> It is the kind of self-awareness expressed in the 'I think' and no determinate self-conception is involved in the self-awareness represented by this 'I think'.<sup>45</sup> As argued above, I think the kind of self-awareness we have in apperception is best expressed as being an awareness or an articulation of the unity of consciousness, that is the fact that all my conscious representations belong to the same subject of consciousness.

The second difference between empirical self-awareness and apperceptive self-awareness lies in the *kind* of self-awareness referred to. Empirical self-awareness, in Kant's view, involves a form of observation of inner states and it is thus

---

<sup>43</sup> Brook (1994) pp. 56-57.

<sup>44</sup> Brook (1994) pp. 56-57.

<sup>45</sup> See A381-2 and A355. Obviously it does contain a self-conception of oneself as a subject, but this is not a determinate self-conception.

equivalent to outer sense in this respect. However, the self-awareness involved in apperception does not involve any act of observation, inference or identification of oneself. The self-representation involved in apperception cannot involve acts of observation, inference or identification for two reasons: first, as no determinate self-conception is given in apperception (in the 'I think'), there is no mark or criterion by which the self could observe, infer or identify itself by. Second, self-awareness by observation, inference or identification is dependent on the existence of some self-representation prior to any act of reflection. In order for the subject to be able to observe or identify some state as its own, it must already have an idea of itself. A form of pre-reflective self-representation is thus a necessary condition for self-awareness by observation, inference or identification. Rather, apperceptive self-reference must be prior to self-reference via identification. Brook concludes that Kant's work on apperception is an early source of Shoemaker's self-reference without identification.<sup>46</sup> This seems to give a good account of the kind of self-awareness we have in transcendental apperception.

Brook's account of the kind of self-awareness we have in transcendental apperception supports my reading of the identity claim as a purely formal claim based on the unity of consciousness. What is expressed in apperceptive self-awareness (the 'I think') is simply that the subject of consciousness is the same in all the representations that are part of the unity of consciousness. The identity claim is thus restricted to the subject qua subject of consciousness, and it can be established simply on the basis that this representation does not involve any determinate representation of the subject. Brook's account of apperceptive self-awareness also supports the claim that our awareness of the spontaneity of thought does not give us an insight into ourselves as 'self-active beings'. This means that the fact that we are aware of spontaneity in thought does not tell us anything about the mind or the self that grounds this thought of spontaneity, because apperceptive self-awareness never gives us any determinate representation of the self.

---

<sup>46</sup> Brook (1994) pp. 70-72.

### **1.i. Transcendental Apperception and the Deduction of the Categories**

The ‘Transcendental Deduction’, I have assumed, is about objective experience in general in that it attempts to establish the a priori concepts after which experience must be structured in order for objective knowledge to be possible. The ‘Transcendental Deduction’ is concerned with describing the conditions and features of objective experience, rather than with proving objective experience. The fact that the ‘Transcendental Deduction’ is about experience in general has two main consequences, first, that the subject of experience in the ‘Transcendental Deduction’ is not necessarily human, but can be any finite, discursive maker of judgements. Secondly, that the deduction is in the first instance not a deduction of the specific categories (causation etc), but of the categories in general, simply as a priori rules after which intuitions are structured so that an objective experience is created about which we can make knowledge claims, which have objective validity. The ‘Transcendental Deduction’ thus contains an attempt at showing that the experience of a maker of judgements with objective validity must be structured after a priori rules.

Given my focus on the principle of transcendental apperception, I will, in what follows, concentrate on the role of transcendental apperception in the deduction of the categories. My aim with this section is therefore not to evaluate the success of the deduction of the categories, but rather to evaluate the use of the principle of transcendental apperception in reconstructions of the deduction of the categories. The importance of this in relation to my aims is to show that the interpretation, which I have advanced of the principle of transcendental apperception, can play its required role in the deduction of the categories. If this would turn out not to be the case, then it would follow that my reading cannot be exegetically accurate.

Henrich, Guyer, and Allison all believe that the role of transcendental apperception in the deduction of the categories supply further evidence for the interpretation they offer of this principle. In the following I will therefore abstract from the fact that the reconstruction these scholars offer, are arguments of a progressive, not regressive, form, and evaluate their use of the principle of transcendental apperception as a demand for knowledge of identity in a reconstruction of the deduction of the categories. I will also show that there are at least two possible reconstructions of the deduction of the categories based on the text, which make use of the principle of transcendental apperception only as a

principle of unity. As two reconstructions based on the text are possible, I do not think that the fact that transcendental apperception has an intended role to play in the deduction of the categories, shows that Henrich, Guyer, or Allison's interpretation of transcendental apperception are at an advantage over the interpretation I have given here.

Guyer and Henrich have identified two different starting points for a deduction of the categories in the 'Transcendental Deduction', which they have named respectively the objective and the subjective deduction. These two form the main strategies for a deduction of the categories, but further subdivision is, of course, possible within them.

The strategy of the subjective deduction covers versions of the argument, which deduce the categories from transcendental apperception. Henrich, Guyer, and Allison all give different versions of this argument.

Henrich defines the categories as: "the functions of synthesis without which the manifoldly given contents cannot 'stand together' [zusammenstehen]. They are necessary conditions of synthesis".<sup>47</sup> His deduction of the categories starts from the identity of self-consciousness, which he claims is a principle that combines unity and identity<sup>48</sup> (conceived of in a moderate sense. The fact that it is identity in a moderate sense means that the subject can be related to different representations while still being the same subject).<sup>49</sup> Henrich argues that the knowledge of the identity of the subject is only possible through the knowledge of the transitions of the same subject from state to state. As our knowledge of our identity is held with Cartesian certainty so must Cartesian certainty attach to everything which this knowledge is dependent on. The conditions of transition from one state to another must therefore be known with Cartesian certainty. The categories are seen as "the form of progression with which all transitions must comply".<sup>50</sup> Our Cartesian certainty of our own identity therefore makes necessary our knowledge of the categories as a priori rules of transition. Knowledge of the categories is a necessary condition for our identity of self-consciousness, of which

---

<sup>47</sup> Henrich (1994) p. 167.

<sup>48</sup> Henrich (1994) P. 186.

<sup>49</sup> Henrich (1994) p. 179-183.

<sup>50</sup> Henrich (1994) p. 197.

we have a priori certainty, and the categories must therefore have an a priori status.

Guyer's version of a successful deduction resembles Henrich's in that it also starts from self-consciousness, but where Henrich's self-consciousness has the status of incorporating a piece of given necessary knowledge, Guyer's deduction starts from the fact of empirical self-consciousness.<sup>51</sup> Such empirical self-consciousness is, he argues, dependent on our consciousness of the existence of our different states in time, but the representation of time depends on a representation of objects as existing independently of inner sense, as Kant argues in the Refutation of Idealism. Guyer's argument goes as follows:

To make determinate judgements about the temporal succession of subjective states at all, which is presupposed even by empirical knowledge that my continuing self actually has a manifold of representations, I must link these representations in some way to objects in space which are capable of both continued existence and yet determinate change. Only thus can I determine that my present state actually represents a 'sequence of one impression upon another'. Yet to make such connections require precisely that I make judgements about the continued existence of objects regarded as distinct from mere modifications of inner sense and judge the changes of such objects. What can this be but to apply the concepts of inherence and subsistence to things regarded as objects in a strong sense and to apply to such objects the further dynamic categories of causality and dependence and perhaps even reciprocity of action? Thus, making the temporal judgements presupposed by any self-knowledge requires the use of the categories, and it is at least necessary that *if* I make the former, then I must use the latter. (Guyer (1987) p. 152).

This version of the subjective deduction is very close to being an objective deduction, as the categories are deduced as the forms of synthesis of representations in the concept of an object. Guyer simply tries to base this deduction on something, which he finds less controversial than our experience of objects, by making empirical self-consciousness the starting point of his deduction. However, as he has not argued for the necessity of this self-consciousness his argument is at most conditional: If I make judgements about my

---

<sup>51</sup> This is Guyer's preferred reconstruction of the deduction of the categories, but in fact he thinks that this version of the deduction is only hinted at in the 'Transcendental Deduction' and comes closest to being realised in the 'Refutation of Idealism'. (Guyer (1987) pp. 75, 85).

temporal identity then I must make judgements about objects structured by the categories.

Allison argues that a special kind of self-consciousness is involved in the principle of transcendental apperception. He explains that this should not be understood as involving the consciousness of the self as an object, but rather the consciousness of the self as an activity.<sup>52</sup> I think, as I have argued above (in 1.g) that Allison's account puts too much emphasis on the *activity* of the subject. In normal perceptual states we are not in fact conscious of our own activity, as it is not a consequence of our own activity what representational states we are in most of the time.

Allison's version of the deduction is interesting because it combines the objective and subjective versions and uses both the principle of transcendental apperception and the concept of an object to deduce the categories. Allison's reciprocity thesis, which really is a thesis of necessary and sufficient condition, interlinks the principle of apperception with the concept of an object:

Since it follows from the apperception principle that the unity of consciousness is impossible apart from a synthetic unity of representations, and since this synthetic unity can only be achieved by uniting these representations under a concept, and since (by definition) any such synthetic unity counts as an object, it also follows that the representation of an object is a necessary condition for the unity of consciousness. But this is equivalent to saying that the unity of consciousness is a sufficient condition for the representation of an object, which is just what the reciprocity thesis asserts. (Allison (1983) p. 146).

This thesis is possible because of the broad definition of object, which Allison ascribes to Kant. Allison interprets Kant's definition of an object as "that in the concept of which the manifold of a given intuition is united" (B137) as meaning that any a priori synthetic unity counts as an object. However, Allison has not yet convinced us that an *a priori* synthesis is necessary for the unity of consciousness. This he does by identifying apperception with the faculty of the understanding and the working of the understanding with judgements (all on the basis of textual evidence):

---

<sup>52</sup> See his interpretation of spontaneity in 1.g.



The remainder of the argument merely unfolds the implications of this analytic principle. We see first that synthetic unity is a condition of the understanding (analytic unity), and consequently of the representation of an Objekt. The act of understanding is then identified with judgements. It follows from this that the manifold must conform to the logical functions of judgements if it is to be brought to the objective unity of consciousness, or conceptualised. If we accept the results of the Metaphysical Deduction, it also follows that this manifold is necessarily subject to the categories. (Allison (1983) p. 148). The presupposed judgement of self-identity thus presupposes judgements of objects. As a judgement in itself implies knowledge of necessary and universal connections, the status of the categories as a priori is secured.<sup>53</sup>

The principle of transcendental apperception as a principle of our awareness of our own identity plays therefore a central role in Henrich, Guyer, and Allison's versions of the deduction of the categories.

However, there are many places in relation to the deduction of the categories, where Kant refers only to the *fact* of the unity of consciousness or apperception and does not demand that the subject is *aware* of this fact. I have identified two main versions of the deduction in which Kant uses transcendental apperception only as a principle of unity.

The strategy called the objective deduction covers different versions of the argument, which attempts to deduce the categories as a priori rules from either the concept of an object or from the fact that experience contains objects.

An object is defined as "that in the concept of which the manifold of a given intuition is united" (B137). According to Kant nothing can be represented as united unless it has been so united by the subject in an active synthesis of the manifold. One can thus see how the principle of transcendental apperception plays a role in the objective deduction because it, as a principle of unity, is what makes the synthesis of a manifold possible. However, this synthesis must happen according to a priori rules in order to create an experience, which contains objects (and thus has objective validity). This argument thus deduces the categories from the concept of an object, as a priori rules of synthesis the manifold of intuitions after which this synthesis must happen in order to synthesise them in the concept of an object.

---

<sup>53</sup> "the very idea of judgement itself implies knowledge of a necessary and universal connection" (Guyer (1987) p. 94).

This deduction rests on several idealist assumptions, most importantly that any representation of something as combined must have been synthesised by the subject. The principle of transcendental apperception's role in the objective deduction is limited to the principle of unity, which makes synthesis possible. However, transcendental apperception does play a role in this interpretation of the deduction and this would explain why Kant should include a discussion of this principle in the chapter on the deduction of the categories, even if he meant to deduce the categories from the concept of an object.

There is also evidence of a version of the subjective deduction that uses only the unity of apperception as a premise, without relying on the necessity of the subject's becoming aware of its own identity. §20 is an example of the evidence of such a deduction. In §20 Kant deduces the categories as the logical functions of judgements, where judgements are the form in which representations are brought under one consciousness. This version of the deduction thus makes use of the principle of transcendental apperception only as a principle of unity. The argument here claims that the categories are necessary because they are "the functions of judgements, in so far as they are employed in determination of the manifold of a given intuition" (B143). Earlier Kant has said that representations, in so far as they allow for being combined, must conform to the conditions under which they can be brought to the unity of consciousness (B136). I think this fits with this version of the subjective deduction, which Kant hints at further in §20.

The principle of transcendental apperception as a principle of unity can thus play an important role in at least two different versions of the deduction of the categories. In a purely objective deduction it plays a role as principle of unity, which makes synthesis possible. In the final version, which I have discussed briefly, it plays a role as a necessary analytic unity, which is dependent on a synthetic unity. The synthetic unity consists of a manifold brought under the analytic unity of apperception via judgements, the functions of which are the categories.

### **1.j. Conclusion of Part 1**

Kant's transcendental treatment of the subject in §16 tells us that any subject of experience will have unity of consciousness, weak, formal identity of subjects of

consciousness, and the ability apperceptively to self-ascribe its representations. We thus know that there must be a unity of consciousness between different representations in order for synthesis between the representations to be possible and in order for us to be able to say that the representations in question form part of one experience. As experience is necessarily synthesised (which Kant argues for in §15), the mind is necessarily unified. Following from the unity of consciousness is also a formal identity claim, which tells us that the subjects of consciousness are identical to each other within one unified experience. We have also established the necessity of a possibility of apperceptive self-ascription of representations, which involves what Shoemaker would call self-reference without identification. It has also been argued that transcendental apperception is an act of spontaneity. This spontaneity shows itself in two areas: first, in the fact that higher order I-thoughts arise 'spontaneously' from first order conscious representations, which do not contain any representation of the subject prior to this act of self-ascription. Second, our spontaneity shows itself in the synthesis which takes place in the understanding, because the kind of conceptualisation it undergoes also does not lie in the given. Finally, I have shown how the principle of transcendental apperception as I have interpreted it plays a role in at least two possible reconstructions of the deduction of the categories, namely in the objective deduction and in a version of the subjective deduction which is based on the principle of unity.

The account of transcendental apperception that we find in the 'Transcendental Deduction', can be thought of as describing certain capacities, the performance of which are naturally and necessarily ascribed to a mind in so far as it possesses objective knowledge. Central to transcendental apperception is the necessity of the possibility of apperceptive self-ascription. I think that this capacity can only be ascribed to a mind, so that a creature that has the capacity for apperceptive self-ascription is necessarily also a minded creature. (Apperceptive self-ascription is thus a sufficient but not a necessary condition for mindedness). Likewise the capacity for spontaneous synthesis and judgement making are the capacities of a mind. Kant's principle of transcendental apperception describes capacities of the mind. I thus think that the analysis of transcendental apperception in the 'Transcendental Deduction' has made room, conceptually, for us to talk about the mind in a way that was not open to us prior to the analysis. Prior to the analysis

we could talk only of the subject of experience in relation to the possibility of objective knowledge, but given that some of these preconditions turned out to be necessarily ascribable to a mind (as the unity of consciousness and active synthesis are necessarily capacities and properties of mind (conceptually speaking)) the question arises of whether this forms the basis of a Kantian theory of mind and if so what status this theory will have. This is the question I will be dealing with in the second part of this dissertation.

## **2. A Theory of Mind?**

### **2.a. Introduction**

My analysis of the principle of transcendental apperception in the ‘Transcendental Deduction’ has opened up conceptual room for talking about the mind transcendently, which was not the case prior to the analysis. Prior to the analysis we could talk only of the possibility of objective experience and its preconditions. However, the fact that certain of these preconditions, namely the unity of consciousness and spontaneous synthesis, are necessarily ascribable to a mind opens up the prospect of formulating a Kantian theory of mind on the basis of this analysis. In this section I will examine Kant’s own answer to this question, which can be found in the first edition ‘Paralogisms’. I think Kant’s treatment of the issue here is clear and coherent, however, in this section I will also deal with some of the common objections and issues that arise from my analysis of the ‘Paralogisms’.

### **2.b. The ‘Paralogisms’**

In my examination of the paralogisms my focus will be on what they tell us about Kant’s principle of transcendental apperception, and therefore not on establishing the fallacy, which each paralogism commits. As the aim of this analysis is to expand my analysis of Kant’s theory of the self and the mind, I have chosen not to give an analysis of the fourth paralogism, as the aim of this paralogism is not to describe the self, but rather to explore the relationship between inner and outer sense.

The paralogisms are invalid generally because they draw conclusions about how the soul (in the Kantian sense of the world) is noumenally. Part of Kant's aim in the 'Paralogisms' is to show that the rationalist view of the self as a soul which is a simple substance with personal identity over time, is not certain or necessary, because there are alternative accounts of the self which also make sense of the data, that is, the different aspects which we have established belong to transcendental apperception (the second premise in all the arguments). As such alternatives are available, the rationalists are not justified in their claims about the self as a thing in itself.

My aim with this analysis is twofold, firstly and negatively to identify what Kant thinks we cannot know about the subject, and secondly, and positively, to identify anything new that Kant may say about transcendental apperception, which can help to expand our account from the 'Transcendental Deduction' of this way of looking at the self. I will therefore look at two main questions with regards to each paralogism. First, what it is that Kant here argues we cannot know about the noumenal self, and second what his analysis tells us about the self viewed transcendently. Of these two questions the second is primary, as the aim with this chapter is to establish whether Kant's account of transcendental apperception amounts to a theory of mind. What is interesting in a Kantian analysis of the self is to establish what kind of knowledge of the self an analysis of possible experience gives us and what this knowledge amounts to. I therefore hope to establish not only what, as Kant in the 'Paralogisms' argues, we cannot know about the self as it is in itself, but also what he thinks we *can* establish about the self when viewing it transcendently.

The purpose of the 'Transcendental Deduction' and of the 'Paralogisms' is obviously different, and consequently the notion of 'the subject of experience', which Kant uses in the two chapters, is also slightly different. The aim of the 'Transcendental Deduction' is to deduce the necessary features, which make objective experience in general possible. The results of the 'Transcendental Deduction' should therefore hold for creatures with other sensible conditions of experience than ours. The subject of experience in the 'Transcendental Deduction' is, therefore, not specifically us. The 'Paralogisms', in contrast, are concerned with refuting the arguments of the rational psychologist. The subject they are concerned with is, therefore, the subject which the rational psychologist is

concerned with, namely us. These two differences, of course, mean that the positive results from the ‘Paralogisms’ cannot be considered as valid for the subject of the ‘Transcendental Deduction’. However, Kant’s positive claims in the ‘Paralogisms’ are all concerned with the transcendental aspect of the self. There are two ways in which this concern can be formulated. Either we can understand the phrase ‘the way the self must be represented to itself’ as making claims about how the self must necessarily view itself pre-philosophically in order to have experience, or we can understand it in the weaker sense of referring to the way in which we must represent the self in a philosophical reflection. Given the aim of the ‘Paralogisms’, I find the second interpretation most plausible. Kant is not here concerned with establishing any necessary conditions for experience, but simply with distinguishing two non-empirical ways of looking at the self: transcendently, as the way in which we must represent the self, and noumenally, as the way the self is in itself. The first of these approaches is legitimate whereas the other is not.<sup>54</sup>

### **2.c. The First Paralogism – The Self as Substance**

The first paralogism is stated in the first edition as:

- 1: that, the representation of which is the *absolute subject* of our judgements and cannot therefore be employed as determination of another being, is substance
- 2: I, as a thinking being, am the *absolute subject* of all my possible judgements, and this representation of myself cannot be employed as predicate of any other thing.
- 3: Therefore I, as thinking being (soul) am *substance*. (A348).

This argument is repeated in the second edition ‘Paralogisms’ as:

- 1: That which cannot be thought otherwise than as subject does not exist otherwise than as subject, and is therefore substance.
- 2: A thinking being, considered merely as such, cannot be thought otherwise than as subject.
- 3: Therefore it exists also only as subject, that is, as substance. (B410-411)

---

<sup>54</sup> Powell (1990) p. 132.

In both editions the problem with this argument is identified as the lack of a middle term. The first premise is an analytic categorical proposition<sup>55</sup> giving a definition of the concept of 'subject'. The second premise, however, is only about the subject from a transcendental point of view, and thus tells us only how we must represent ourselves to ourselves, not how we actually are. The second premise cannot therefore establish that we fall under the concept of the first premise. The conclusion thus does not follow and the syllogism is invalid. This argument is dependent on the acceptance of a hidden premise, namely that we cannot infer from the way in which we must represent something to the way this thing really is. The acceptance of this hidden premise is in fact the basis of all four paralogisms.

We cannot apply the empirical category of substance to the 'I' of transcendental apperception, because we do not have any intuition of the self as an object. All we have in the 'I think' is the unity of consciousness as a way in which we must represent ourselves, but to which there belongs no intuition. The fact that we must represent ourselves as the subject of all our judgements does not mean that we know anything about what underlies this representation of ourselves.<sup>56</sup> The material of the second premise thus does not allow us to conclude that the self is (schematised) substance.

The first paralogism thus tells us; negatively, that we cannot know if the soul is a substance, because our knowledge of the self transcendently does not extend to knowledge of the self viewed noumenally, and so although we know that we must represent ourselves always as subject, this does not mean that we fall under the concept of substance from the first premise, if this is taken as schematised category.

Positively, the first paralogism tells us that the self must be *represented* as subject, and that "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,

---

<sup>55</sup> Powell (1990) p. 66.

<sup>56</sup> Ameriks (2000) pp. 69-70.

the everlasting duration of the human soul in all changes and even in death (...)” (A350-351).<sup>57</sup>

Kant bases his claim that the self must be represented as subject on the observation that: “Now in all our thoughts the ‘I’ is the subject, in which thoughts inhere only as determinations; and this ‘I’ cannot be employed as the determination of another thing. Everyone must, therefore, necessarily regard himself as substance (...)” (A349).

This clearly refers to the doctrine of transcendental apperception from the ‘Transcendental Deduction’. Powell (1990) argues that the first clause of the second premise: “I, as thinking being, am the absolute subject of my judgements” refers to the role of the unity of transcendental apperception, within which all representations must occur to be experienced. It is correct to say that “I, as thinking being am the absolute subject of my judgements”, because of this necessary unity of apperception. However, it is important here to draw a distinction between being a subject of judgements in the sense of being the subject of the content of that judgement, and in the sense of being the maker of that judgement. Powell draws this distinction and calls subject in the first sense Subject(1) and subject in the latter sense Subject(2). It should be clear that the ‘Transcendental Deduction’ establishes that the ‘I’ is the subject of all my judgements in the sense of Subject(2), because it is as the maker of the judgement it is necessary, not as the subject of the topic of the judgement.<sup>58</sup> I think this distinction further supports my analysis of §16, both in the sense that the ‘I think’ is simply the subject of conscious representations but may contain no representation of the self, and in the sense that it is not the existence of an actual identity claim which is important to transcendental apperception. The ‘I’ of transcendental apperception is the subject in the latter sense of being the necessary maker of all my judgements, not in the former sense of being the necessary subject of the content of all my judgements – the possibility of self-ascription is thus a possibility that follows from the transcendental unity of apperception rather than what makes this unity possible.

On the basis of his distinction Powell sets up the first paralogism in such a way

---

<sup>57</sup> Ameriks (2000) p. 67.

<sup>58</sup> Powell (1990) pp. 76 and 79-82.



that it is clearly fallacious, due to the lack of a middle term:

- 1: That which is a subject(1) is substance
- 2: the 'I' of the 'I think' (= subject(2)) must be represented as a subject(1)
- 3: Therefore the 'I' of the 'I think' is substance. <sup>59</sup>

The first paralogism can thus be seen to emphasise the analysis of apperception I have given in my analysis of the 'Transcendental Deduction'. The 'I' must always consider itself the absolute subject of all its judgements, and must therefore represent itself as (unschematised) substance. However, it also shows that the fact that the subject must represent itself as the common subject of all its representations does not mean that we can know that the subject is a substance.

#### **2.d. The Second Paralogism – Of Simplicity**

The second paralogism is given its formal presentation in A351:

- 1: That the action of which can never be regarded as the concurrence of several things acting, is simple.
- 2: Now the soul, or the thinking 'I', is such a being
- 3: Therefore the soul is simple. <sup>60</sup>

The central line of argument in the second paralogism is not easy to identify because of the way Kant sets up this argument. However, I think both Kant's negative and positive conclusions stand out clearly nonetheless. The paralogism is invalid because it has an ambiguous middle term. The second premise establishes only that the soul must be represented to the subject as simple, while what we need to establish the conclusion the rationalist is looking for, is that the soul *in itself* is a being the actions of which can never be regarded as the concurrence of several things acting. <sup>61</sup> Kant expresses this confusion in A355: "But the simplicity of the representation of the subject is not *eo ipso* knowledge of the simplicity of the subject itself". The point is the same made in the first paralogism, namely that our knowledge of the self viewed transcendently does not give us knowledge of

---

<sup>59</sup> Powell (1990) p. 82.

<sup>60</sup> A very similar formulation is given in B407-408.

<sup>61</sup> Powell (1990) p. 101.

how the self actually is noumenally.<sup>62</sup> Kant thus wants to deny that we have any knowledge of the self as simple, as this cannot be established empirically, as the self is not an object, or conceptually because it is not an analytic proposition. Kant makes a second negative point in the second paralogism, namely that even given the soul's simplicity the rationalist cannot argue for the soul's immortality. Kant therefore allows for the sake of the argument that the rationalist has proved that the soul is simple, and he also allows that the soul is non-corporeal (which he actually agrees with if the soul is taken as inner sense), but even given these two premises the rationalist cannot show that the soul *in itself* is non-corporeal and thus indestructible.<sup>63</sup> Kant's argument for this is again that the representation of the soul in inner sense and material objects in outer sense tells us nothing about the substratum of these appearances (A358-359).

Positively, the second paralogism tells us that the self must be represented as simple. The second paralogism thus gives us an elaboration on a point we have already established in the 'Transcendental Deduction', namely that the I of apperception is represented as simple. Kant expresses this fact in statements like: "This proposition ['I think'], however, is not itself an experience, but the form of apperception, which belongs to and precedes every experience; and as such it must always be taken only in relation to some possible knowledge, as a *merely subjective condition* of that knowledge." (A354). And later in: "Nor is the simplicity of myself (as soul) really *inferred* from the proposition, 'I think'; it is already involved in every thought. The proposition '*I am simple*', must be regarded as an immediate expression of apperception, (...). '*I am simple*' means nothing more than that this representation, 'I', does not contain in itself the least manifoldness and that it is absolutely (thought merely logical) unity." (A355). And finally: "It means a something in general (transcendental subject), the representation of which must, no doubt, be simple, if only for the reason that there is nothing determinate in it. Nothing, indeed, can be represented that is simpler than that which is represented through the concept of a mere something." (A355).

I think these quotes support Brook's interpretation of the reflectivity of the 'I think' as self-reference without identification. The two last quotes tell us that

---

<sup>62</sup> Powell (1990) p. 97.

<sup>63</sup> Powell (1990) p. 105.

there is nothing determinate in apperception and as the transcendental self is represented as a mere something, this compares the transcendental subject to the transcendental object = x, which is also done in the ‘Transcendental Deduction’.

As the first paralogism, the second paralogism also has a weak reading in which the argument is valid. Kant thus says that we can conclude that the self is simple if we take this to mean nothing more than that the self must be transcendently represented as simple.

### **2.e. The Third Paralogism – Of Personality**

The third paralogism is formulated in A361 as follows:

- 1: That which is conscious of the numerical identity of itself at different times is in so far a person
- 2: Now the soul is conscious of the numerical identity of itself at different times
- 3: Therefore it is a person.<sup>64</sup>

Again I think it is difficult to identify Kant’s central line of argument, but relatively easy to identify his negative and positive conclusions. Negatively Kant wants to deny that we have knowledge of the numerical identity of the soul as it is in itself just because it is a precondition for thought that we are represented as having numerical identity. In the footnote to A363 he shows the possibility of the self in itself failing to have numerical identity while the self is still represented as having this quality.<sup>65</sup>

If, then, in analogy with such bodies, we postulate substances such that the one communicates to the other representations together with the consciousness of them, we can conceive a whole series of substances of which the first transmits its state together with its consciousness to the second, the second its own state with that of the preceding substance to the third, and this in turn the states of all the preceding substances together with its own consciousness and with their consciousness to another. The last substance would then be conscious of all the states of the previously changed substances, as being its own states, because they would have been transferred to it together with the

---

<sup>64</sup> See also B408.

<sup>65</sup> Ameriks (2000) pp. 35-136.

consciousness of them. And yet it would not have been one and the same person in all these states.

The negative conclusion on the third paralogism is thus the same as on the first and second: the transcendental representation of the self does not allow us to draw any conclusions about how the self is in itself, and so the noumenal ignorance thesis is sustained.<sup>66</sup> The positive conclusion is, however, much more interesting than the positive conclusion of the first and second paralogisms, because Kant goes much further in his claims about the numerical identity of the transcendental subject than he does in the 'Transcendental Deduction'. In the 'Transcendental Deduction' identity is introduced as an analytical principle following from the transcendental unity of apperception. It is a merely formal fact of the unity of consciousness, but not one the subject needs to be aware of to have experience. However, here Kant links identity to self-consciousness and claims (in the second premise) that the soul is conscious of the numerical identity of itself at different times. The difference between the identity claim in the 'Paralogisms' and the 'Transcendental Deduction' may of course be explained by the different aims of these two chapters of the *Critique of Pure Reason*. Where as the subject in the first part of the 'Transcendental Deduction' is the subject in general in the sense of being any maker of judgements, the subject in the 'Paralogisms' is the subject, which the rational psychologists are concerned with, that is, a human self-conscious subject. Kant's references to self-consciousness or identity in the 'Paralogisms' can therefore simply be interpreted as a reference to an ability we humans have as a matter of fact, and it therefore does not involve any claim as to the necessity of this ability. I therefore do not believe that the second premise in itself refutes my analysis of the identity claim in the 'Transcendental Deduction'. However, there is another difference between the 'Paralogisms' and the 'Transcendental Deduction', which I think can be used in an argument for the claim that the ability to become aware of one's own identity is a necessary for certain kinds of experiences, and thus that these experiences are based on a cognitive achievement. The subject from the 'Paralogisms' is in time, and I think that certain experiences involving the awareness of the succession of time are

---

<sup>66</sup> Ameriks (2000) pp. 131-134.

dependent on the cognitive achievement to be able to ascribe successive representations to the same 'I' over time. The process of counting, for instance, is dependent on such a cognitive achievement. Unless one is able to recognise a past representation as one's own and to connected present representations to former representations as being the representations of one self, one would not be able to perform the act of counting.<sup>67</sup> Certain forms of diachronic synthesis, such as counting, thus require that one is able to self-ascribe past and present representation to the same identical 'I'. However, such a cognitive achievement seems only to be necessary for certain kinds of diachronic synthesis (and possibly also for some kinds of synchronic synthesis). However, the kinds of synthesis that require self-ascription to an identical 'I' are all deliberate acts of synthesis and all acts that require a high level of conceptualisation and cognitive power anyway, so to say that these acts are also dependent on the ability to self-ascribe representations to an identical self is not to say that all synthesis requires this ability.

Kant's positive claims in the third paralogism thus do not contradict the claims he makes about identity in the 'Transcendental Deduction'. The third paralogism tells us that the formal identity principle established in the 'Transcendental Deduction' will, for beings like us who have the ability to become self-aware in time, lead to a representation of the self as identical to itself in all the representations that are synthesised together in one unified experience. This, however, only tells us how a self-conscious being must necessarily represent itself to itself and it does not therefore establish any substantial identity claim about the self as it is in itself. Finally, the third paralogism points to certain forms of (diachronic) synthesis, which requires that the self ascribes all the representations involved in the act to one identical self. I think counting is a good example of such an act.

---

<sup>67</sup> Powell (1990) p. 133.

## **2.f. Conclusion on the 'Paralogisms'**

The minor premise of the three paralogisms thus tells us that the self must be represented as a simple substance and as a person. The latter feature is, however, dependent on the subject in question making claims about its own identity at different times, and does not therefore have the same general application as the other two. That the self must be represented as a simple substance is a conclusion Kant reaches from facts about the transcendental unity of apperception, established in the 'Transcendental Deduction'. That the I of apperception must be represented as simple, follows from the unity of consciousness: the I of apperception relates to a manifold but cannot itself be divided. The similarity between the I of apperception and the transcendental object =  $x$ , is also emphasised in the 'Paralogisms': the I of apperception is simple, because there is nothing determinate in it. That the self must be represented as a substance, follows from Kant's claims about the personal form of all conscious representations: the self is the necessary subject (in the sense of being the maker of judgements) of all its judgements and must therefore be represented as a substance.

However, the fact that the self must be represented as a simple substance does not mean that we have any positive knowledge about what, if anything, underlies this representation. Kant can refute the rational psychologists because he can show that their conception of the soul is not the only thing that can explain the fact that we must represent ourselves as simple substances.

## **2.g. Noumenal vs. Transcendental Considerations of the Mind**

One way of expressing the result of the analysis of the 'Paralogisms' is via the distinction between noumenal and transcendental accounts of the mind. What the 'Paralogisms' have shown is that we cannot formulate a Kantian theory of mind as it is in itself (noumenally), but we can formulate a Kantian theory of mind considered transcendently. A transcendental consideration of the mind shows how the mind must necessarily be represented in a reflection of the necessary conditions of objective experience. The analyses of transcendental apperception in the 'Transcendental Deduction' and in the 'Paralogisms' have shown that the mind must represent itself as unitary and simple, and must have the capacity to

self-ascribe its conscious representations and to unite its representations in spontaneous synthesis.

The transcendental perspective is not to be identified with a perspective on either phenomena or noumena. It is thus a mistake to argue, as Kitcher does, that since Kant thinks that we can know nothing about the noumenal mind, transcendental apperception must be about the phenomenal mind.<sup>68</sup> Kitcher's mistake lies, I think, in trying to identify an *object*, which the transcendental consideration of mind is a theory of. The answer to how transcendental apperception should be placed within the Kantian framework depends, I think, on how we interpret the distinction between noumena and phenomena. Especially, important is the question of whether the distinction should be interpreted as involving a distinction between two worlds or two objects, or if it should be interpreted as a distinction between two ways of considering the world or objects of our experience. Only once we have established the interpretation of the phenomena/noumena distinction can we place the transcendental consideration within this framework. The two-object interpretation can be found in Strawson among others, while the two-aspect interpretation has been defended by Bird, Allison and others. I will not go into this discussion here, but simply state that I believe that the distinction should be understood as a distinction between two ways in which we can consider objects.<sup>69</sup> When we talk about noumenal objects or the noumenal mind in distinction from phenomenal objects or the phenomenal mind we are therefore not making a distinction between two objects, but rather between two ways in which we can consider objects or the mind: either phenomenally, that is, under the epistemic conditions for our experience, or noumenally, that is, in abstraction from these epistemic conditions.<sup>70</sup> The most plausible versions of the two-aspect reading of transcendental idealism are given by Bird and Allison. In this version it is important to note that the object of consideration is picked out in its empirical form and that it is this object that we attempt to consider in abstraction from our epistemic conditions when we consider it noumenally.

---

<sup>68</sup> Kitcher (1990) pp. 21-22.

<sup>69</sup> Matthews (1982) gives a good discussion of these two ways of interpreting the distinction, and argues persuasively in favour of the 'two-aspects' interpretation.

<sup>70</sup> Here noumenally is taken in its negative sense as meaning simply non-sensible, as such Kant uses it as equivalent to 'thing in itself' and 'transcendental object=x'. See below.

To ask whether the 'Transcendental Deduction' gives us knowledge of the noumenal mind or of the phenomenal mind is therefore misleading, because the noumenal mind is not an object in itself distinct from the phenomenal object. Instead we have to examine the way in which we consider the mind in the 'Transcendental Deduction' and to what extent this gives us any positive knowledge of the mind.

A qualification is necessary here. Kant uses the term 'noumena' in two different ways in the *Critique of Pure Reason*. In the positive sense of noumena, this concept refers to the experience of an actual intellectual intuition. The way objects are to such an intuition can never be known to us because we do not ourselves have the capacity for intellectual intuition. The second and negative sense of noumena, however, refers to a consideration of objects in abstraction from our sensible intuitions of them, and so means simply 'non-sensible'. Noumena in the second, negative sense is equivalent to the terms 'things in themselves' and 'transcendental object = x',<sup>71</sup> in the sense that these terms are also ways of referring to the objects of our experience in abstraction from the form given to them by us. The positive sense of 'noumenal' specifies the way things would be experienced by a being with intellectual intuition, and we cannot know what objects would be like to such a being. Kant is known for saying that knowledge requires both intuitions and concepts, and because our positive talk of mind or objects in abstraction from our sensible intuition of them does not (obviously) involve any intuition of such objects (as that would require that we had intellectual intuitions), he wants to distinguish it from intuition-involving knowledge. He makes this distinction by calling our considerations of objects, in abstraction from our sensible intuition, *thoughts* rather than knowledge. A further distinction is necessary here, namely the distinction between the legitimate knowledge we gain from doing transcendental philosophy and the claims to knowledge which are of a transcendent nature, and which are thus illegitimate. Again the knowledge we gain from a transcendental reflection is linked to the negative sense of 'noumena', where as the illegitimate transcendent knowledge makes claims about noumena in the positive sense.

---

<sup>71</sup> The connection between the negative sense of noumena and the transcendental object=x is explained by Allison (1983) pp. 245-246.



The two-aspect reading of the distinction between noumena and phenomena can explain how Kant can talk about things in themselves without violating the noumenal ignorance thesis. I want to illustrate how it deals with this problem by presenting two of Allison's arguments about objects considered as things in themselves. It is often argued that Kant violates his own noumenal ignorance thesis when he claims that things as they are in themselves affects us and provide the content, though not the form, of our appearances/representations. Allison argues that Kant is allowed to make this claim, because it should not be understood as a positive ontological claim about noumenal objects. So Kant is not claiming about noumenal objects that they affect us. Rather he says that when we consider the something, which affects the mind (the topic is thus not empirical affection), we must think of it as a thing in itself, and in abstraction from its empirical form, because this form is given to it by us,<sup>72</sup> and thus we cannot think of that which affects the mind under a description which already involves the mind. Kant is thus not making any positive claims about how things in themselves *are*. Rather he describes how we must think of the something, which affects the mind, when we consider the necessary affection of the mind.

Similarly it is often argued that Kant violates the noumenal ignorance thesis when he claims that things in themselves are not in space and time. Again Allison shows how Kant's claims are legitimate, when we remember that he does not mean to make ontological claims about distinct noumenal entities here, but rather to make clear that when we think of objects in abstraction of our sensible intuitions of them, then we have to think of them in abstraction of space and time, which are the forms of our intuition.<sup>73</sup>

However, Kant's claims about the mind in the 'Transcendental Deduction' are not about the noumenal mind (as the 'Paralogisms' show) or the mind noumenally considered. If they were, Kant's claims would have to be purely negative claims, like the claims about things in themselves discussed above, which simply make reference to the mind considered in abstraction from space and time. Kant's claims about transcendental apperception are clearly not like that. Rather than being a consideration of the mind considered noumenally or a consideration of the

---

<sup>72</sup> Allison (1983) p. 250.

<sup>73</sup> Allison (1983) p. 252.

mind considered phenomenally, it is a consideration of the mind considered transcendently, that is, under its necessary representation in a transcendental reflection. I think the two-aspect view of transcendental idealism can help us in making sense of what kind of theory of mind a transcendental consideration of the mind can be. Most importantly the two-aspect theory shows that claiming that Kant gives a theory of mind considered transcendently does not commit us to the existence of a third object – the transcendental mind. All it introduces is the idea of a third way of considering the mind, namely transcendently in terms of the necessary representation of mind when we consider it in a reflection of the conditions for objective experience. On this reading the ‘Transcendental Deduction’ does not make any claims about a distinct object, but is a description of the mind under a special consideration associated with transcendental reflection.

One way of stating the difference between a transcendental theory of mind and a noumenal theory of mind (a theory of mind as it is in itself) is that the latter but not the former can tell us something about the ontology of the mind. While Kant cannot say anything about the realisation of the functions or capacities of mind, he can and does say something about these functions and capacities. This is one moral that the ‘Paralogisms’ seem to support: there is more than one way in which we can account for the functions and features of mind established as necessary for experience, and the rational psychologists are therefore not justified in making claims about how these features and functions are realised in us.

The fact that spontaneity is a necessary part of transcendental apperception and the fact that the subject’s representations are normative means that the kind of system that can realise these functions must be a system that can produce normative representations. Sellars’ argument has, however, shown us that there is nothing in the conception of spontaneity that means that this capacity could not be realised by a causal system.<sup>74</sup>

However, the fact is that it is not even possible to show that the capacities described *need* to be realised by or grounded in something else, and so the idea of a realiser or grounding of the capacities described in transcendental apperceptions is a mere possibility that we cannot say anything informative about. Instead we

---

<sup>74</sup> Sellars (1971), see 1.g above.

may just have to think of Kant as describing a set of capacities and functions. These functions may, or may not, be realised by something else, which we, in the former case, would be unable to say anything informative about.

The ‘Paralogisms’ can be read as supporting this interpretation: since it is possible for the capacities contained in the principle of transcendental apperception to be grounded in different ways, speculations about the grounding of these capacities are not a legitimate part of a transcendental reflection.

## **2.h. ‘The Imaginary Subject of Transcendental Psychology’**

Strawson (1966) describes the part of Kant’s claims about the transcendental apperception, which describes the activities of the faculty of the understanding (such as synthesis and spontaneity) in the following way:

It is useless to puzzle over the status of these propositions. They belong neither to empirical (including physiological) psychology nor to an analytic philosophy of mind, *though some of them have near or remote analogues in both*. They belong to the imaginary subject of transcendental psychology, a part of the Kantian model.<sup>75</sup>

While I don’t agree with everything Strawson says here, I think he is correct in at least the first part of his statement. Kant’s claims about transcendental apperception do not belong to empirical psychology – because they are not about the mind empirically considered, nor are they about our minds under a noumenal consideration. They also do not belong to an analytic philosophy of mind, because Kant is not concerned with establishing the necessary conditions of mindedness in the ‘Transcendental Deduction’, and his principle of transcendental apperception should not be seen as describing the conditions for mindedness. However, the principle will have consequences for a theory of mind in both philosophy and in psychology, because any theory, which attempts to describe the mind of a creature with objective experience, should conform to the results of the transcendental consideration of the mind in the ‘Transcendental Deduction’.

The last part of Strawson’s claim is, however, problematic. It is problematic because the claim that Kant’s claims belong only to an imaginary subject of

---

<sup>75</sup> Strawson (1966) p. 97.

transcendental psychology ignores the importance of these claims. It is also problematic because Kant clearly does not believe that the subject of transcendental apperception is imaginary. He continuously claims that we are confronted with something real in transcendental apperception. What Strawson is ignoring here is Kant's insistence that transcendental apperception necessarily involves an awareness of our own spontaneous activity, and therefore a point of contact with ourselves, though this point of contact does not involve any determinate conception of ourselves.

The claims Kant makes about transcendental apperception are not just about an imaginary subject which is a side product of Kant's model, rather they tell us something about the capacities and functions necessary for objective experience in general. As these capacities and functions are necessarily ascribable to the mind, it tells us something about the capacities of the mind that are necessarily common to all creatures which have objective experience. One way of expressing this is to say that transcendental apperception stands in a relationship of determinate to determinate to all the different possible minds of creatures with objective experience including human beings.<sup>76</sup> It does not constitute a theory of mind in the sense of describing the ontology and metaphysics of the 'real' mind or of mind as it is in itself. However, given that it identifies some of the necessary capacities of minds of creatures with objective experience, it is a theory of mind in that sense of the term. Transcendental apperception is thus not just a feature of the Kantian model to be discarded once that model has served its purpose. It is an important principle, which describes the capacities, which all minds enjoying objective experience have in common.

## **2.i. Transcendental Apperception and Cognitive Science**

In her book, *Kant's Transcendental Psychology*, Kitcher argues that Kant's transcendental psychology has been unjustly ignored or rejected by Kant scholars. She argues that the result of this is both that Kant has not been recognised as making a valuable contribution to contemporary cognitive science, and that the

---

<sup>76</sup> See Yablo (1992): "some properties stand to others as *determinate* to *determinable* – for example, *crimson* is a determinate of the determinable *red*, *red* is the determinate of *colored*, and so on." (Yablo (1992) p. 252).

rejection of this part of Kant's philosophy has led to misinterpretations of important parts of the Critique, most notably the 'Transcendental Deduction'.

According to Kitcher, the 'Transcendental Deduction' should be read as describing the necessary features of a mind, which engages in cognitive processes. She sees Kant as arguing that it is only possible for something to have cognitive states if it realises a system of interrelated states necessarily synthesised together. Minds, she concludes, are "no more than contentually interconnected systems of cognitive states"<sup>77</sup> according to Kant. I think Kitcher is right in claiming that Kant is describing some psychological processes in the 'Transcendental Deduction' – especially when she describes the three forms of synthesis in the first edition 'Transcendental Deduction': these forms of synthesis are psychological processes because they take place in time. Moreover, Kant is certainly concerned with explaining the workings of the faculty of the understanding, and the principle of transcendental apperception contains capacities and functions, which can only be ascribed to the mind. However, this latter aspect of the 'Transcendental Deduction' describes mental capacities but not psychological processes, because these capacities are considered in abstraction from time. Kitcher thus goes too far in her interpretation of Kant. She takes him to be concerned throughout the 'Transcendental Deduction' with our cognitive processes specifically, whereas in fact many of Kant's claims in the first edition 'Transcendental Deduction' and most of his claims in the second edition 'Transcendental Deduction' explicitly abstract from one of the very important features of the working of our minds – namely that our cognitive processes are in time. Kitcher's view, that Kant is concerned with describing psychological processes throughout the 'Transcendental Deduction', leads her to reject the ideality of time on the grounds that it is inconsistent with Kant's theory of mind as a system of cognitive processes (because processes take time).

Brook, in his book, *Kant and the Mind*, also puts forward a positive Kantian theory of mind. I think he makes a mistake similar to Kitcher, though the mistake is less obvious in his work. Like Kitcher he takes Kant to be giving an account of how we must consider *our* minds in transcendental reflection. This is a mistake for two reasons: as I have earlier explained, the 'Transcendental Deduction' is a

---

<sup>77</sup> Kitcher (1990) p. 122.

consideration of the mind in abstraction both from our epistemic conditions and from any contingent feature of the mind. It describes some of the necessary functions of the mind when the mind is considered simply as the mind of a cognitive experience in general, and so the results of this consideration will not tell us everything about the functions of our mind considered under its empirical description. The two different considerations of the mind can be seen as standing in a relation of determinable to determinate, so the 'Transcendental Deduction' does not directly contain a theory of our minds considered empirically. However, transcendental apperception is relevant to the attempt to formulate a theory of mind empirically considered. Because transcendental apperception and a theory of our minds stand in a relation of determinable to determinate we can use the principle of transcendental apperception as a check on our theory of the human mind. If our favourite theory in cognitive science cannot be made to cohere with the principle of transcendental apperception, then something must be wrong with this theory of our minds.

## **2.j. Conclusion of Part 2**

Kant's claims about transcendental apperception do not constitute a theory of mind in the sense of making ontological and metaphysical claims about the mind as it is in itself. However, the principle of transcendental apperception does constitute a theory of mind in the sense that it describes some capacities (necessarily ascribed to mind), which are necessary for objective experience in general and which are therefore common to all minds, which have objective experience. It is a transcendental theory of mind, which describes how the mind must necessarily be represented given that we have objective experience.

Transcendental apperception thus stands in a relation of determinable to determinate to a theory of mind considered phenomenally, and is therefore of relevance to an empirical theory of mind because it can work as a check on it. If we were capable of formulating a theory of the mind considered noumenally the principle would also be of importance, but obviously we are not, and so the principle in itself does not violate the noumenal ignorance thesis because it does not make any positive claims about the noumenal mind.

### **3. Conclusion**

In the first part of this dissertation I have argued first, that in view of the assumption that the 'Transcendental Deduction' should be read as giving a regressive argument, the principle of transcendental apperception should be interpreted as being part of the conditions for objective experience, which Kant identifies in the 'Transcendental Deduction'. I have argued, secondly, that the principle of transcendental apperception is most importantly a principle of unity of consciousness, of which the claim that it must be possible for the 'I think' to accompany all my conscious representations, is an expression. The unity of consciousness is not deduced from the fact of I-thoughts, nor is it constituted by a collection of I-thoughts, because unless we have already established that the I-thoughts in question belong to the same experience, we have no grounds for claiming that they have the same reference. The unity of consciousness is rather shown by the synthesis of representations into more complex representations, and this is why Kant claims that the analytic unity of transcendental apperception is dependent on a synthetic unity – the unity of synthesised representations. However, the unity of consciousness is also what makes synthesis possible, as two representations can only be synthesised if they are united in some way. There is therefore a double dependence between unity and synthesis. Once the unity of consciousness has been established we can see that a weak identity claim follows from it. Within the unity of consciousness the subjects of different conscious representations are identical simply as the subjects of a unified consciousness. However, this is a weak identity claim, which cannot establish anything substantial about personal identity considered either noumenally or phenomenally.

I thus disagree with Allison and Henrich who see the identity claim as primitive and use it to deduce the unity of consciousness. I think Hurley has shown convincingly that this is not possible, and I think the text supports the view that the identity claim is something, which follows from the unity of consciousness, and not the other way around. I have also argued that transcendental apperception is a principle of spontaneity, first, because the addition of the 'I think' to any representation is not simply an expression of something which is already given in the content of the representation, but is added to it 'spontaneously'. Secondly, because the synthesis of representations into more complex representations is undetermined by the given it requires the spontaneity of the mind. Finally, I have

shown that the interpretation I give of transcendental apperception plays a role in at least two of the identifiable strategies for the deduction of the categories, namely in the objective deduction and in a version of the subjective deduction. I conclude that Kant's analysis of the necessary condition of experience have made it conceptually possible for us to talk about the mind considered transcendently in a way that was not open to us prior to the analysis. Because some of the conditions of experience, which he describes in the 'Transcendental Deduction', are necessarily ascribable to mind, since they describe mental capacities and functions, it is now open to us to talk about the mind considered transcendently. In the second part of my dissertation I have shown that the claims Kant makes in the 'Transcendental Deduction' can form the basis of a Kantian theory of mind. I argue that the 'Paralogisms' can be read as showing positively how the mind is necessarily represented in transcendental reflection given that it is the mind of the subject of an objective experience. Negatively, the 'Paralogisms' establish that the fact that we can show how the mind must necessarily be represented does not allow us to conclude anything about how the necessary capacities and functions of mind are realised. In fact whether or not they are realised by something else at all is not even something we can know. The 'Paralogisms' thus show that a transcendental theory of mind can establish simply the necessary features, functions, and capacities of the mind when we consider it in transcendental reflection, but nothing else. The two-aspects theory's explanation of the phenomena/noumena distinction further gives us a model on which to understand the status of a transcendental theory of mind. It shows us that having a transcendental theory of mind does not commit us to a third object distinct from the phenomenal mind and the noumenal mind, but rather that it commits us simply to a third way of considering the mind, a way given by how we think of the mind in a transcendental reflection of the necessary conditions of objective experience. Because the first part of the 'Transcendental Deduction' describes the conditions for objective experience in general, the transcendental consideration of the mind is meant to have a wider application than a theory of the mind phenomenally considered, for instance, which is specifically limited to describing the human mind under the conditions of human experience. If the first part of the 'Transcendental Deduction' really describes only the necessary conditions for objective experience in general, then its results should be valid for other minds



than our own – given that these minds also have objective experience. The results of the ‘Transcendental Deduction’ and the ‘Paralogisms’ are therefore relevant to other kinds of consideration of the mind, for instance they are relevant to the attempts by cognitive science to describe the mind. Finally, I have argued that though the transcendental consideration of mind does not commit us to the existence of a third object – the transcendental mind Strawson is not right in claiming that transcendental apperception is only ‘the imaginary subject of transcendental psychology’. In transcendental apperception we are confronted with something real both in the unity of consciousness and in our awareness of our own spontaneity.

I conclude that Kant’s principle of transcendental apperception does amount to a Kantian theory of mind, but a theory of a special kind. It amounts to a theory of how the mind must necessarily be thought of when we consider it via the mental capacities and functions, which are part of the necessary conditions of objective experience. It is a theory of mind considered transcendently and while it can tell us something real about the functions and capacities of the mind considered transcendently it cannot give us any information about how, or even whether, functions and capacities are realised.

#### 4. Bibliography:

- Allison, Henry: *Kant's Transcendental Idealism*. Yale University Press 1983
- Allison, Henry: 'Kant's Refutation of Materialism' (1989) in Allison, Henry: *Idealism and Freedom*. Cambridge University Press 1996
- Allison, Henry: 'On Naturalizing Kant's Transcendental Psychology' (1996) in Allison, Henry: *Idealism and Freedom*. Cambridge University Press 1996
- Allison, Henry E: 'Apperception and Analyticity in the B-deduction' from Allison, Henry E: *Idealism and Freedom*. Cambridge University Press 1996
- Ameriks, Karl: "Understanding Apperception Today" in Parrini: *Kant and Contemporary Epistemology*. Dordrecht 1994
- Ameriks, Karl: "From Kant to Frank: The Elimenable Subject" in Ameriks, K. and Sturma, D: *The Modern Subject. Conceptions of the Self in Classical German Philosophy*. Albany: State University of New York Press 1995
- Ameriks, Karl: "Kant and the Self – A Retrospective" in Klemm, D. and Zöllner, G: *Figuring the self: Subject, Absolute, and Others in Classical German Philosophy*. Albany: State University of New York Press 1997
- Ameriks, Karl: 'Kant's Transcendental Deduction as a Regressive Argument'. Klemm and Kuehn: *Immanuel Kant vol. II*. Ashgate and Dartmouth 1999
- Ameriks, Karl: *Kant's Theory of Mind*. Clarendon Press 2000
- Bennett, Jonathan: *Kant's Dialectic*. Cambridge University Press 1974
- Bird, Graham: *Kant's Theory of Knowledge – An outline of One Central Argument in the Critique of Pure Reason*. Routledge and Kegan Paul 1962
- Bird, Graham: 'Kantian Myths'. *Proceedings of the Aristotelian Society* 1996
- Brook, Andrew: *Kant and the Mind*. Cambridge University Press 1994
- Gardner, Sebastian: *Kant and the Critique of Pure Reason*. Routledge 1999
- Guyer, Paul: *Kant and the Claims to Knowledge*. Cambridge University Press 1987
- Henrich, Dieter: 'Identity and Objectivity' in Henrich, Dieter: *The Unity of Reason* Harvard University Press 1994
- Henrich: "The Identity of the Subject in the Transcendental Deduction", in Schaper and Vossenkuhl (eds.): *Reading Kant: New Perspectives on Transcendental Arguments and Critical Philosophy*. Basil Blackwell 1989

- Hurley, Susan: 'Kant on Spontaneity and the Myth of the Giving'. Proceedings of the Aristotelian Society 1994
- Kant, Immanuel: *Critique of Pure Reason* (translated by Kemp Smith) The Macmillan Press 1929 (second edition 1933)
- Keller, Pierre: *Kant and the Demands of Self-consciousness*. Cambridge University Press 1998
- Kitcher, Patricia: *Kant's Transcendental Psychology*. Oxford University Press 1990
- Matthews, H. E: 'Strawson and Transcendental Idealism' in: Walker, R. C. (ed.): *Kant on Pure Reason*. Oxford University Press 1982
- Pippin, Robert: 'Kant on the Spontaneity of Mind'. Canadian Journal of Philosophy 1987
- Powell, C. T: *Kant's Theory of Self-consciousness*. Clarendon Press 1990
- Rosenthal, David M: 'Two Concepts of Consciousness' in Rosenthal, David: *The Nature of Mind*. Oxford University Press 1991
- Sellars, Wilfrid: '...This I or he or it (the thing) which thinks...'. Proceeding and Addresses of the American Philosophical Association 1971
- Shoemaker, Sidney: *The First-Person Perspective and Other Essays*. Cambridge University Press 1996
- Strawson, P: *The Bounds of Sense*. (1966) Routledge 1995
- Thomas, Alan: "Kant, McDowell, and Consciousness" European Journal of Philosophy 1997
- Walker, R. C: *Kant*. Routledge and Kegan Paul 1978
- Yablo, Steven: 'Mental Causation'. The Philosophical Review, vol. 101, No 2. April 1992