KANT ON SELF-CONSCIOUSNESS

A thesis submitted to the Board of the Faculty of Arts of
University College London in partial fulfilment of the
requirements for the degree of Ph.D. in Philosophy

André Nilo Klaudat

University College London

July 1997
The purpose of this thesis is to provide an account of Kant’s view on self-consciousness. I give an answer, from the perspective of Kant’s philosophy, to the questions of what are the features of self-consciousness and how it participates in making knowledge possible.

I argue that Kant’s view on self-consciousness is developed along two lines of argument. In the first, self-consciousness is constrained by the account of how the categories necessarily agree with the objects of experience. Thinking - the manifestation of self-consciousness par excellence - is shown to be objective insofar as the categories establish a relation between representations and its objects. This point helps to explain the main feature of Kant’s view on self-consciousness: its self-sufficiency. Self-consciousness does not need, for its possibility, empirical criteria for the identification of a subject of experience.

The second line of argument concerns what conditions the categories in their work. Kant, in this respect, claims that transcendental apperception is the a priori ground of all concepts, and this issue involves the notion of a transcendental subject. I show that transcendental apperception, in the guise of the transcendental subject, is that ground by being what makes the performance-aspect of the movements of a corporeal agent possible. This agent must, however, be conceived as corporeal, and thus far his movements reveal the ability to develop techniques necessary for the judgement of independently existing objects.

The study is carried out against the background of Cartesianism and Classical Empiricism. The claims that emerge from Kant’s analysis of transcendental apperception in the two lines of argument permit one to clarify, finally, Kant’s point that his predecessors confuse self-consciousness with self-knowledge.
For Fabiana
ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Jerry Valberg for his guidance and encouragement during the course of this thesis. I am most thankful for his advice and useful criticism.

I would also like to thank Dr. Marcus Giaquinto and Prof. Malcolm Budd for their supervision at the early stages of this research.

Special thanks go to Marco Frangiotti for discussions of Kant, Strawson, Allison, and others. I also thank him for his moral support.

I am very grateful to Andrew Kirker, Tim Witherow and Fabiana Horn for proofreading this thesis, and for suggestions as to how to improve it.

I hereby duly acknowledge the financial support by CAPES, the Brazilian research support agency, which gave me the opportunity to embark on this project.
## CONTENTS

**Introduction**  
6

**I Self-Consciousness in Cartesianism**  
10

**II Self-Consciousness and Personal Identity in Empiricism**  
21

1. Locke  
2. Hume

**III Strawson on the Paralogisms**  
42

**IV Kant on Self-Consciousness: Context and Constraints**  
61

1. Introduction  
2. The Question about the Conditions of Representation  
3. The Problem as Concerning Intentionality  
4. The Autonomy of Meaning  
5. The Beginning of Kant's Solution to the Problem of Intentionality: Synthesis  
6. Kant's Idealism and the Transcendental View of Content  
7. An Illustration of the Transcendental View of Content in Contemporary Debate

**V Kant on Self-Consciousness: Self-Sufficiency and Externalism**  
113

1. Introduction  
2. Strawson's Account of Self-Consciousness: the Lack of Self-Sufficiency  
3. Kant's View of Self-Consciousness as Self-Sufficient  
4. Allison: Between Subjectivity and Objective Unity  
5. Henrich: the Deduction from the Identity of Self-Consciousness  
6. Kant's Externalism  
7. Self-Consciousness and Kant's Externalism

**VI Kant's Critique of Cartesianism and Empiricism**  
170

1. Introduction  
2. The Critique of Cartesianism  
3. The Critique of Empiricism  
4. Kant's distinctive Metaphysics of Mind  

**Bibliography**  
193
The present work is about Kant's conception of self-consciousness. It started with the belief that Kant had something really new to contribute to the discussion of this philosophical topic. There can be little doubt that self-consciousness is central to Kant's whole philosophical enterprise. Both his positive contributions - the deduction of the categories - and his negative contributions - the criticism of the paralogisms - are based on his understanding of the 'transcendental apperception', his phrase for self-consciousness. As Henrich remarks, this notion represents one of Kant's unexamined 'foundational thoughts', which receives attention only 'on account of what it achieves, not for its own sake'. Focusing an investigation on this conception promises to be rewarding in the attempt to understand Kant's philosophical achievements in general.

A further motivation for embarking on this research came from the study of Allison's work on Kant, which, in a way, determined the scope of my project. My interest in Kant on self-consciousness became an issue in what could be called Kant's metaphysics of mind. So, epistemological issues as such - scepticism, the external world - were not my concern. Allison played a role in shaping my interests as follows.

Among Kant's recent interpreters, Allison must be regarded as one who has done a great deal in fostering the cause of Critical Philosophy. His interpretation and 'defense' of Transcendental Idealism is an achievement any student of Kant has to take into account. It is Allison's decision to give central stage to the Kantian claims that constitute Transcendental Idealism and make it a general philosophical perspective which represents such an advance over the early outstanding achievements of Strawson. But not everything is clear in Allison. Crucial assumptions behind his explanations and interpretations of the transcendental idealist position are not without obscurities.

After saying that we have to distinguish sharply between an empirical and a transcendental version of two very important Kantian distinctions - ideality/reality and appearances/things in themselves - Allison presents the transcendental account of ideality.

He begins by reminding us that the transcendental level is the level of 'philosophical reflection upon experience (transcendental reflection)'; then, he goes on to say that 'ideality' at the transcendental level is used to characterize 'the universal, necessary, and, therefore, a priori conditions of human knowledge'\(^3\).

In the continuation, however, Allison presents what is intended as an explanation of transcendental ideality with respect to the a priori conditions of human sensibility: space and time are transcendentally ideal 'on the grounds that they function as... subjective conditions in terms of which alone the human mind is capable of receiving the data for thought or experience'. My dissatisfaction with Allison arises from an ambiguity about the status of this 'human mind'.

Allison had just explained what 'ideality' considered empirically means: it characterizes the private data of an individual mind, or 'any mental content in the ordinary sense of “mental”'. What he means to designate is the 'private domain of our own representations'. He then states the transcendental version of ideality I mentioned above. But, at the level of philosophical reflection upon experience, at the level where universal and necessary, that is, a priori, conditions are claimed to be fundamental, what can the import be of a reference to the 'human mind', to its (subjective) 'conditions'? Supposedly the empirically ideal items of our minds are the subject-matter of psychology, but what about the mind which is mentioned in transcendental or philosophical reflection? Allison speaks in another place of the a priori conditions as reflecting the 'structure of the cognitive apparatus'\(^4\). But then he links the mind which philosophy is interested in with universal and necessary conditions, rather than with the evolutionary and genetic approaches. So, are we dealing here with two minds, or two aspects of the same thing? Allison's way of characterizing transcendental ideality - with its reference to the human mind - having defined an empirical sense of mind as comprising private mental contents puts a suspicious gloss on his 'philosophical reflection'. It makes it almost impossible to resist the temptation to construe the universal and necessary (a priori) conditions as the most general psychological conditions. Allison perhaps approximates unnecessarily what should be kept apart as a matter of principle. The indefiniteness of his references to the mind at the

---

3. (1983), p.7. Quotations that follow are from pages 6-7 of this work.

transcendental level is due to his view of the mind in general, a view which I will argue is problematic and which can lead to misunderstandings about Kant's *transcendental* claims. Unfortunately, even Allison, who did so much to vindicate Transcendental Idealism as a metaphilosophical or methodological standpoint⁵, is in danger of fostering a misleading account of what the transcendental perspective in philosophy is. As a final example I would like to quote Allison's comment on an 'implicit' principle in the *Critique*: 'Whatever is necessary for the representation or experience of something as an object... must reflect the cognitive structure of the mind (its manner of representing)'. It is almost impossible not to give a psychologistic reading to this claim.

So, the present dissertation is about Kant's conception of self-consciousness paying special attention to Kant's *philosophical* enterprise. From my study, two claims can be attributed to Kant: firstly, that self-consciousness is self-sufficient, it does not need for its competent exercise to appeal to any reference to subjects given empirically in experience; and, secondly, self-consciousness plays a fundamental part in the conception of the transcendental subject, which has to be construed in externalist terms, *viz.* as the subject that is a corporeal agent capable of engaging in activities which possess a performance-aspect. I argue for these two general points using the following structure of exposition.

The first two chapters explore the background against which Kant had to present his views. These chapters have the limited aim of making clear what the fundamental tenets of Cartesianism and Classical Empiricism are concerning self-consciousness. In dealing with this task, two features became of interest in relation to Kant: Cartesianism's way of securing personal identity through our unfailing awareness of the very substances each of us is, and Empiricism's commitment - especially in Hume's case - to the distinctness and separation of 'perceptions'.

Chapter three concerns one of the most interesting approaches to Kant's fundamental conceptions, that of Strawson. This author was and still is a major influence on any attempt to understand what Kant means by his connection of the possibility of objectivity and the possibility of self-consciousness. Strawson's views on this issue also have an important influence outside Kant interpretation. I have chosen to address Strawson's

---

⁵ Cf. *op.cit.* p.25.

account in its presentation of Kant's negative contribution, *i.e.* the Paralogisms, because I believe there it is easier to see where Strawson distances himself from Kant, and what the real issues are. I argue that Strawson is still too committed to Empiricism to convey Kant's meaning.

Chapter four addresses the Kantian view that self-consciousness has objectivity. I argue that it is fundamental to appreciate that Kant himself faces this issue constrained by his explanation of how the understanding agrees with appearances necessarily. This is a point which determines, as a consequence, what a transcendental view of the content of our concepts and judgements can be. And this, in turn, helps to explain, at the beginning of chapter five, the self-sufficiency of self-consciousness. Chapter five is also dedicated to explaining what Kant means by the claim that transcendental apperception is the *a priori* ground of all concepts. I argue that Kant's position is externalist, that is, Kant appeals to the conception of a subject as a corporeal agent whose actions are immersed in a normative environment.

Chapter six examines, using the points made in the previous three chapters, Kant's critique of Cartesianism and Empiricism. I then return to the two main features I presented in chapters one and two as being of interest for Kant, and attempt to clarify further Kant's position and its differences with Cartesianism and Empiricism. I conclude chapter six and the dissertation with what I think has emerged in my study of self-consciousness as new in Kant's metaphysics of mind. I argue that Kant's conception of thinking, as distinct from experiencing, his insistence on the first person perspective and his metaphilosophical view of the limitation of the investigation are all distinctive of *critical* philosophy.

Finally, I would like to point out that references to the *Critique* are made in the main text simply by 'A' or 'B', for the first and second editions respectively, followed by the page number in the original German edition. In addition, I make reference to the Academy edition of Kant's works in one of two ways. Sometimes I use 'Ak' followed by the number of the volume given by a roman numeral, giving the page reference with an arabic numeral. On a few occasions I make references using 'p.Ak283', which refers to page 283 in the Academy edition of the work. In these cases the volume in which the work is to be found has already been mentioned or is well-known.
CHAPTER I

SELF-CONSCIOUSNESS IN CARTESIANISM

Rational Psychology was at the time of Kant a metaphysical discipline pursuing as its ultimate goal the proof of the soul’s immortality. Kant in the first *Critique* mentions, in relation to Rational Psychology or Metaphysical Pneumatology, only Moses Mendelssohn’s attempt to prove the permanence of the soul, that is, the soul’s absolute continuity beyond this life (cf. B413-26). But the discipline had other famous exponents at the time in Christian Wolff, Alexander Gottlieb Baumgarten and Christian Crusius\(^1\). Its basic arguments, however, did not originate with those metaphysicians, but are part of a tradition which goes back through Leibniz at least to Descartes.

We can call ‘Cartesianism’ the outlook derived from the theses advanced by Rational Psychology. In the *Critique* Kant presents them as establishing that the soul is *substance*; then, from its substantiality, that it is nonetheless *immaterial*; furthermore, that because it is a *simple* immaterial substance it is *incorruptible*; and that because it is identical to itself through time it carries *personality*, that is, continuity of personhood; and from all this that it is *spiritual*, which because of its *commercium* with bodies is the principle of life in matter, that is, the ground of *animality*; the latter when properly limited by spirituality produces the concept of the soul’s *immortality* (cf. A345/B403). Cartesianism seems to have most of its core concepts in the writings of Descartes himself. The *Meditations*, second edition (1642), seek to prove, according to the subtitle, the existence of God and the distinction of the soul from the body. The first edition (1641), however, had in its subtitle the promise to prove the existence of God and the *immortality of the soul*. The thesis of immortality is never considered in the *Meditations*, but quite clearly the real distinction would have been a necessary preliminary had Descartes wanted to prove the thesis there\(^2\).

---


My purpose is to identify the views about self-consciousness in Cartesianism as one particular philosophical approach to these matters, concentrating on Descartes' position, especially in the Second Meditation. I will confine myself to the presentation of the nature of Descartes' claims, of their meaning, leaving their criticism to much later on, in fact to chapter six. It is important first to get clear about the **real** central points of Cartesianism, because it is in contrast to this position that Kant presents his view.

By registering that there are many illusory sense perceptions Descartes was led to consider doubting the senses as a whole. But soon he realized that particular errors of sense perception were not enough to lead him to doubt his sitting by the fire, wearing a winter cloak, holding a paper in his hands, etc. He then considered the possibility of dreaming exactly those things. He claimed that we all sometimes have in dreams the strong conviction of being somewhere, doing something, etc, when in fact we are only dreaming those things. Descartes' argument here depends, firstly, on there not being a criterion to distinguish being asleep from being awake from the sole point of view of what are normally taken as the contents of what is seen, heard, touched, etc. But, secondly, for the 'dream-argument', as it is known, to be plausible, Descartes must be committed to the view that there really is no way of answering appropriately the question 'Am I awake or am I dreaming?' Descartes certainly believed both things, consequently he considered as the only adequate retort the one concerning 'simple natures'.

Things in sleep are not what they appear, but they cannot appear like that if they were not formed in the likeness of real objects. Dreamers, and painters, can create sirens and satyrs, but not with wholly new natures: they have to use colours, represent their creations as corporeal (that is, as extended, shaped, sized objects), these objects have to be represented as in a certain number, as being in a certain place, as enduring through a certain time. Descartes consequently agreed with this retort to the effect that simple natures are real. But his reading of this situation is quite unique, and in his hands 'real' assumes a rather special sense.

---

3. Cf. *Meditations on First Philosophy* in: Anscombe & Geach (1970), pp.59-124. What follows immediately below in the main text is a summary of the First Meditation, all the points are to be found in *op.cit.*, pp.61-5. When talking about the Second Meditation, I will refer to this edition by pages in brackets only.

Simple natures are real, according to Descartes; but because they are not sufficient to distinguish, by their mere presence, sleep from waking life, they cannot be used to claim that the senses as a whole do not deceive us. For Descartes this means that Physics, Astronomy, Medicine and all the sciences depending on the consideration of composite objects (as these are the only ones given through the senses) are doubtful; whereas Arithmetic, Geometry, etc, the proper sciences of the simple natures, whether they really exist in nature or not, are certain. Anthony Kenny explains that simple natures can be studied in their necessary connections independently of whether they exist or not because existence is itself a simple nature. So, given that for a simple nature to exist it must be combined with another simple nature (i.e. existence), and because this combination is contingent for all things except God, a science that studies necessary connections will not study things that are composite\(^5\). Simple natures are real, then, independently of whether they exist or not, and this is enough to make us doubt about the objects in nature, but not whether two plus three make five, or whether a square has only four sides.

To doubt this kind of claim to certainty Descartes introduced the possibility of an omnipotent God deceiving him into believing those things. However, because of theological worries about God, supposedly good, deceiving him, he immediately changed the possibility to one of an evil spirit, which is supremely powerful and intelligent, and does his utmost to deceive Descartes. So, even two plus three making five is not certain any more, and that is the situation of inextricable darkness in which Descartes finds himself at the end of the First Meditation. It is interesting to notice that the rationale for the hypothesis of the evil spirit, from the Cartesian point of view itself, is the thesis that eternal truths, like those of Mathematics, are created, which is in sharp contrast to the view that not even God could make the mathematical propositions in question false and so avoid our being certain about them\(^6\). This is an indication that the Cartesian doubt is really not so radical, and makes one wonder how much else is not assumed without further questioning in the procedure of doubting.

Descartes starts the Second Meditation pitying his situation. He recapitulates the reasoning that put him there and then realizes that he was and is convinced by the reasoning,

---


and that this certainly requires at least that he must have existed, and insofar as he continues to believe so, he exists. 'But if I did convince myself of anything, I must have existed' (p.67). This is the closest the Second Meditation comes to the famous formulation of the Discourse, i.e. 'I am thinking, therefore I exist'. In Latin it runs cogito, ergo sum, hence the cogito argument.

Descartes' view is that 'I am', or 'I exist', whenever uttered or conceived is necessarily true (cf.p.67). This seems to support the interpretation that the cogito argument has the nature of a recognition that such utterances have a self-confirmatory character. This would imply that the cogito argument is not really an argument, and 'cogito' is not really a premise from which existence is derived. But the progress in the Second Meditation gives the Cartesian view about 'I am' or 'I exist' a specific import. They are necessarily true whenever they are thought. The position is actually the conclusion from the reasoning that if the evil spirit deceives me at least the object of deception must exist. This reasoning requires that the thoughts I entertain, whether true or not, are immediately present to me, that is, are indubitably given to me. Descartes doubts, and convinces himself that nothing in the world exists, but about his convincing himself there is no possible doubt at all. So the reasoning in the Second Meditation is 'I have convinced myself that..., but if I did convince myself of anything, I must have existed'. The necessary truth about existence, that whenever one thinks one must exist, derives its strength from the certainty that when one thinks, one simply knows that, and of what, one thinks.

Now Descartes asserts that at the time of thinking he is something and not nothing (cf.p.67). A bit later he equates 'really to exist' with 'being a real being' (cf.p.69). He talks about the 'I' by means of expressions like 'this "I"', 'the "I"', and admonishes himself not rashly to 'take something else for the "I"' (cf.p.67). What I want to call attention to is the fact that Descartes infers his existence as the existence of a thing, a substance. In the Principles, Descartes, having defined 'thought' as 'everything which we are aware of as


happening within us, in so far as we have awareness of it', proceeds to argue that the mind is better known than the body, and in the process he states how he sees the situation: 'Whenever we find some attribute or qualities [as our thought, for example, when we judge that the earth exists], there is necessarily some thing or substance to be found for them to belong to\(^9\). To make it clear that we are dealing here with an *inference* for the existence of a substance it is appropriate to remind ourselves of Descartes’ retort in the Fourth Replies: 'We do not have immediate awareness of substances... We know them only by perceiving certain forms or attributes which must inhere in something if they are to exist; and we call the thing in which they inhere a “substance”’\(^10\).

It is this inferred substance that Descartes calls ‘I’ and takes to be *himself*, and is willing to know more about, to understand better. He even asks: ‘What am I?’ (cf. p.67). From this it is safe to conclude that ‘I’ and ‘self’ for Descartes are expressions which refer to a thing, a substance. From this point on Descartes calls, without any hesitation, this substance he is going to research an *ego*. One could ask, however, what right Descartes has to use ‘I’ or ‘self’ to designate this substance? This is to ask why Descartes should think that ‘I’ or ‘self’ are used to refer to this substance of which we up to this point do not know the nature.

We can see an answer to this question in the Cartesian reasoning that leads him to assert that he is ‘a conscious being’. Descartes, firstly, reaffirms his unshakeable commitment to the ‘*cogito*’ premise. It is indubitable that I think, or it is a ‘fact that there is consciousness (or experience: *cogitatio*)’ (p.69). From this he then jumps to ‘of this and this only I cannot be deprived. I am, I exist; that is certain’ (loc.cit.). So, from the certainty of the ‘*cogito*’, that there is consciousness or experience, he goes to answering ‘What am I?’, assuming without examination that the substance which is the *substratum* of the thinking is an ego-substance. As I am sure that there is consciousness, the thing that exists is a *conscious being*! It is quite clear that the ‘I’ comes in because of Descartes’ conception of the fact that there is consciousness. To assert that is to assert that there are *activities* (thinking, doubting, etc) and *states* (being convinced, etc) which in requiring a substance of

\(^9\) *Principles of Philosophy* in: Cottingham et al. (1985), pp. 195, 196, respectively, or AT VII A 7, 8.

inherence actually require an ego. To the question ‘What is a conscious being?’ (p.70) Descartes answers ‘A being that doubts, understands, asserts, is willing, is unwilling, further, that has sense and imagination’ (loc.cit.). These are normally taken to be the responsibilities of a human mind, a self. As Descartes says: ‘I am’ precisely taken refers only to a conscious being; that is a mind, a soul (animus), an intellect, a reason’ (p.69)\(^\text{11}\).

This helps us to understand how the first major feature of Cartesianism is arrived at: the ‘I’ is a referential expression. The reason is that cogitatio involves activities, states, which can only be ascribed to a subject, a substance. Now because these activities and states are mental, the substance is cogitans. As the definition of thought mentioned above implies, thoughts, as happenings within us of which we are immediately aware, require that there is a basis for the happenings, and this is the mind or self referred to by ‘I’, ‘self’, ‘myself’, etc.

But Descartes moves on: ‘What now? I will use my imagination’ (p.69). The next problem is, given that there are doubts, understandings, assertions, denials, wishes, etc, that ‘these are a good many properties’, how to guarantee that they apply to Descartes’ self. ‘If only they all belong to me’ he muses wishfully (cf.p.70). This is a question of identity. Descartes wants to show how we can be certain of personal identities. The answer is ingenious.

To answer this question Descartes builds on his view of cogitatio and of the ‘I’. He puts it very rashly that the mentioned properties simply cannot fail to belong to me. He rhetorically asks himself: ‘But how can they fail to? Am I not the very person who is now “doubting” almost everything; who is “willing” to know more, and is “unwilling” to be deceived; who “imagines” many things, even involuntarily, and perceives many things coming as it were from the “senses”?’ (p.70). This is a ‘who’ question, and the answer is that in thinking or undergoing cognitiones I, as it were, infallibly meet with the person who I am. This answer is presented not only as a conception of persons, of what they are - they are minds, souls, etc - but also as a view of how the numerical identity of these substances is to be established, that is, how number is to be determined. When I think my thoughts I cannot fail to meet in the process the thing, the person, that I am. Descartes thinks he has his identity as a person dependent only on the workings of the ‘I’ in his doubtings.

understandings, etc, and the ultimate ground for his viewing the ‘I’ as working to guarantee that these doubtings, understandings, etc, give the identity of the thinking thing that he is, is that he believes that he knows without any possibility of error that he is the person who is now doubting, understanding, etc. This seems so straightforward for him because he takes the knowledge that, and of what, he is thinking to be indubitable. We could represent Descartes’ reasoning as follows: I know that and what I am thinking; this requires that this thinking is the activity of a thinking substance; and this implies that when I know that and what I am thinking, I also know that it is I, the thinking substance that underpins my thoughts, who really does the thinking; so in knowing that I am thinking I know that it is I as a person, the person I am, that is doing the thinking. According to Descartes I always ‘meet’ the person I am, ‘my very self’ (p.71), when I think something. In thinking ‘I am aware of’ it (loc.cit.). Descartes envisages backing this account of what is really best known to each of us (our very selves) with his analysis of how we establish the identity of a piece of wax. I shall present this very briefly.

A piece of wax has many sensory features: colour, shape, smell, etc. Put by the fire it changes completely, it loses its colour, form, fragrance, etc. How do we establish that it is the same wax that is still there? Descartes’ view is that our knowledge of what this wax is is really not provided by the senses, because all that I got through the senses about the wax has changed but the wax is still there (cf.p.72). The consideration of what is properly attributable to the wax, that it is something extended, flexible and changeable, is also not something that could be established by the senses or by the imagination. Basically, because its extension also changes, grows greater with increasing heat (cf.pp.72-3). Descartes concludes: ‘It remains then for me to admit that I know the nature even of this piece of wax not by imagination [here he could have mentioned also the senses], but by purely mental perception’ (p.73)\(^\text{12}\).

\(^{12}\) It would be interesting to compare Descartes’ view on the relation between the identity of objects and the senses with Kant’s position about these identities and ‘outer sense’. It is remarkable how Descartes plays down the senses as a source of knowledge: ‘It must be observed that perception of the wax is not sight, not touch, not imagination; nor was it ever so, though it formerly seemed to be; it is a purely mental contemplation (inspectio); which may be either imperfect or confused, as it originally was, or clear and distinct, as it now is, according to my degree of attention to what it consists in’ (p.73). Kenny (1968, p.71) remarks that Descartes identifies sensation, the actual sensation or consciousness of seeing, with thought, and thus alters the meaning of sensus. For Descartes even when one sees an object one actually thinks one sees an object. The first thing to notice in relation to Kant’s position is that for him there is no assimilation of the sensibility to the understanding (cf. A50/B74 and ff., also the Amphiboly, A260/B316-A292/B349).
So, Descartes is aware that in dealing with a piece of wax he is dealing with an extended, flexible and changeable thing, but his view is that nonetheless identity can only be established by a purely mental contemplation orientated by judgement, by the intellect, not by the senses or the imagination. Only through looking at how we understand things, not how we see or touch them, are we going to discover the naked real self of the wax (cf.pp.73-4).

Now, the moral Descartes extracts from this story is that the intellect is even more essential for establishing the identity of the mental thing each of us is. If to establish the identity of the piece of wax through its changes (through time) the intellect or human mind is needed in the form of a purely mental perception or contemplation, it is only natural to assert that the intellect or mind is even more needed to establish its own identity as the substance which is the *substratum* of all his thinkings. Descartes presents this point as the thesis that the mind is better known than the body (cf.pp.74-5). But the thesis is in this context an answer to a question of identity. A human mind is needed to perceive the wax in its identical naked self, but when Descartes does just that, what are the conclusions to be drawn about his mind, about himself? Descartes answers saying that if he judges about the wax, it is much clearer that he himself exists as an identical self because he sees the identical wax. He then concludes that he must surely be aware of himself as an identical self, 'not only much more truly and certainly, but also much more distinctly and manifestly' (p.74). He also writes: 'I thus clearly recognise that nothing is more easily or manifestly perceptible to me than my own mind' (p.75, my italics: AK)\(^\text{13}\).

This gives us the second major feature of Cartesianism, *i.e.* the way it guarantees personal identity\(^\text{14}\). I will summarize this very briefly. I judge things about the wax because I see it, but much clearer than this is that I exist in a determined identical form because I see the identical wax. Descartes' argument concerns the immediacy in which mental properties stand to my mind, so that actually they could not be things distinct from my consciousness.

\(^\text{13}\) My point is *not* that the 'I' of the *Meditations* stands for the man René Descartes, rather that his view is that identical personhood can be met with by each of us in thinking and using 'I'. Cf. Wilson (1978): pp.4-5.

\(^\text{14}\) This point escapes Wilson, cf. 1978, pp.59-60, 75-6. Although this author notices Descartes' worries about the 'continuing identical entity' being referred to by 'I', especially clearly in the 'Synopsis', she does not say anything about how Descartes thinks he can succeed in warranting this identity. For a view like mine cf. Mendus (1984), pp.135, 138.
That I doubt, understand, etc, is no less of a fact for Descartes than my own existence (cf.pp.70-1). This accounts for the incorrigible description I can give of the state of my mind, and of my experience according to Cartesianism. That I see, hear, etc, may be false, but not that I seem to see, to hear, etc (cf.p.71). Consequently, there is no way in which I could fail to meet the person I am when I think, doubt, etc, as being in those states. This for Descartes is sufficient to establish not only the mental nature of my being but also its personhood.

The conclusion I want to draw from this is that for Cartesianism, because of the awareness I have of my very self in the process of doing things like thinking, doubting, etc, of which I am incorrigibly certain, self-consciousness (that I think, etc) is knowledge of the self (as a thinking thing independent of bodies): knowledge of the true nature of the self and its identity as a being of this nature. So, the Cartesian model of self-consciousness is that of the consciousness of a thing mediated however by the consciousness of it as being in states of thinking, doubting, etc. Properly understood, this is knowledge of the thing as being in those states. This exhausts what we need to know about the self's true metaphysical nature. Thus far, I side with those who think that all the materials which Descartes uses to conclude that he is not a body, but rather a soul, are already given with the Second Meditation. The real distinction established by the Sixth Meditation does the job of validating as clear and distinct the claim about that distinction, but we already know at the time of the Second Meditation what our real nature is, and this is achieved by assuming that what the dream-argument establishes is true, that is, that there are no external objects, including my own body\(^\text{15}\). (This is certainly also implied by the way the evil spirit hypothesis is introduced. As we saw above, simple natures did not even have to exist for them to be certain, which implies that extensionality can be studied without there being anything extended). But not only do we know about our nature, we also know the way 'I' refers to one's own self, a way which implies at least that 'I' refers to something distinct from what 'you' refers to, or so Descartes believes\(^\text{16}\).

\(^{15}\) On the debate about what is in fact established by the Second Meditation and what needs to be supported by the Sixth, see Kenny (1968), pp.62, 90-5, also Williams (1978), chapter 4, pp.102-29.

I want to finish this summary of Cartesianism on self-consciousness by calling attention to a difficulty concerning the knowledge of the self which that consciousness constitutes. It is not uncommon to see such knowledge being represented as introspective. Armstrong presents Descartes as holding the doctrine of 'self-intimation': 'everything in a mind at a time is available to consciousness, introspectively available, at that time'\(^{17}\). According to Armstrong, this means that in Cartesianism nothing escapes inner sense, and more, the introspective awareness is infallible. The deliverance of inner sense in Cartesianism is self-evident. This author then sees in the Cartesian tradition the view of self-consciousness which is self-knowledge as stating that 'the current contents of the mind all intimate themselves to an infallible faculty'\(^{18}\).

Such a completely different philosopher as Hacker also goes down this path and characterizes Cartesianism as involving claims to the effect that (1) if someone has experience, then he knows he does; (2) one's knowledge of one's experience is certain; (3) this knowledge is incorrigible. Now, because in Cartesianism to have experiences is to be in a relation to activities, states, etc, a relation which produces knowledge, there must be an inner equivalent of perception. This would be inner sense, introspection, consciousness or awareness. This is the view that one's mind is transparent to oneself. The activities, states, etc, are immediately known by introspection\(^{19}\).

The difficulty I want to point out is that, as we saw, Descartes says nothing about inner sense in this context. Actually, from the analysis of the piece of wax, what turns out to be crucial is a purely mental contemplation oriented by judgement, by the intellect only. Nothing of help comes from the senses. But on the other hand, those authors seem to have a point, because Descartes also calls this contemplation a purely mental perception. The truth is that this difficulty is beyond a definite answer, simply because of Descartes' deformation beyond recognition of 'sensus' in general. As I very briefly indicated above in note 12, even the act of seeing for Descartes is thinking one sees. His claim is that, from the point of view of their 'formal realities', all thoughts have the same status, they are all modes or states of

---


my consciousness, and 'sensus' does not distinguish anything\textsuperscript{20}. So, that it is an inner sense does not matter either, given that for Descartes it is all in the scope of the mind simpliciter.

\textsuperscript{20} Cf. 'Third Meditation' in: Anscombe & Geach (1970), pp.81-1, where 'formal reality' is rendered 'inherent reality'.

CHAPTER II

SELF-CONSCIOUSNESS AND PERSONAL IDENTITY IN EMPIRICISM

It certainly sounds odd to suggest that Locke and Hume, the great patrons of Empiricism, could have helped to foster the cause of Cartesianism. While they are not, of course, to be ranked as Rational Psychologists, and Kant never refers to them in those terms, nevertheless the fundamentals in their philosophical approaches are clearly informed to a large extent by Cartesianism. In a sense then, what Kant has to say against Cartesianism will affect or can also be targeted at Locke and Hume. In this chapter I will present the empiricist view, discussing Locke and Hume together insofar as their similarities permit. My aim is to extract what is crucial for their viewpoints.

1. Locke

Empiricism begins with a belief about the origin of ideas and a generalized suspicion about metaphysical claims. The first empiricist contribution to the matters of my concern is the attack on the claim concerning the clearness and distinctness of our idea of substance as Descartes understood it. Both Locke and Hume maintain that we have no such idea, that the only idea we have is that of a collection of simple ideas or particular qualities\(^1\). Now, both also claim that it does not matter what kind of substance we should be investigating. Locke thinks that our position is the same whether we try to understand the substance belonging to the spirit or to the body; thinking is as clear as extension; and communication of motion by thought, normally attributed to the spirit, is as evident as impulses we ascribe to bodies\(^2\).

This first standpoint of Empiricism constitutes a major blow to the Cartesian metaphysical claim about the soul. But it also creates a problem which Locke in particular was to feel acutely. Cartesianism proceeds from the immaterial substantiality of the soul to its indivisibility, which constitutes the bridge it then employs to get the proof of the soul’s

---

1. Cf. Locke (1979), p.305; and Hume (1978), p.16. Locke’s *Essay* and Hume’s *Treatise* will be referred to, respectively, as E and T.

immortality. Now, Locke gives up the claim to substantiality, and with it all the benefits he could earn from adhering to it. He feels the necessity of producing an alternative account of how our minds persist through time, and how we have personal identities as we believe we do. This problem of personal identity was of central importance for Locke because of its practical and ethical significance. Having abandoned the Cartesian substantial identity, he had to find a sound foundation for 'all the right and justice of reward and punishment', and he sought to achieve that through his account of how we have personal identity.

The problem with Locke’s solution is that while departing from Cartesianism with his denial about substance he still adheres to the basic philosophical assumptions precisely characteristic of that outlook. This has to do with an unexamined adherence to the Cartesian theory of ideas. In Locke’s case it produces a solution that will appeal to a quasi-substantial entity. In Hume, who inherits the problem, we will find a radicalization of the Lockean position which brings to the fore both the empiricist stance and the Cartesian commitment.

Empiricism’s first move created the task of giving an alternative account of persons. Empiricism faced this problem by examining how we ordinarily think about ourselves and, by attacking the difficulty in this way, it was able to clarify an ambiguity which plagued Descartes’ account. Locke tackled the problem with the tools of our ideas of identity and diversity, since it is these which make us think about certain possibilities in a determined manner. A thing that we see to be in a certain place at a certain time is not the same thing which is at the same time in another place. Two things of the same kind do not exist in the same place at the same time, etc. Now, we think about ourselves as a thing which persists through time and is identical with itself, as a particular. This way of facing the problem of persons makes it clear that what is being sought is not something which has the nature of a kind or type; rather, because we think we exist, and exist as particulars, what is being sought is the identity of particulars, that is, something that allows us to talk of number. This is why Locke introduces the discussion under the heading of the principium individuationis and

3. Cf. Locke, E, p.341 for the point about the foundation; cf. also Allison (1966) passim for the practical side of Locke’s worries.


5. Cf. Locke, E, p.328; also Hume, T, pp.253-9 for the same strategy of starting with how we ‘feign’ the identity of ordinary objects.
The task, then, is to discover what the concept of ourselves is given such an unassailable belief in our particularity.

The first remarkable conclusion Locke produces is to claim that what we say about the identity of simple and compounded substances, that is, of atoms and masses, and what we say of living creatures like animals (or brutes) and vegetables is also, with some minor qualifications, what we say of the identity of the same man. The identity we ascribe to the same man is surely of the same sort we apply to animals, or indeed to vegetables, and not to masses, precisely insofar as the important principle informing the ascription is one about a common life, which is understood as the motion - of parts conjoined in an organization - coming from within. ‘The Identity of the same Man consists in nothing but a participation of the same continued Life, by constantly fleeting Particles of Matter, in succession vitally united to the same organized Body’.

It is interesting to notice that the topic of the identity of man was not of major concern to Locke. He only insisted upon the having of a body with a certain shape. This, however, enabled him to refuse two very unpalatable options. The first construed the identity of man as the identity of soul, and the second accounted for the identity of man by appeal to intelligence and rationality.

Locke, then, changes the subject of research. He abandons the idea of man and starts to enquire into that of person. In paragraphs nine, and those following, of the chapter on personal identity he distinguishes ‘person’ from ‘man’, but he also brings into play a further distinction he had made earlier (paragraph seven): ‘person’ and ‘man’ are both distinct from ‘substance’. This picture is not easy to define.

---


7. Cf. E, pp.331-2; also Hume, T, pp.255-8 for the identity of masses, mountains, planets, ships, churches, rivers, animals and vegetables, with life being a sympathy of parts to a common end and the reciprocal causal relations between the parts. The identity of a man is affirmed to be of the same sort as that of living creatures.

8. Cf. E, pp.332-5 for these points. Cf. Hume, T p.257 for how the identity of man remains even if some details of shape change: the same individual can change from a boy either to a fat or to a lean man. But Hume just assumes that the human shape remains.

9. Cf. Allison (1966), p.530 for the implicit recognition on Locke’s part that, because of man being accountable in terms essentially corporeal, this was not the way forward in the pursuit of personality as the basis for moral responsibility.
‘Person’, according to Locke, ‘is a thinking intelligent Being, that has reason and reflection, and can consider it self as it self, the same thinking thing in different times and places; which it does only by that consciousness, which is inseparable from thinking, and as it seems to me essential to it...’. The crucial notions of this account are those of ‘thinking thing’ and ‘consciousness’. In explaining them Locke at once departs from Cartesianism and maintains one of Cartesianism’s fundamental claims. Locke dispels from his account the idea of sameness of substance; but then in explaining the identity of persons he uses a Cartesian view of consciousness, of thinking.

About this aspect of Locke’s views, Allison remarks that it constitutes the adoption of ‘the Cartesian conception of consciousness as necessarily entailing self-consciousness’. I think this is potentially misleading because Locke is actually situated midway between Cartesianism and Hume’s full-blown Empiricism about the self. Allison’s remark brings Locke closer to Cartesianism than is really the case. My view here depends on how the self was seen in Descartes, and consequently what could be understood by ‘self-consciousness’ from that point of view.

According to Allison, Locke’s view of consciousness, in requiring that it is ‘impossible to perceive, without perceiving, that [one] does perceive’, that is, ‘when we see, hear, smell, taste, feel, meditate, or will any thing, we know that we do so’, imports an immediate self-awareness through which ‘every one is to himself, that which he calls self...’. But this is not straightforwardly Cartesian. As we saw in chapter one, for Descartes there is no immediate awareness of a substance, we only know it through its attributes, which are conceived to inhere in it. Self-awareness or self-consciousness in Descartes is a mediate awareness of self through its attributes. But more importantly, it is awareness or consciousness of a thing, a substance. Locke sees both things differently.

It is important to point out that for Locke this consciousness is not of a substance, nor is the knowledge mediate knowledge of the substance-self. Locke observes that it is that consciousness, self-consciousness actually, which makes everyone to be to himself what he calls self, but that does not determine whether the same self so presented continues in the

11. Cf. op.cit., p.530, my italics: AK.
same or in diverse substances. Furthermore, the immediacy involved in this knowledge which is our self-consciousness is of the same kind as the one which accompanies our present sensations and perceptions\textsuperscript{13}.

Let me return to Locke's two crucial notions. Self-consciousness is what always accompanies thinking and what constitutes a self, but it is also what distinguishes oneself 'from all other thinking things', and what alone constitutes 'personal identity, \textit{i.e.} the sameness of a rational Being'\textsuperscript{14}. So, for Locke, 'thinking thing' stands for \textit{a} rational being, distinguishable from all others, which is nonetheless not to be identified with any particular kind of substance. This thinking thing or rational being, 'independent-from-substance', is what Locke understands by a self or a person. Exactly because of this anti-Cartesian stance about 'thinking thing' Locke is not going to equate consciousness with self-consciousness as Allison believes. This equation is asserted by Locke, but it means something different from what Descartes meant. In spite of adhering to the Cartesian conception of consciousness, Locke has a different conception of self-consciousness. I will return to this later on.

The second crucial notion in Locke's account is that of consciousness, of which I said something already. Consciousness is that which is inseparable from and essential to thinking, but it is also that which is able to give identity to the self as \textit{a} self, that is, as 'the same thinking thing \textit{in different times and places}', and which 'distinguishes [a self] from all other thinking things'\textsuperscript{15}. But how does consciousness do that? Locke's answer is: 'As far as this consciousness can be extended backwards to any past Action or Thought, so far reaches the Identity of that \textit{Person}, it is the same \textit{self} now it was then; and 'tis by the same \textit{self} with this present one that now reflects on it, that that Action was done'\textsuperscript{16}. Locke talks in this context of a \textit{presence} to the mind, but also of what hinders the workings of this consciousness, \textit{i.e.} 'forgetfulness', and consequently ponders that even the best 'memories' lose sight of part of their past actions. Paragraph ten is dedicated to this subject\textsuperscript{17}. There are

\begin{itemize}
\item \textsuperscript{13} Cf. Locke, E, p.335.
\item \textsuperscript{14} \textit{Loc.cit.}.
\item \textsuperscript{15} \textit{Loc.cit.}, my italics: AK.
\item \textsuperscript{16} \textit{Ibidem}.
\item \textsuperscript{17} Cf. E, pp.335-6: 'Consciousness makes Personal Identity'
\end{itemize}
two interpretative lines that could be taken here. I will mention the first and hope that by taking into account considerations, which it tries to minimise the importance of, one becomes convinced that Locke’s ‘true’ position is that followed by the second line.

The first line would agree that Locke links the workings of that consciousness with memory, and then that the recognition that memory fails implies ‘the losing of sight of our past selves’, but this, according to this line of thought, backed by Locke’s very words, does not concern personal identity, rather it concerns the problem of sameness of substance. The crucial issue of ‘same person’ is not being addressed by this point, because it does not matter whether there is sameness of substance, as long as that ‘same consciousness that makes a Man be himself to himself’ is present to form one person. What Locke would have in mind here would be a point about ‘self-consciousness in time’, that is, as Allison says, ‘that I can not intelligibly doubt... my identity as the enduring subject of a determining set of representations or mental states; for example, as the subject who has a blue percept at t₁, followed by a white percept and a pain sensation at t₂’. That means, ‘insofar as I am conscious of such a succession, for example, a headache followed by a white percept, I am aware of them both as pertaining to my numerically identical self’\(^{18}\). Locke writes exactly in this vein: ‘For as far as any intelligent Being can repeat the Idea of any past Action with the same consciousness it had of it at first, and with the same consciousness it has of any present Action; so far it is the same personal self’\(^{19}\). Allison uses such a line of thought to silence the sceptical challenge that the identity of a present subject of experience with a past one is only inferred and dependent on memory, that is, that identity cannot be guaranteed this way\(^{20}\). It seems that it is exactly this challenge that Locke has in mind when he discredits the relevance of the fact that memory fails as a point about sameness of substance and not person. This may be so, and it certainly sounds interesting, but more depends on consciousness. Locke makes personal, in the sense of numerical, identity depend on consciousness, and he has a very concrete conception of its workings. Memory will guarantee the identity of person. This is the thrust of the second line of interpretation of Locke’s claims. It is at this point that what Locke says about consciousness, or memory, will


\(^{19}\) E, p.336.

be Cartesian but will also end up producing the distinctive empiricist view of self-
consciousness.

The result of Locke’s view of ‘thinking thing’ is to have produced a distinction he is
very keen on, that between a same substance and a same person. This separation leads to the
questions he formulates in paragraph twelve and tackles in the continuation, \textit{i.e.} what
happens to the sameness of the one if the other changes and vice-versa. The development
that interests me is that which gives memory central stage. Instead of a straight answer to the
question of whether sameness of substance is compatible with distinction of persons, Locke
prefers to point out what makes individuality, sameness of person. Against the believers in
transmigration of souls, Locke makes the common-sensical point that it does not matter
whether one ‘has in himself an immaterial Spirit, which is that which thinks in him, and in
the constant change of his body keeps him the same, and is that which he calls himself
[notice that here Locke allows his interlocutor the Cartesian view]: Let him also suppose it
to be the same Soul, that was in \textit{Nestor} or \textit{Thersites}, at the Siege of \textit{Troy},...which it may
have been, as well as it is now, the soul of any other Man: But he, now having no
consciousness of any of the Actions either of \textit{Nestor or Thersites}, does, or can he, conceive
himself the same Person with either of them? \textsuperscript{21} This rhetorical question reveals Locke’s
position: the criterion is memory. Let a man have the soul of Socrates, but if he is ‘not
conscious of any of \textit{Socrates’} Actions or Thoughts\textsuperscript{22} he could not be the same person as
Socrates. Locke admits the possibility of numerical identity of souls across human bodies,
but he denies that this implies sameness of person, for the same immaterial substance
without sameness of consciousness or \textit{of memories} does not make for sameness of person.
Locke’s view implies that if someone finds himself conscious of any action of, say, the
Pope, then he finds himself to be the same person as the Pope\textsuperscript{23}. The role of memory is
explicitly stated as follows: ‘‘Tis plain consciousness, as far as ever it can be extended,
should it be to Ages past, unites Existences, and Actions, very remote in time, into the same
Person, as well as it does the Existence and Actions of the immediately preceding moment:
So that whatever has the consciousness of present and past Actions, is the same Person to

\textsuperscript{21} \text{Locke, E, p.339.}

\textsuperscript{22} \text{Ibidem.}

\textsuperscript{23} \text{Cf. E, p.340.}
whom they both belong”\textsuperscript{24}. In paragraph seventeen Locke points out that the consciousness in question makes the same person, a self, insofar as it attributes to itself and owns ‘Actions’\textsuperscript{25}. Now, I know that all the passages in which Locke talks of a consciousness of ‘Actions’ such-and-such could be read in accordance with the first interpretative line, that is, as not implying any major commitment to the sound workings of the faculty of memory. Actually, Locke does not use the word memory in this context. But, the following considerations should change this view.

From paragraph eighteen onwards Locke addresses his real worries, \textit{i.e.} the basis for ‘the Right and Justice of Reward and Punishment’. There he states that the all important problem of the responsibility for actions depends solely on the person, that is, on personal identity. It is this notion which offers the basis for the owning of actions and their imputations\textsuperscript{26}. In the subsequent paragraphs Locke discusses the basis for punishment and reward and brings into play his point about sameness of consciousness. He considers a \textit{sleeper} \textit{versus} \textit{a waking} Socrates, that is, a case of a sleepwalker who according to his view disowns his deeds while asleep\textsuperscript{27}. Locke makes his criterion and consequently his official view explicit: ‘If there be any part of its Existence [of that consciousness], which I cannot upon recollection join with the present consciousness, whereby I am now my self, it is in that part of its Existence no more my self, than any other immaterial Being. For whatsoever make my own Thought and Action, it will no more belong to me, whether a part of me thought or did it, than if it had been thought or done by any other immaterial Being any where existing’\textsuperscript{28}. So, the workings of Locke’s ‘consciousness’, that capable of distinguishing individuals, \textit{comes down to memory, recollection}. There will be one person where there is one memory (one continuous and coherent set of recollections coming from the distant past and reaching one’s present thoughts and actions), and the same person where there is same memory (the items remembered continue the same).

\textsuperscript{24} Ibidem.

\textsuperscript{25} E, p.341.

\textsuperscript{26} Cf. E, pp.341-2; also Allison (1966), pp.531-2 for the centrality of Locke’s worries about responsibility.

\textsuperscript{27} Cf. E, pp.342-5 for similar cases.

\textsuperscript{28} E, p.345, my italics: AK, ‘self’ excepted.
Now, why does Locke have a Cartesian conception of consciousness? The answer lies in the way Locke conceives of that of which he is most certain. His basic building block is the conception of a present consciousness whereby I am now a self to myself. This present consciousness is the infallible means through which I own my actions and thoughts, and through this present consciousness I have them ‘before [my] eyes in one view’

This is the Cartesian consciousness which is inseparable from thinking, that is, which amounts to the perceiving that one perceives without possibility of error. As Locke puts it, ‘when we see, hear, smell, taste, feel, meditate, or will any thing, we know that we do so’

What do we know? Locke’s answer is in accordance with a famous tradition: ‘ideas’. ‘[Idea] being that term, which, I think, serves best to stand for whatsoever is the Object of the Understanding when a Man thinks’. Locke is also clear about how we relate to these objects: ‘Whatsoever the Mind perceives in it self, or is the immediate object of Perception, Thought, or Understanding, that I call Idea’. The view of consciousness which results is just this perceiving of ideas: ‘Every Man being conscious to himself, that he thinks, and that which his Mind is employ’d about whilst thinking, being the Ideas, that are there, ‘tis past doubt, that Men have in their Minds several Ideas’

Such views come directly from the Cartesian conception of thinking, and of thoughts: ‘I use this term to include everything that is within us in such a way that we are immediately aware of it. Thus all the operations of the will, the intellect, the imagination and the senses are thoughts’. What is ‘within us’ is, for Descartes, an idea: ‘I understand this term to mean the form of any given thought, immediate perception of which makes me aware of the thought’

Locke thought he could use the element of certainty, of infallible knowledge, in the Cartesian notion of consciousness to account for the identity of persons. He ventured to extend this certain present consciousness to the past, to memories. But he is caught in two

30. E, p.335, my italics: AK.
31. Locke, E, pp. 47, 134 (my italics apart from Idea), 104, respectively. Cf. for the same views Hume, T, pp. 66, 67, 189, 190.
minds in his attempt. As we saw, he claimed that the losing sight of our past selves, the failing of memory, was a problem concerning the sameness of substance, and not the sameness of person (in paragraph ten). However, being really concerned with what guarantees personal identity, or individuality, Locke also thinks that the present consciousness when extended to the past could assure us uniqueness of person. He is also painfully aware of the deadly threat to his theory: false memories. Furthermore, the fact that actions and deeds are ascribed with or without memory of them creates a discrepancy between ‘Humane Judicatures’ and his theory of persons as the basis for the imputation of action. Locke accepts that, for example, in the cases of sleepwalkers and drunkards the ‘fact’ - the action - can be proved against the agent, but not what according to his theory is the case: a ‘want of consciousness’, i.e. the absence of the same person, that is, the presence perhaps of another person as the true agent. Locke feels these difficulties so strongly that he even explicitly considers receding to Cartesianism: ‘I agree the more probable Opinion is, that this consciousness is annexed to, and the Affection of one individual immaterial Substance’. But he takes the heroic stance, which many consider suicidal, and at the end courageously holds on to the view that person, being the name of the self, an entity with an identity over time, is a forensic term, adequate for the appropriation of actions, for designating the being which should be rewarded or punished.

Locke, as it were, delivers. What it is worth is another question. My purpose now is to call attention to the enormity of the difficulty in Locke’s own perception of it. He is willing to ask succour to God: on the Day of Judgement injustices committed against men will be corrected, then ‘hearts shall be laid open’, that is, the consciousness Locke posited as the basis of personhood will be all-pervasive in its whole character. Consciousness then will accuse or excuse without the possibility of any mistakes. To those who want to be judged fairly in the present, Locke offers the worn out argument that the Goodness of God will not permit fatal errors to occur in attributing those consciousnesses to men.

34. E, p.345.
The truth is that Locke never faced up to these problems with his criterion, and that the resort to the Goodness of God is an admittance of failure\(^{37}\). But it is also true that the classical lines of attack on Locke’s view - accusing it of circularity - employed by Butler, Reid and Berkeley, never quite realized the problems involved in the issue. Locke has certainly the merit of being the first to question consistently the obscurities involved in the Cartesian notion of substance\(^{38}\), but his account in appealing to a ‘thinking thing’, which is not permitted to be understood as an immaterial substance \textit{simpliciter}, actually leaves much to be explained, especially when it is posited as distinguishing \textit{individuals}. That is why its status is that of a ‘quasi-substantial entity’. However, in spite of this, Locke is praised, by Allison for example, as having had some valuable insights, in accordance with the ‘self-consciousness-in-time’ interpretation of paragraph ten (‘Consciousness makes Personal Identity’), which influenced Leibniz and ended up being more clearly articulated by Kant with his transcendental (formal) unity of apperception\(^{39}\).

I am not convinced by this. I think Kant’s views constitute a greater break with the tradition, and that Locke’s account, which attacked the Cartesian conception of a thinking substance, while at the same time maintaining that position’s view of consciousness, actually provided a fertile ground for a typical empiricist conception of self-consciousness to grow, which would be formulated more distinctively and forcefully by Hume. I turn now to the contributions of this famous Scot.

2. \textit{Hume}

Hume’s work on the problem of personal identity extracts from Locke’s views the new empiricist stance. Accordingly, Hume is more radical in his criticisms of Cartesianism. However, at the same time, the basic commitments of Empiricism come in Hume more openly to the fore. On balance Hume gives a clearer formulation to the empiricist account, and he also \textit{feels} the \textit{true} difficulties for the account, not so much with the specific solution to the problem of personal identity, but with the fundamentals of the empiricist outlook. Hume is led to suspect the very notion of ‘perceptions’.


\(^{38}\) Cf. \textit{op.cit.}, pp.534-5.

Hume follows Locke, as we saw, to a great extent. However, he starts his section ‘Of Personal Identity’ with a fiercer and more lucid attack on Cartesianism. The idea of self or person, as a being of which we feel the existence and continuance in introspection, and of which we then are certain as enjoying perfect identity and simplicity, Hume proclaims, is contrary to the very experience intended to support the view. According to the principles of his science of human nature it is fair to ask ‘from what impression cou’d [this] idea be deriv’d?’, and to answer as he does: ‘For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can observe any thing but the perception’.

This is the famous thesis of the elusiveness of the self in introspection, that is, in introspection no constant and invariable impression that could account for that idea of the self is to be met with; there we only meet ‘perceptions’.

As I mentioned above, Hume also sets the problem in the context of the notions of identity and diversity, also with the consequence that man, as an ordinary living object, has its identity determined without major problems. I will say something about this later on. As I noted, very much like Locke, Hume also draws on shape and figure to ascribe identity to man.

But in sharp contrast to Locke’s account, in Hume, the worry about providing a basis for responsibility, for the imputation of actions, disappears completely. Personal identity in Hume is a true metaphysical problem. Hume has no qualms in dispelling the Cartesian solution through the thinking substance. Thus, he bravely puts forward his ‘bundle theory’: ‘I may venture to affirm of the rest of mankind, that they are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement. ...The mind is a kind of theatre, where several perceptions successively make their appearance; pass, re-pass, glide away, and mingle in an infinite variety of postures and situations. There is properly no simplicity in it at one time, nor identity in different; whatever natural propension we may have to imagine that simplicity and identity. The comparison of the theatre must not mislead us. They are the successive perceptions only, that constitute the mind; nor have we the most distant notion of


the place, where these scenes are represented, or of the materials, of which it is com- posi’d.\textsuperscript{42}

Following such an assertion of his view Hume puts the problem in his perspective: why do we believe we are those Cartesian selves? Hume is now in a position to explain through his principles of association how we acquire such a belief even though it is not true, that is, how we psychologically come to feign the identity of the mind of man. This point is inescapable. If all ascription of identity is fictitious it would not be different with the mind\textsuperscript{43}.

All ascriptions of identity to objects are fictitious because the mind confounds the ascription of strict identity to objects when they are observed uninterruptedly with the ascription of identity in cases where there is interruption\textsuperscript{44}. This is also what happens in the case of the mind. Although, in agreement with his principles, the identity of mind is not capable of ‘run[ning] the several different perceptions into one, and make them lose their character of distinction and difference, which are essential to them’\textsuperscript{45}, we still believe in it. That is, for Hume the true nature of perceptions is of ‘distinct existences’, which are different, distinguishable and separable from each other, but notwithstanding that ‘we suppose the whole train of perceptions of be united by identity’\textsuperscript{46}. This in Hume’s philosophy of human nature is explainable by the principles of association or natural relations. He investigates resemblance and causation. This implies that what is going to be discovered is not a real bond among perceptions but an association in imagination\textsuperscript{47}.

According to Hume, supposing we could look into the breast of a fellow human we would only observe the succession of perceptions which constitute his mind and would find that many perceptions are actually images of past perceptions. Now, because memory is


\textsuperscript{43} Cf. op.cit., p.259.

\textsuperscript{44} Cf. op.cit., pp.253-5.


\textsuperscript{46} Loc.cit..

\textsuperscript{47} Cf. op.cit., pp.259-60; also Stroud (1977), p.260n1 for one account of why contiguity is disregarded.
only the faculty of producing images of past perceptions, images which as such resemble their objects, the past perceptions, what we have is a chain of mainly resembling perceptions. This is the situation we had in the observation, sometimes interrupted, of an identical object. That is, a set of many similar perceptions. In this respect for Hume memory helps in producing the resemblance between perceptions, which then naturally leads the mind to feign identity.

Causation comes into the Humean account, firstly, because it is natural enough to conceive of our perceptions as producing, influencing, modifying new perceptions, that is, as interacting causally. In this regard the mind is a republic, which continues with the sons and daughters of the old members. Now, when we have this relation among our perceptions in view, we can see that memory actually, as it were, discovers personal identity, insofar as it reveals the underlying causal relations. Again, because the mind runs without hindrances along such a causal chain, identity is feigned. In this respect, causation is present in Hume for a second reason. It is actually our belief in the holding of causal relations among our perceptions which prevents us from withdrawing identity from cases where memory fails. We think that causation holds among perceptions beyond what we can remember. My action on the 12th of December of 1993 is a member of the set of the perceptions that make me what I am even if I do not recollect it. This is the main line of argument Hume employs to escape at least one of the shortcomings of Locke’s account. Hume uses more than memory to present the developed empiricist view in its full strength, certainly because, without major obstacles, we ‘extend our identity beyond our memory’ .

Even if Hume gets round Locke’s problem with the limitation of memory, his own account is not without problems. Stroud objects to the particular functioning of resemblance and causation. Resemblance among perceptions could not per se individuate one set of perceptions from others. Several people could have the same (obviously, ‘similar’ in this account) perception of the Eiffel Tower. Causation is also not a clear-cut case, we certainly experience transpersonal causation of perceptions. Strawson questions the point of the whole approach through association to the problem of personal identity. According to him


the psychological mechanism Hume envisages is completely idle, it is the real fiction. Hume’s search for it is as futile as ‘the search for criteria of subject-identity to be applied in the field of inner experience to determine whether a current experience is or is not one’s own’⁵⁰.

But what is really interesting is that Hume himself was profoundly dissatisfied with his explanation. In the Appendix to the Treatise he realizes that his account is ‘very defective’. The problem, however, which has exercised philosophers since this admission, is to say what exactly is wrong with the account in Hume’s own perception of matters.

In the Appendix Hume admits that he finds himself in trouble concerning personal identity, and he proposes to explain the difficulties, suggesting that he has no idea as yet how to solve them. He does not despair however, and he pleads the ‘privilege’ [sic] of the sceptic ‘to entertain a diffidence and modesty in all [his] decisions’, and confesses that the difficulties are too hard for him⁵¹. The problem for the reader does not lie here, in this mea culpa, and one could certainly anticipate the consequent rhetorical turns of phrase. The challenge is to clarify what the problem really was, not so much for the sake of establishing a historical point about Hume, but rather because one is lead to suspect by Hume’s cleverness that there is something flawed with an account which otherwise did appear to be so powerful. Furthermore, such a philosophical outlook is still very influential, making it worthwhile to query this particular point.

Hume, firstly, recapitulates his views⁵². Then, he states what his position has in common with ‘most philosophers’, viz. that ‘thought’ alone finds personal identity, or that personal identity arises from ‘consciousness’, which is nothing but a ‘reflected thought or perception’.

Hume finally states which of his hopes are not realized, viz. to explain ‘the principles that unite our successive perceptions in our thought or consciousness’. He suggests that the source of the problem is the consistency of two principles which he simply cannot abandon in his philosophy, ‘viz. that all our distinct perceptions are distinct existences, and that the mind never perceives any real connexion among distinct

In the very next sentence he states what it would mean to abandon these principles, we would either conceive of perceptions as inhering in something simple and individual, or the mind would perceive some real connection among perceptions. Hume obviously cannot accept any of these positions. But what is baffling is that these positions do not contradict each other in Hume's philosophy. Supposing that perceptions inhere in something would require us to conceive of them as having real connections with, say, possible perceptions of what they inhere in. Now, the claim about the distinction of perceptions is fundamental for Hume's whole account, being actually used to repudiate the Cartesian position. I think Hume's willingness to refute Cartesianism cannot be doubted, which makes his 'pursuit' of the first position a mere rhetorical move. About the second position however, the consequences could be not the necessary acceptance of Cartesianism, but the recognition of a flaw in the key-stone of his philosophical edifice. I think his complete silence about the matter after the publication of Book III of the Treatise, with the Appendix to the whole book, is symptomatic: the Enquiry simply skips the subject. So, it seems that the real problem cannot be captured only by what Hume says or admits.

Stroud considers a different possibility: the insufficiency of the principles to adequately produce the belief in personal identity. It is certainly not obvious why Hume should have relied upon these principles to explain our beliefs in necessary connections and external objects, but should suddenly doubt them when dealing with personal identity. Following Stroud, this is one way they may be deemed insufficient.

The general technique of appealing to natural relations, principles or dispositions of the mind, works well in the case of the explanation of our fictitious beliefs in 'necessary connection' and 'external object'. The explanation uses a notion of how the mind works in feigning these beliefs. Now, in stating that our belief in 'personal identity' is actually also a fiction, we could rightfully ask what mistakenly leads us to think that we are individual, enduring selves, instead of clinging to the fact that we are bundles of perceptions. Hume is not allowed to answer 'the mind', because this is really what has to be explained, but the official retort 'a bundle of perceptions' is baffling. How could a bundle commit mistakes,


54. Cf. Stroud (1977), p.128 and ff. In the following discussion the idea that here there is a basic flaw in Hume's philosophical position is Stroud's, my discussion here depends heavily on Stroud's development of this point, cf. loc.cit., but there are some disagreements as well.
feign things or believe in them? Or, what does it really mean to state that the mind itself is a fiction? Hume's answer to the 'what' question is 'a bundle of perceptions'. This, following Stroud, is not enough because of the role 'individuation' plays in any minimally intelligible account of personal identity. Hume has to explain why perceptions occur in some 'personal' bundle or other. That is, a belief occurring cannot be explained as occurring 'in a bundle' simpliciter. The explanation, to be satisfactory at all, must imply that a 'to-which-bundle-it-occurs' question could be asked. For the sake of intelligibility the idea of individuation has to be built into whatever account of personal identity, and the suspicion is that Hume fails in this count. Hume sets himself to explain how we acquire the belief in personal identity, and this belief is one in a personal identity, even if it is of 'bundles' of perceptions, but he actually says nothing about individuation or about what makes a 'personal' bundle. Stroud put the objection like this: 'We cannot say that what is the case when I believe that \( p \) is that the lively idea or belief that \( p \) occurs in some "personal" bundle or other. It must occur in a certain particular bundle, viz. the bundle that I am, in order to constitute my believing that \( p \)'.

Now Stroud envisages clarifying this problem by presenting his account of Hume's 'naturalism'. He suggests that Hume would not be in the business of providing 'analyses' of concepts, or logical synonyms of certain phrases, and he would not have done it in the case of 'necessary connection' - always the application of the concept surpasses the objective conditions for applying it, that is, its application does not imply an objective necessary connection between two events. Consequently, he would not have done it in the personal identity case either. I am not sure about this. The two cases seem to be different. Firstly, Hume certainly believed that there was something 'fictitious' about 'necessary connection', but he at the same time offered criteria for using such a phrase. He did not do the same about personal identity. He called it a fiction but did not offer any criteria for the application

\[55 \text{ Cf. op.cit. p.132.}\]

\[56 \text{ Cf. op.cit. pp.132-3, 134-6.}\]

\[57 \text{ Cf. T., section 'Rules by which to judge of causes and effects', pp. 173-6. That Hume provided these rules suggests that he really believed in some objectivity about causal judgements. This suggests that the 'fictitious' character of certain beliefs has a philosophically deflationary force only. For the exploration of this view cf. Buchdahl (1969), Chapter 6, especially pp. 337-48; Klaudat (1991), Chapters 4 and 5, pp. 61-121. For an example of lack of consideration of these strata, like Stroud's, cf. Goodman (1979), pp. 59, 62 and 82.}\]
of the notion. However, more importantly, assuming Hume did not offer any analysis of 'necessary connection', but produced rather 'natural descriptions' of how we come to believe in it, such an approach cannot answer the 'what' question in Stroud's own way of presenting it. That is, the naturalistic approach just supposes what we now have to explain, even if no analyses are offered. In a naturalistic account of necessary connection, what was simply supposed without any further ado is now the object of explanation. The individuation problem is not simply a demand for an 'analysis'. Rather the point is that because of the centrality of the notion of mind or self this notion cannot be left unexplained, even by someone with a bent for naturalism, because what is being investigated now is one central element of the 'descriptions' themselves. So, this attempt by Stroud to help the position via naturalism is off the mark here. Stroud actually goes on to pursue a different line that at the end will come back to the centrality of the problem of the mind. Let's follow it.

The real problem is the conception of the uniting of our successive perceptions in our thought or consciousness. This is really a problem about what unites perceptions together in our thought, 'or what makes us think of them as tied together'. That is, the problem is about 'what makes us think of certain bundles of perceptions as constituting continuous selves'. It is in this respect that Hume says that his difficulty is with the determination of the thought to pass from one object (perception) to another, with the thought discovering personal identity when reflecting on past perceptions. Stroud at this point suggests that what is really troubling Hume is the nature of experience itself, that is, something about the conception of experience makes the job of explaining personal identity impossible. This is the account of experience in terms of the theory of ideas. Hume's problem would be to explain the 'data' upon which he bases all his naturalistic descriptions, that is, he could not account for 'the undeniable fact that the only "data" available to a person for the formulation of his ideas and beliefs are his own perceptions'. According to Stroud this is a crucial point about the self or self-consciousness.

60. Cf. T. p. 635.
Hume cannot account for this 'connectedness', and for him the *theory of ideas* dictates that all perceptions are completely distinct, mutually, and from everything else. They are independent existences, with no real connections of any sort between them. This obviously makes it impossible for him to explain the *basic* 'connectedness' he actually relies upon in explaining 'necessary connection' and 'external object' but which now has necessarily to come into focus in a minimally intelligible explanation of how we acquire the idea of personal identity.

Stroud at the end speculates that although Hume certainly never dreamt of abandoning the theory of ideas he actually half recognized in the *Appendix* a fatal deficiency in the theory in his understanding of it. Hume would have recognized that there is a need to posit *intelligible* connections between perceptions and minds, but then perceptions would not be completely independent existences from minds. This would explain why Hume connects his 'very defective' account with his having 'loosen'd all our particular perceptions'.

It is interesting that Stroud's own way out of this difficulty is through insisting that thinking and psychological phenomena in general require a *subject*, a someone thinking or being in a mental state. The problem with the theory of ideas is that it gives undue prominence to the mental 'objects', and loses sight of the 'subject', only within which those 'objects' can exist. Hume would have had to stretch the theory too much, to the point of revealing its shortcomings. In Stroud's opinion, *Kant* would have put things in order, by insisting that there must be something about perceptions which makes them belong to a particular self or subject. Stroud suggests that it *could* be the requirement that it must be possible to think of perceptions as belonging to oneself, but note 8, at p.261, goes further and adds a reference to *bodies* of people as being on the right track for such a requirement. Accordingly, in respect to what is 'primary in human thought', Kant would be nearer to a suitable position by abandoning the theory of ideas and bringing in the notion of 'judgement'.

---


63 Cf. *op.cit.* p.140, and note 13, p.262. Cf. also Stroud's account of how the theory of ideas hinders Hume's naturalism, and how Kant comes closer to a satisfactory account of these matters, pp.224-39.
This is not the place to assess Stroud's view that there is something about perceptions that makes them belong to particular selves or subjects. In his book *Hume* this question actually remains open. More will be said about this, although not in connection with Stroud, in the next chapters. I think, however, that Stroud is right in his indication of the direction. In relation to what selves are, and what the nature of self-consciousness is, he only gestures towards Kant, but about the crucial *logical* point concerning the limitations of the Cartesian framework about ideas Stroud is more emphatic: 'There is good reason to suppose that the task is not an easy one, but at least a few steps away from the restrictive theory of ideas would be steps in the right direction'\(^64\). The right direction is one which leads from the conception that we have a 'perfectly hard, clear and distinct conception of "what is actually the case"' when we concentrate upon discrete events in our minds, to the employment of *judgements* (as statements that have a truth-value) in accounting for our capacities, abilities or even competencies, that is, to the employment of *judgements* as primary in the account of what it is to 'possess' certain concepts, *i.e.* to use them to judge\(^65\).

This point is very important, and will be pursued in later chapters. To conclude this chapter, what remains to be stated is the conception of self-consciousness that comes out of Empiricism’s worries with ‘personal identity’.

Self-consciousness is consciousness, not of a self, but of perceptions or experiences, which then, in forming a group or bundle constitute a self. Self-consciousness is knowledge through introspection, not of a self, but of different perceptions which succeed each other in a constant flux. As for the identity of these bundles we have seen Hume's difficulties. Locke, not so radical in eschewing the single mental thing, but moving towards a bundle view, had difficulties of his own, as we saw. In this regard the bundle theory has no element to correspond with the second major feature of Cartesianism (identified in chapter one), which guaranteed identity and consequently gave a secure indication of the metaphysical status of the soul. Hume’s insistence on the discrete perceptions, conducive actually to their reification, creates an ambiguity. This can be stated in contemporary terms in the question: are perceptions mental private objects accessible only to myself or are they material public objects accessible to others as well? However there can be little doubt that the Cartesian


influence upon the Empiricists' thinking led the bundle theory to be about mental private objects.

Now, in parallel to Cartesianism's first major feature, Hume's position is that the 'I' stands, not for a substance, but for something nonetheless, *i.e.* a collection, a heap, of perceptions. Kant did not single out the empiricist account for explicit criticism probably because it does not indulge in the kind of excesses of Cartesian Rational Psychology. In a certain sense it is deflationary already, surely to Kant's taste\(^6\). But there should be no doubt that because of the account's reliance on introspection, on the *knowledge* of what is this time *inside*, as a model for self-consciousness, Kant would have condemned it as also committing the mistake of confusing self-consciousness with self-knowledge\(^7\).

---

\(^6\) Cf. Cassam (1989), *passim*, especially p.78n18, for agreements between Kant and reductionist strategies (lower case 'r'), those that affirm that a person's existence consists in the existence of a brain and body plus the occurrence of physical and mental events.

\(^7\) This is not Cassam's line in presenting the 'Kantian' criticism of Reductionist views (capital 'R'), like Parfit's, which accept the view about a person's existence as in the note above, but claim that it is possible to give a complete description of reality without asserting that persons exist. This leads to the view that it is possible to explain the unity of consciousness in *impersonal* terms.
CHAPTER III

STRAWSON ON THE PARALOGISMS

The goal of this chapter is to present a path to the heart of Kantian views on self-consciousness. I have chosen the indirect approach of starting with Kant's criticisms of Cartesianism, instead of presenting his positive views, because of Strawson's elegantly compressed discussion of it, which allows one to come near the centre and to see clearly one way of trying to reach it, although one which I will argue is neither Kant's nor the best. Besides this, there is a special interest in Strawson's views about, as Williams put it, the macro-level of the problems of mind and body, that is, the conception of persons, because it is a Kantian view. What distinguishes it from Kant's own view are the elements of empiricism it contains, especially concerning experiences.

In the overall structure of this thesis this chapter serves the purpose of pointing out the difficulties to be tackled in relation to Kant. Notions like that of the 'transcendental unity of apperception' will only receive an introductory account here; Kant's statements about it and their full-fledged alternative interpretation will be discussed later on. So, this chapter does not intend to exhaust the message to be drawn from the 'Paralogisms' (i.e. the Kant's text). From the forthcoming discussion I also hope that at least some hints can be obtained as to why Kant introduces a wedge between self-consciousness and self-knowledge.

Strawson starts with three revealing general points. Firstly, the doctrine to be attacked by Kant is identified as being Cartesianism, which is, as we saw, the conception that we are - and we know it through mere conscious experience - each of us, an individual thinking substance, that is, an immaterial, persisting substance, identical with itself; an individual soul, not dependent for its existence on the body it is connected with and ultimately from matter in general.  


2. Cf. Strawson (1966), p.162. The following discussion is all contained in the chapter 'Soul' of Strawson's book, pp.162-74. Only points made elsewhere will be referred by pages. I will also draw on views presented in (1974b), where Cartesianism is said to be the view that a human being is really a complex of two one-sided things: a mind and a body, cf.p.170.
Secondly, the way to attack Cartesianism is to appeal to the ‘principle of significance’, a principle Strawson finds in Kant. Roughly, it says that knowledge of an object as an existent with a certain property, or as a particular falling under a certain concept, requires empirically applicable criteria for using that concept, or for ascribing that property. According to Strawson, and he is correct about this, sensible intuition plays this role in Kant. The problem is, however, what the import of this notion is in Kant. We shall have to leave this question for when we come to deal with Kant himself. In the present case, what is required, according to Strawson, is that the crucial concept of a numerically identical persistent being should be connected with reality, or that there be an intuition (empirical awareness) of such an object. And here Cartesianism fails, because it does not specify any such intuition. Obviously, Strawson would have us believe, it could not, because there is no possible intuition of a Cartesian immaterial subject of experiences. It is important, so the claim goes, to notice that, given that Cartesianism turns the self into an object, it is fair to ask for an intuition of it.

Strawson’s third initial remark concerns the diagnosis of the error committed by Cartesianism in its account of the self. Kant’s famous explanation is oracular: the Cartesian Rational Psychologist confuses ‘the unity of experience with the experience of unity’ (cf. B413, 421). Strawson agrees that there is something to Kant’s diagnosis, but much has to be explained if the unity of consciousness, understood by Strawson as the connectedness of experiences, is to be linked to the awareness of an immaterial object which is the unitary subject of the experiences. Strawson believes nonetheless that with his ‘reconstruction’ of the exposure of the illusion Kant’s views can be rescued.

We can divide Strawson’s reconstruction into three stages. At first he makes some basic points: his interpretation of the transcendental unity of apperception, and its necessarily correlated notion, the Strawsonian account of ‘person’. He then moves on to explain where the problems start for Cartesianism. That is, given the basics, some facts obtain concerning self-ascription of experiences, and Cartesianism misunderstands these facts. Finally, Strawson states what the limitations of Kant’s exposition are, but also how it can be repaired so that we can give a coup de grâce to Cartesianism. I will follow this route of Strawson. My goal here, however, is to stress the centrality of the notion of the transcendental unity of apperception in these matters. I will present in an introductory fashion the Strawsonian account of it, which is closely linked with his conception of an
empirical subject of experiences'. This latter conception has in my opinion empiricist elements, which at the end lead Strawson to misrepresent Kant's view. These elements will be called into question here, firstly, insofar as they lead Strawson to commit this mistake about Kant; and, secondly, on their own merit. How one can present an alternative account of the unity of consciousness will be left for chapter four and five.

Strawson begins by presenting his working understanding of the transcendental unity of apperception. According to him, Kant with this conception presents the requirement that there must be a unity of consciousness. This is a seminal requirement for the thesis of the objectivity of experience. That is, the thesis 'that experience must include awareness of objects which are distinguishable from experiences of them in the sense that judgements about these objects are judgements about what is the case irrespective of the actual occurrence of particular subjective experiences'\(^3\). Strawson does not talk much about this thesis in the context of his discussion of the Paralogisms. In this context he is concerned with explaining what the unity of consciousness requires, and what it is required for. It requires, given the unanalysed conception of a series of temporally extended experiences, that this series should have 'a certain character of connectedness and unity'. Now, this is obtained through the means of the 'concepts of the objective'. This needs clarifying, and I will return to it. What is important for the moment is that the connectedness and unity of experiences condition the possibility of empirical self-consciousness, that is, the ascription of experiences by an empirical subject of them to himself, and that is what it is required for. This notion of empirical self-consciousness, in turn, is very important indeed in Strawson's defence of the thesis of the objectivity of the experience. However here, in the context of the pursuit of a concept of self, the connectedness and unity of experiences have an implication that is really crucial. If we assume for now, as Strawson does, that experience is experience of an objective world, that is, that experience is of objects which are distinct and independent from it, then connectedness and unity give rise at least (i.e. minimally) to the idea of one series of experiences of the world. This is potentially, according to Strawson, an autobiography. And this is a simple but important result in the search for the concept of the self, i.e. as the subject of such an autobiography.

Now, everything said up to this point only offers absolutely minimal features of a potential autobiography. The connectedness and unity only represent the basic condition of any talk about selves. For Strawson, more has to be said, the full conditions have to be spelt out, before this can lead to the concept of a person, which is the empirical concept of a subject of experience. Significantly it is this ‘filling out’ that will actually account for what was left unexplained above, the point concerning ‘the concepts of the objective’ I said needed clarification. With the exposition of the full conditions of any talk about persons, Strawson explains how these concepts bring about the required connectedness and unity\(^4\).

Strawson requires, in accordance with the principle of significance, that talk of a series of experiences should be linked with empirically applicable criteria of identity of such a series. Only in this way can we realistically expect this concept of a series of experiences to have any meaning whatsoever. Now this can only be done by appealing to our ordinary concept of personal identity, which is conceived as involving empirically applicable criteria for numerical identity through time and space of what is a subject of experience, i.e. a man, or a human being, for Strawson. These criteria, in spite of not being identical with criteria for bodily identity, make an essential use of the notion of the identity of a human body. So, what Strawson says is that the concept of a person, or an empirical subject of experience, is not a concept of a body, but it involves such a concept in an essential way. This is a very important point, which is more fully developed in the third chapter, ‘Persons’, of Individuals\(^5\). This concept can have a crucial role to play in another context, viz. in assessing Kant’s position, as a means of avoiding the mistake committed by materialists who think that persons are no more than bodies. But here a particular aspect of the Strawsonian general position comes to the fore. What is needed is a way of linking the experiences with reality, to the realm of the objective, or of objects. The way to do this is to connect such a series indirectly through the concept of a person (in the end because there

---

4. Notice that there seems to be a circularity here, which I only mention now. I said that unity of consciousness was required for objectivity, but now we have that unity accounted for by the means of ‘concepts of the objective’, that is, the unity of experiences will be accounted for in terms of - as we will see - an object in the world, one among others whose conception we can entertain objectively. Perhaps an illuminating way of putting this would be to say that an epistemological thesis about the objectivity of experience is going to be defended through a metaphysical thesis about the character of experiences as particulars pertaining to an object in the world.

must be a *subjectivity* side to experiences in general) to a particular *body* (*numerically* identical through time and space). So, it is this body-component that comes to the fore: this appeal to a subject as a physical thing\(^6\). It is the body which in the end warrants the conception of *one* unified experience of the world, *one* experiential route through the world. It is in this respect that Strawson boldly asserts that he has a concept which satisfies 'the most stringent critical requirements', offering 'an absolutely firm basis for a genuinely *object*-referring use [my italics: AK]' of all *personal* terms, including proper names and pronouns, the foremost being 'I'. And here we see what the threat Strawson wants to dispel is. It is with such a use of these personal terms that Strawson intends to eschew the notion that states of consciousness, inner experiences, etc, are being ascribed to something immaterial, really unknowable, very suspect indeed, or even to nothing at all. These terms *refer* to something objective, to an *object* in the world which is a 'perceptibly permanent, persistent and identifiable' *object of intuition*. For Strawson this is the secure basis for the talk of *persons*. What I would like to call attention to in this way of solving the difficulty of connecting a series of experiences to reality is the nature of the initial problem: *experiences*, a *series* of them, had to be anchored to something objective somehow, otherwise Cartesianism would be in a dialectically advantageous position. I will focus on this conception of experiences below.

We can enter now the second stage of Strawson's reconstruction. Given the view of persons advanced so far there are some important facts which have to be taken into account to see clearly where Cartesianism goes astray. Firstly, when a subject of experience ascribes experiences to himself, it is not necessary for him to identify himself as the subject to which these experiences are being ascribed. That is, it is not necessary for him to identify the subject referred to by 'I' in saying 'I see such-and-such'. The reason given is that it does not make sense to entertain the thought 'this inner experience is occurring, but is it occurring to *me*?' The point is the general one that, as Strawson puts it, 'there is nothing that one can thus encounter or recall in the field of inner experience such that there can be any question of one's applying criteria of subject-identity to determine whether the encountered or recalled experience belongs to oneself- or to someone else\(^7\). This point has also been

---


referred to as the immunity to referential failure through misidentification of the first person pronoun\textsuperscript{8}.

But, in line with his account of persons, Strawson has more to say about the use of 'I'. In spite of the irrelevance of identification for its use, this pronoun nonetheless does not 'lose its role of referring to a subject'. That is, in spite of the impossibility of misidentification, 'I' is still with absolute invariance referential. Even in 'criterionless' self-ascription of experiences reference to a subject is not in fact lost. This feature of the use of 'I' is also known as its immunity to referential failure tout court\textsuperscript{9}. For Strawson, then, even if I do not have to identify myself when self-ascribing experiences, my use of the 'I' does not lose its referential power 'because - perhaps - it issues publicly from the mouth of a man who is recognizable and identifiable as the person he is by the application of empirical criteria of personal identity'\textsuperscript{10}. What happens is that the use of the 'I' does not lose its links with the empirical criteria of subject-identity. What will have importance later on is the fact that here 'in practice' means from a third-person perspective. That is, 'I' refers to a subject because others listen to its issuing from a mouth to whose owner they can refer and to which an experience can be ascribed. The bottom-line in this view is that I do the same, although not observationally, about myself in self-ascription, in spite of its criterionlessness. In this connection, Strawson talks of a 'rule' which is 'general': even in self-ascription (and that is what we are doing) of experiences we use the concept of a subject of experience as one which necessarily requires the identity of such a subject as 'an object of outer sense'. The plausibility of these claims will be challenged in chapter 5 below.

Now, the Cartesian account of these facts, of this situation, is deeply mistaken. The criterionlessness of self-ascription is explained philosophically by the Cartesian as an immediate contact with something of 'singular purity and simplicity - a pure, individual, immaterial substance'. What happens is that the Cartesian recognises the criterionlessness of self-ascription, but he gives it a wrong basis: he thinks that in these cases (of self-ascription) there is an immediacy, a kind of internal direct contact with the experience, which leads him to give a purely inner basis for self-ascriptions. In 'Self, Mind and Body' Strawson identifies

\begin{itemize}
  \item \textsuperscript{8} Cf. Powell (1990), p.207 and ff.; Evans (1982), p.215 and ff..
  \item \textsuperscript{9} Cf. Powell loc.cit..
  \item \textsuperscript{10} Strawson, loc.cit..
\end{itemize}
this basis as 'a certain experience of intense looking within, or introspective concentration...
One is tempted to say in such moments that one has direct experiences of oneself as a
conscious being'\textsuperscript{11}. This 'experience' produces the conviction that, if one 'sees' how things
are with oneself, or even the awareness of oneself in undergoing the 'experience', this is a
completely indubitable piece of knowledge, so secure as to make unnecessary any
explanation of the identity and difference of souls, basically because this direct experience is
an experience of my individuality and identity\textsuperscript{12}. But, the Cartesian error is compounded in
this way because he not only recognizes the criterionlessness of self-ascription, he also
wants to maintain the ascription of experiences to a \textit{subject} of them. So, 'I' for the Cartesian
is also referential, as we saw in the first chapter, but obviously because of the inner basis of
self-ascription it will refer to the special sort of being I already mentioned. Strawson’s
explanation of this mistake is as follows.

The Cartesian falls prey to the illusion of 'a purely inner yet subject-referring use for
the “I”’. What he does not realize is that criterionless self-ascription of experiences is
possible without the ‘I’ losing its customary role of referring to an empirically identifiable
subject of experiences. The Cartesian ends up disconnecting the ‘I’ from the ordinary
criteria of personal identity, but still believing that it is referential, believing accordingly in a
reference to a special being. But, for Strawson, this move cuts the ‘I’ from any referential
use whatsoever. For him, what the Cartesian is really talking about is what Kant calls the ‘I’
expressing ‘consciousness in general’. But Strawson is not serious about the ‘I’ losing its
referential force; what the Cartesian really does in cutting the use of the ‘I’ from its normal
surroundings in his philosophical reflection is to use mistakenly the bare \textit{form} of reference
to persons. Noticeable here, then, is that the ‘I’ is just assumed by Strawson, without any
other possibility, to be \textit{par excellence} referential.

In the final stage of his reconstruction Strawson explains where Kant was right and
where his account was limited. Basically, that Kant criticized Cartesianism the way he did
was brilliant, but he failed to stress the relevance of that crucial component of Strawson’s
view itself: the body, an object of outer sense. In this respect, Kant made too little of the
empirical criteria of subject-identity. This makes for an incomplete exposure of

\textsuperscript{11} . (1974b), p.175.

\textsuperscript{12} . Cf. \textit{op.cit.}, pp.175-6.
Chapter 3

Cartesianism's error, but the Kantian diagnosis can yet be justified. The linkage between necessary unity of experience and believing that this constitutes the special experience of one's own unity is through the idea that the 'I', used mistakenly in a referential way, 'expresses' nonetheless that unity. But this unity is only the absolutely minimal condition for any conception of experience; it is far from the totality of conditions necessary for objective experience. However, it is plausible to think that, no matter how far-fetched it is, the entertaining of that abstract thought of a minimal connectedness of experiences leads the Cartesian to conceive of a pure subject of experience, one to which nothing empirical, especially concerning criteria for identification, matters. But the entertaining of this minimal connectedness is not enough; one must also disregard the normal referential conditions for the use of 'I' while maintaining that it nonetheless is referential. We saw in the first chapter how Cartesianism follows this path. As could be expected, this is not the last point that troubles Strawson.

Now, in Strawson's opinion, in spite of Kant's shortcomings, the framework within which he locates his views allows one to deliver a final blow to Cartesianism. In this final criticism Strawson makes much of Kant's famous image of the series of elastic balls (the footnote at A364-5). Its moral is, for Strawson, that the Cartesian guarantee that through the inner basis of self-ascription we are really each of us in contact with our souls, conceived as immaterial substances which maintain their identity throughout a succession of different states, is compatible in fact with the supposition that there are several souls. In a series of such self-ascriptions nothing in the Cartesian story rules out this possibility. Just as motion can be passed from one to another in a series of elastic balls, so that self-consciousness, which is self-ascription, could be passed from one to another in a series of souls. Kant's point would be that the Cartesian claim to the uniqueness of the soul supposedly guaranteed by the internal direct contact is, in fact, compatible with the postulation of a series of souls. Thus far I can agree with Strawson. Kant's point is that it is futile to claim that we experience or know a self in self-consciousness, be it conceived as any sort of thing. But what Strawson wants is to fill out this attack, and it takes the specific form of an assault using the weaponry of the principle of significance (a Strawsonian principle). It is doubtful, however, that this is Kant's strategy.

This principle states that 'there can be no legitimate, or meaningful, employment of ideas or concepts which does not relate them to empirical or experiential conditions of their application'\(^{14}\). What those empirical or experiential conditions are, or what is to be understood by specifying an 'experience-situation', has still to be seen. Here Strawson commends the principle to *empiricist* philosophers as the tool that Kant uses to repudiate all of transcendent metaphysics. In the context of the Cartesian conception of self-consciousness, the result is the following.

The Cartesian *expression* of self-consciousness, in sentences which self-ascribe experiences, has no way of guaranteeing the uniqueness of the self to which these experiences are supposedly being ascribed: there could be 'a thousand souls simultaneously thinking the thoughts his words express'. Strawson's point is that only *our* criteria of singularity and identity for subjects of experience can do that, which in fact is the way of identifying people, men, and in this the crucial component is that we are physical objects, not immaterial substances.

Strawson objects to the Cartesian account in the same way in 'Self, Mind and Body'\(^{15}\). The problem with the account is that it fails to give a proper dualistic *analysis* of sentences involving persons as subjects. The Cartesian fails to make the talk of a mind or soul, as the subject in a sentence, genuinely *referential* to something which is metaphysically independent of a person. The point is that, if you want to refer to things you have to provide means to *count* them. Furthermore, if they are conceived of as persistent, they have to be able to be re-identified. Strawson again presses the point that the Cartesian is unable to explain the *numerical* identity and difference of individual selves or souls, making his account *incoherent*. Strawson claims that *he* can easily count selves or souls: each *person* is one. He makes the notion of person fundamental in his account, the notion of a self, a mind, being derivative in being a *person's* mind. Now 'person' does not stand for a body as such, but Strawson recognizes, at least once\(^{16}\), that his way of counting persons depends on the identity and identification of particular *things*. However, concerning the criticism, is there anything wrong with this manner of attacking the Cartesian? And is this

---


Kant's way? These are difficult questions to be answered and I will start with the second. Then, I will say something about the first, arguing for the view that there are elements of empiricism in Strawson.

I do not think that Strawson's attack on Cartesianism demanding criteria for counting souls is Kant's also. Now, I do not mean this to imply that Kant would not count souls by counting people, or that Kant would not take the concept of person to be fundamental. Furthermore, I do not want to suggest by the criticism I am about to make of Strawson that he is wrong, from a Kantian point of view, in maintaining that a person is not a body. Strawson's empiricist motivations, however, show up in his interpretative reconstruction of Kant's exposure of the Cartesian illusion, as I hope to show in the continuation.

Kant's point against the Cartesian, especially in the elastic balls image, can be understood through a comparison with one of Wittgenstein's in the 'private language arguments', viz. that which purports to show the irrelevance or redundancy of a private object as a private sample that underpins the use of a word. In paragraph 271 of the Philosophical Investigations, Wittgenstein asks the reader to imagine a person who can not remember what (what thing) 'pain' means. That is, supposing that he has grasped the meaning of the word by the use of a private sample, we now suppose that he cannot remember that private ostension, and he ends up, according to the story, calling different things by that name. Now, for someone who maintains that it is by private ostensions that one really gets to understand the meaning of words, it must be possible that this unfortunate state of affairs obtains, but also, at the same time, that the person in question in fact uses 'pain' in such a way that it fits with all the usual symptoms and surroundings of pain. Actually, that he uses it as we all do. Wittgenstein's comment on the original story is that the supposedly private sample does the same work as that of a wheel which turns but does not move anything in a mechanism, actually of a wheel that is not part of the mechanism, that is, an idle wheel. Now, the important and illuminating part of this story is not that we are not given empirical criteria for identity and identification, so that we could count per impossibile 'private objects'. Rather, the point is one about the irrelevance of such a notion.

to our *use* of the word. Wittgenstein's point about the person and his faulty memory aims to show that the idea of a private object underpinning the use of 'pain' is actually as explanatory as the supposition that there are *many* private objects underpinning 'pain' while the person is using the word as we all do. Structurally, that is the same point as Kant's point about the elastic balls. Kant's image intends to show the irrelevance of a soul, or for that matter, many souls, for our use of 'I'. This bears on Strawson's views, such that his alternative *physical object* is not required by Kant to explain either the meaning of 'I', or how self-consciousness is to be understood. But why is Strawson led to think about these matters in the way he does? The basic line in Strawson's arguments is that empirical self-consciousness requires the possibility of conceiving of experiences themselves in an objective manner. So, a lot depends on this conception of *experiences*.

Strawson starts his philosophical project of justifying the objectivity of experience with an 'unquestionable datum', that experience 'essentially exhibits temporal succession'. This, properly understood, is just the conception of 'a temporally extended series of experiences'. But, as I said, this in Strawson's view is not enough; what is required is the filling out of this notion with the full conditions for the application of the concept of a subject of experience. What is really built into the starting notion of experience that leads Strawson so smoothly to the idea of such a subject, with the possibility that at the end what really matters is only the notion of a physical object, *i.e.* a body? The answer is that Strawson has a view of experiences as *particulars*. In *Individuals* the basic model to deal with them is that of a subject which *has* them. This model, with its reliance on *possession*, is presented through the idea of self-ascriptions of experiences, that is, what one does when one genuinely ascribes one's states of consciousness to something, *viz.* oneself as an *object in the world*.

19 . Cf. Mendus (1984), p.140: 'Kant's argument against the rationalists is not that their conclusions are meaningless, or nonsense, but only that they are empty, that they will do no honest philosophical work for us'.

20 . Cf. Strawson (1966), pp.89-112, especially pp.100-1, 105, 106, 110-1; also Cassam (1989), pp.85-94. Cassam is in the business of explaining what the kernel of the 'objectivity requires unity of consciousness' argument is, but he admits that the fundamental notion is that of a single, subjective, experiential route through an objective world, a notion which cannot as such stand on its own. It has to be explained by the fully-fledged notion of a subject of experience, which turns, in the end, on the notion of the subject as a *physical thing* tracing a path through space and time.

I can now present the basic tenet that lies behind any talk of a series of experiences, those that will make way for the conception of a subject of them. We talk about a series of experiences, we then say that such a series is a potential biography, or autobiography, and from this we want to derive the conception of a subject or a self. Now, for this conception to have any meaning we have to have a way of identifying series of experiences. A body, an object in the world, that is, a physical object with a spatio-temporal path through the world will suffice for this. However, this is a plausible solution only if we start with the Strawsonian tenet that it must be impossible that one particular experience should be part of the history of two separate selves. Bird construes this point as the view that ‘one person cannot have intuitive access to the empirical experiences of another’22. Strawson states the tenet in several places, but in no way more cogently than when he makes it depend on the particularity of experiences.

States, or experiences, one might say, owe their identity as particulars to the identity of the person whose states or experiences they are. From this it follows immediately that if they can be identified as particular states or experiences at all, they must be possessed or ascribable in just that way which the no-ownership theorist ridicules: i.e. in such a way that it is logically impossible that a particular state or experience in fact possessed by someone should have been possessed by anyone else. The requirements of identity rule out logical transferability of ownership23.

So, the ‘requirements of identifying reference’ to experiences as particulars make the logical, that is, the possible, transferability of ownership of experiences really impossible. For Strawson, the ‘identical’ pain (a particular) one owns is, in fact, one’s own, and cannot ‘logically’ (now understood as involving metaphysical analysis) be of somebody else, be another’s pain. Strawson, in presenting this view, actually equates ‘particular states of consciousness’ with ‘private experiences’24. Now, this can be called a view of experiences as particulars which are essentially (metaphysically) privately owned25.

For Strawson, this conception of experiences as particulars is the building block of the very big and complex edifice of objective experience. This conception eventually

---

provides the basis for persons to meet ‘the most stringent critical requirements’. Let us see how this is brought about.

Strawson’s view of experiences as particulars to which we can identifyingly refer seems to turn these subjects of discourse into items about which it is plausible to ask questions of numerical identity. Strawson seems willing to accept as perfectly intelligible an account of experiences in the following mould: two persons may have qualitatively the same or identical headache, but to do justice to the fact that there are two aches, that we can count them, we have to claim as well that they are numerically distinct, that is, they are in fact two headaches. For Strawson, to guarantee the logical impossibility of someone having another person’s headache we have to claim that, for example, my headache is my numerically identical headache. You can have a headache as well, but it is yours, and your headache although exactly like mine is another headache. So, what Strawson seems to be willing to explain is the way in which we can say that two persons cannot have the same headache. That is, headaches, pains, sensations, etc, can have both qualitative and numerical identities. How can the basis of this view, the claim about the numerical identity of experiences, be secured, and be made to meet stringent critical requirements?

The numerical identity of material or physical objects is normally explained by appealing to the notion of location in physical space at a certain time. We saw very briefly in the second chapter how Locke and Hume applied the ideas of identity and diversity to things. The conclusion was that no matter how similar two things are in terms of their qualities, as long as they are in different places at the time of inspection, they are, we can conclude, two things. If I saw a chair in my study room yesterday and I see another today, I may intelligibly ask whether they are the same numerically. If answered that the one I saw yesterday is in the next room, I may safely conclude that there are two similar chairs. Can one, or Strawson, resort to the same expedient concerning ‘experiences’?

Philosophers are tempted to answer ‘yes’ to this question. Although Strawson does not say anything which can be taken as a straightforward answer to this particular question, I

believe his position, if examined in terms of its unstated consequences, especially in relation to certain connected problems, implies an affirmative answer\(^{27}\).

Let us take pain as an example. We can have the same pain, an acute pain in the arm. Now it seems that we can talk of \textit{two} pains here because what is really said is that I have my pain in my arm and you yours in your arm. That is to say, the pains are in different \textit{places}, given that the two arms are. This view makes the location of a sensation the criterion for numerical identity, and, accordingly, the location the criterion of possession, of ownership as well. As it were, X's pains are the pains in X's body. So, what we say when we utter 'I have an acute pain in the arm' should be understood in the following sense: I have a pain in a numerically identifiable arm, which is referred to by me in my uttering the sentence, and which has a specific location; and you say that you have a pain in your arm, which occupies

\[\text{27 . Cf. Cassam (1989), pp.89-102, for the development of the Strawsonian position involving claims about how objectivity requires that \textit{perceptions} be based on the \textit{location} and orientation of the perceiver. This is the view that perception is egocentric and that objectivity is only guaranteed if perceptions are episodes or items in an experiential route through the world, and this in turn only makes sense with the notion of a physical thing tracing a spatio-temporal path through the world. In dealing with a different problem Strawson shows, I believe, his commitment to the view presented here. By supposing the unanalysed point about experiences as particulars, that is, dependent particulars which can be identifiably referred to, Strawson moves on to claim that in self-ascription of experiences one has in many cases 'an entirely adequate basis for ascribing a P-predicate to oneself, and yet in which this basis is quite distinct from those on which one ascribes the predicate to another' (1959), p.107). What is this basis? With his example of depression (\textit{op.cit.}, pp.108-10), Strawson makes two points related to this: (1) we do not need to observe our behaviour for self-ascribing depression, feeling it is sufficient; (2) this does not create a problem about other minds, because the piece of language used to self-ascribe depression has a \textit{structure} which is also present in ascription of depression to others. So this \textit{publicly} structured language guarantees the correlation between self- and other-ascription, thus avoiding \textit{epistemic privacy}, the claim that if only I have only I know for sure (cf. Hacker (1993), pp. 25-33). However, Strawson wants to stick to the 'only I have' part of his story. Consequently, 'X's depression is something, one and the same thing, which is felt, but not observed, by X, and observed, but not felt, by others than X' (Strawson, \textit{op.cit.}, p.109).

Now, leaving aside the problem that the feeling of depression is \textit{not} the same thing as depressed behaviour and so that it is in fact \textit{not} the case that what is felt by X can be observed by others, Strawson goes further and says something about the \textit{having} of depression, X's having of a something. Earlier on, in criticising the no-ownership theory, Strawson agreed that perceptual experiences are contingently (causally) dependent on the perceiver's body, but he contended that this does not imply, just because of the contingency, the no-ownership of experiences, but rather, according to Strawson, it helps to account for \textit{genuine} ownership, if we accept beforehand the view that experiences are \textit{privately owned}. That experiences, not \textit{all}, but some, \textit{mine} in fact, can be identifyingly isolated for the pinning down of 'the identical pain which [is] in fact one's own' by describable objective causal relations between objects and the body I am, this is Strawson's way out to the problem of having to single out the series of experiences that I am, now not contingently so. That is why, I believe, Strawson states 'that a necessary condition of states of consciousness being ascribed [recognised as being owned] at all is that they should be ascribed to the \textit{very same thing} as certain corporeal characteristics, a certain physical situation \&c' (\textit{op.cit.}, p.102, my addition: AK).\]
a different location in space. Notice that we are here talking of a physical space. So what we have are two numerically distinct pains\textsuperscript{28}. Is this a defensible position? I do not think so.

There are several ways of challenging this position. It is possible to explain how one comes to consider such a position by appealing to how we forget differences in talking about experiences. In talking of ‘a pain in my arm’, does ‘in’ have the same meaning as in ‘I have a pin in my arm’? When I say ‘I have a pain’, does ‘have’ have the same meaning as in ‘I have a car’? And how do we count pains? If you have a headache and I have a headache, does it follow that there two headaches in the room?\textsuperscript{29}. Another fruitful path is to compare how we talk of identity of things and, say, colours, opinions, mental events, sensations\textsuperscript{30}. I will concentrate on the identity of sensations.

The position I am considering holds that the identity of a sensation can be given by its location. Furthermore, the location of the sensation determines the possession of it. A pain, say, is identifyingly referred to by where it is located. If it is a headache, and the head is A’s head, then the headache is A’s headache. This is used to account for my and your headache. My headache is in my head whereas yours is in your head, and those objects are in different places in physical space. Now it is in fact a mistake to believe that when a sufferer of a pain locates her pain, by saying for example ‘I have a pain in my arm’, she is locating the pain in physical space. Rather, as Malcolm puts it, what one is doing is to locate one’s pain ‘in a space of sensations that has one’s own body as its frame of reference\textsuperscript{31}. If I say ‘I have a pain in the arm’ and you say the same, then we have a pain in the same place. Now, if I locate my pain, say, in my leg, then we would have pain in different places. What happens is that in saying ‘I have a pain in my arm’ I am not giving a different description of my pain from your description when you say the same. As far as pains are concerned, the way in which we identifyingly describe them, e.g. ‘in my arm’, does not determine a place in physical space, and so when you say it and I say it, we are not talking of different places.

\textsuperscript{28} ‘Since A’s pain is in his body and B’s in his distinct body, they are, by parity of reasoning, numerically distinct’ (Hacker (1993), p.20).

\textsuperscript{29} Cf. Hacker (1993), pp.20-5.


So, it is an error to maintain, as Strawson seems to, that sensations are identifyingly referred to by their location in their owners, which are also identified through their locations in physical space. *Sensations* do not have their identity dependent on the identity of the persons (or bodies) who own them. Rather, they have their identity dependent solely upon their description, including eventually those in terms of places which are apparently in physical space. If your sensation can have the same description as mine, then, for all purposes concerning the identity of sensations, we can have the same sensation. Sensations, as other contents of consciousness as well, seem to have only *generic* identity and not *numerical* identity\(^\text{32}\). That is, sensations have their identity determined by their intensity, phenomenal characteristics, even by their location as indicated, say, by a sufferer of pain. So, two cancer sufferers may indeed have the same pain, as two chairs may have the same colour, two people the same opinion or the same sudden thought. A more perspicuous account, then, gives these matters a different order. To have a pain is not to own anything, it is to be in pain, to suffer, and the person who has a pain, the sufferer, is she who *manifests* pain. Consequently, the concept of the location of pain is in fact dependent on the concept of the sufferer of pain, and the location of it is where the sufferer says it is. As Hacker puts it, the ‘Whose is it?’ or ‘Who is hurt?’ are *prior* to ‘Where is it?’\(^\text{33}\)

The interesting question to ask now is why would Strawson hold the view that experiences are particulars with numerical identity? Why is he led to maintain such a position to achieve his goal of proving the objectivity of experience? It is here that Strawson’s empiricist commitments appear more clearly I believe. In *Individuals*, page 98, he says that denying that we could *refer* to particular states or experiences is just a ridiculous position to hold. So it seems that his view is one about what is *conceivable, intelligible*, to maintain about experiences in general, as it were, a truly *metaphysical* claim. Another way to put this would be to say that he believes that experiences can, insofar as they are particulars, be *named*. Obviously, they would be *dependent* particulars, but nonetheless particulars. However, what informs Strawson’s view of these matters is the ‘principle of significance’. The first thing to notice is that the principle is applied at different levels, being

---

\(^{32}\) Cf. *ibidem*, pp.119-20. In the next chapter attention will be given to thoughts and their identities.

\(^{33}\) Cf. (1993), p.22. The point about the right order will be very important concerning thoughts below.
pervasive in Strawson's philosophy. Above I called attention to how Strawson deals with the problem of maintaining, on the one hand, 'that peculiar non-transferable kind of possession' of experiences, and, on the other hand, that we do not thereby create a problem of other minds. Strawson believes in 'a very central thought': 'viz. that it is a necessary condition of one's ascribing states of consciousness, experiences, to oneself, in the way one does, that one should also ascribe them, or be prepared to ascribe them, to others who are not oneself'\textsuperscript{34}. We are here talking about 'minds', and in this connection we have one of the first appearances of the principle of significance: the applicability of predicates ascribing states of consciousness defines a major logical type or category of individuals. Now this is the case because of a 'purely logical' point: 'the idea of a predicate is correlative with that of a range of distinguishable individuals of which the predicate can be significantly, though not necessarily truly, affirmed'\textsuperscript{35}. Here we have the typical Strawsonian qualification: we are not concerned with truth, but with what is significant to affirm. The question, however, is: is the idea of a 'predicate' always correlative with that of a range of distinguishable individuals? As we saw above, concepts \textit{in general} were thought legitimate and even meaningful only insofar as they were related to empirical or experiential conditions of application, but is this correct? We also saw that Strawson appealed to a 'general rule' to the effect that even in self-ascription of experiences we use the concept of a subject of experience as one which necessarily requires the identity of such a subject as 'an object of outer sense'.

In relation to Kant we saw that the belief that the Paralogisms criticized the Cartesian account of the mind by appealing to a principle of significance is not correct\textsuperscript{36}. Where this leaves Kant's positive views about persons is still an open question, especially in connection with Strawson's account of the meaning of this concept.

It is certainly not the case that Strawson's position has empiricist motivations merely because it involves treating experiences as particulars and the self as an object (or 'I' as

\textsuperscript{34} (1959), p.99.


referential *simpliciter*). Cartesianism has similar claims. The distinguishing feature of Strawson’s position is *how* it purports to make references to experiences and make the self meet ‘the most stringent critical requirements’, which are both assumed to involve the ‘absolutely firm basis’ of ‘object-reference’. It is in this connection that Strawson introduces *empirical* criteria for the numerical identity of persons and, as I tried to show, of experiences. As for the self, its concept will have a meaning for Strawson if it stands for an object in the world of physical space accessible through *sensible intuition*, that is, *empirical awareness*. Strawson uses this approach for whatever concept he is dealing with, even the concept of experiences in general. So, conceiving of experiences as particulars with their numerical identity determined by their location in objective space allows Strawson to use their concept meeting rigorous standards. The experiential or empirical conditions required by the talk of experiences admittedly have to involve the recognition that they are dependent particulars. However, location in an objective physical space which is available for empirical awareness gives this talk its non-objectionable objectivity. I see in such a move the ancestry of Locke’s and Hume’s empiricism. We saw, nevertheless, how the approach gets things in the wrong order, in terms of how experiences are referred to and located.

I want to conclude with a remark about Kant. It is the case that Kant requires that a concept must have a sensible application if it is to have objective reality, which is conceived of as empirical significance, capable of leading us to knowledge. I am not sure whether this part of Kant’s view is captured by Strawson’s ‘principle of significance’, or by Bennett’s ‘concept empiricism’. Because, as Allison warns the readers of Kant, this does not

---

37. Evans’ (1982) philosophical project seems to involve arguing for a ‘realist’ framework (for the theory of thought in general) purged from verificationism (in fact, Dummett’s ‘ideal verificationism’, cf. pp.93 and ff.). ‘Verificationism’ is understood as making the claim that the fundamental mode of identification of an object is the demonstrative mode, and that this involves being in the position to determine the applicability of a basic range of decidable properties to the object. Concerning the self and experiences in general, Evans will develop the Strawsonian approach, using even an inherited ‘generality constraint’. The self will continue to be an object and experiences will be among other things particular informational and cognitive states (cf. Chapter 7, pp.205-66). Against Strawson, Evans will argue that it is misleading to restrict the immunity to error through misidentification to claims to knowledge involving only mental properties. I do not know if Evans succeeds in developing his views along these lines, but then the motivation behind his project is not clear to me.


account for the full breadth of Kant's dealings with 'possible experience'. This notion involves, instead of hypothetical perceptual episodes in the history of some sentient being, or even the appeal to empirical laws, a link with formal, *a priori*, conditions in Kant's hands. Even when Kant seems to be closer to verificationism, as in the Antinomies for example, his main point is to show how certain views neglect or completely misunderstand these conditions; the claims they end up making are in fact *not* unverifiable, but are rather futile.\(^{40}\).

---

CHAPTER IV

KANT ON SELF-CONSCIOUSNESS: CONTEXT AND CONSTRAINTS

1. Introduction

It is sometimes claimed that the main goal of the Transcendental Deduction of the categories is to prove against the sceptic the possibility of objective knowledge. It is from this perspective on the Deduction that, in recent years, a discussion started on the nature and powers of transcendental arguments. When the discussion is kept within Kantian territory, one notion is acknowledged to be essential to the workings of such a kind of argument: the notion of self-consciousness or of transcendental apperception. I am not going to deal here with Kant’s arguments against scepticism. I will not even discuss whether this is the main objective of the Deduction. So, I will not be interested in the objectivity of knowledge as an issue between Kant and the sceptic.

I want to concentrate my efforts on getting a clearer picture of what Dieter Henrich called Kant’s ‘foundational thoughts’. This author remarks that Kant uses his analysis of self-consciousness as ‘the supreme principle to which, in his opinion, all forms of rationality (“employment of the understanding”) can be traced back’. Kant’s analysis of self-consciousness is the core of his ‘foundational thoughts’. The trouble is that, according to Henrich, this central part and the text dedicated to it (the Deduction) ‘could only be given the attention and the space warranted by an important auxiliary theorem; and this only on account of what it achieves, not for its own sake’

I think it is reasonably clear to any reader of the first Critique how Kant wanted to apply his critical principles to the theoretical tradition he inherited. What is certainly not so clear are the foundations which were to yield the critical instruments Kant used to examine that tradition. At A106-7 Kant claims that the transcendental apperception is what grounds all necessity. It is said to be the original and transcendental condition which makes possible the unity of consciousness present in the synthesis of the manifold of all intuitions and in the concepts of objects in general. This turns it, in fact, into a special condition of all the objects of experience. The notion of transcendental apperception is, therefore, obviously crucial; but

unfortunately it is also very obscure. This is why it requires 'the highest degree of interpretation\textsuperscript{2}. So, what I intend is to present a picture of self-consciousness in Kant, a notion which is, without any doubt, central to Kant's critical enterprise. When the notion is approached as designating a human faculty, as the consciousness of self, I am going to claim that Kant's 'analysis' reveals that it has a feature that is most important: self-consciousness is self-sufficient. Even though I will argue this claim only in the next chapter, it is important to make clear right from the beginning what I mean by the notion of the self-sufficiency of self-consciousness. This will be presented later on as one crucial result of Kant's investigation of self-consciousness. Its importance will appear when we come to face the claim that the unity of self-consciousness is the \textit{a priori} ground of all concepts (cf.A107).

Self-consciousness will be shown to be self-sufficient for Kant in the sense that its exercise does not presuppose knowledge of the empirical criteria for the identity of a subject of experience or person. This is equivalent to saying that I-thoughts do not presuppose for their mastery knowledge of the empirical criteria of identity on the basis of which we employ, for example, terms like 'she', 'he', 'you'. This is not to deny, however, that I-thoughts presuppose awareness of 'she', 'he' and 'you'; rather, 'I' is different from them as regards its use by a subject precisely insofar as its mastery does not require knowledge of empirical criteria for the identity of a subject of experience.

As is clearly noticeable, my working explanation of the self-sufficiency of self-consciousness is in terms and ideas first presented in relation to Kant by Peter Strawson. I would like to explore further (cf. also the previous chapter) some of Strawson's observations on Kant which bear on the transcendental apperception, with the aim of making clear how my view differs from his despite being couched in his jargon. The shortcomings of Strawson's view of self-consciousness, which he presents as a rational reconstruction of Kant's, will be discussed only in chapter 5, section 2.

Firstly and foremost, one should not, according to Strawson, identify transcendental self-consciousness with empirical self-consciousness. Because the former is only a kind of abstract or philosophical consideration of the only satisfactory notion of self-consciousness, \textit{viz.} the latter, which is our ordinary empirical self-consciousness\textsuperscript{3}: the self-ascription of

\textsuperscript{2} Cf. Henrich, \textit{loc.cit.}.

diverse states of consciousness or the awareness of one's inner states. In this respect, transcendental self-consciousness represents only the necessary conditions for the possibility of a satisfactory notion of self-consciousness, it does not include the sufficient conditions for the explanation of the actual occurrence of self-ascription of experiences. It represents the requirement, stated in general terms, that experiences, if they are to be experiences, have to be conceptualized 'as to determine a distinction between the subjective route of his [the subject's] experiences and the objective world through which it is a route'. So, transcendental self-consciousness represents only the requirement that the thought of a 'temporally extended point of view on the world' be possible. This is the thought of an abstraction and we are in need of much more besides if we are to be able to explain fully the notion of self-ascription of experiences. But the requirement, nonetheless, states the 'more fundamental part of those full conditions', 'the essential core of personal self-consciousness', 'the fundamental basis of the possibility of self-ascription of experiences'. For without that distinction and this thought we would not have the space to think of experiences as experiences, that is, as items which represent how things are but that are independent of those things, as items which are not absorbed into the things they represent. And without such a notion of experiences, talk of self-ascription of experiences would be rendered dubious.

However, in second place, the context in which the notion of transcendental self-consciousness appears in Kant, according to Strawson, is one of examination of the conditions of 'a possible experience in general'. Now, for Strawson (even though he discusses it as a possibility, it represents his position), these conditions include 'the

---

5. Cf. op.cit., p.103.
10. Strawson talks about this latter point as 'the necessary reflexiveness of experience' (p.111), as 'the necessity of providing room, in experience, for the thought of experience itself' (p.110, cf. also p.107).
satisfaction of the full conditions for the empirical self-ascription of experiences, \textit{i.e.} the existence of empirically applicable criteria for the identity of a subject of experience\textsuperscript{11}. This is taken to imply that 'the requirement of transcendental self-consciousness [the requirement expressed by the abstract or philosophical consideration of the empirical self-consciousness] is derived from the requirement of the possibility of empirical self-ascription, and ultimately derives its intelligibility from the latter'\textsuperscript{12}. What Strawson has in mind here can be gathered from what he excludes: to imply, as he does, that the conditions of a 'possible experience in general' do include the satisfaction of the full conditions for the empirical self-ascription of experiences is to \textit{exclude} the possibility that the requirement of transcendental self-consciousness is derived solely from the conditions of a possible experience in general according to which self-consciousness is \textit{intelligible independently of the empirically applicable criteria of the identity of a subject of experience}\textsuperscript{13}.

So, \textit{pace} Strawson's own admonition not to identify transcendental self-consciousness with empirical self-consciousness, he is willing to maintain that the very intelligibility of transcendental self-consciousness is \textit{dependent} on 'the empirically applicable concept of the identity of a subject of experience'. This means that the requirement which was referred to as an \textit{abstraction} has its rationale in this other conception. That is, the necessary reflexiveness of experience, which is said to be identical with 'the necessity of saving the recognitional component in an experience from absorption into its sensible accusative (and thereby saving the status of the experience as experience)'\textsuperscript{14} is \textit{dependent} on the conception of the identity of a subject of experience. This dependency is explained by Strawson as follows.

Experiences are experiences, which are independent from the objects they are experiences of, only because they are \textit{self-ascribed} items. Self-ascription, however, can only be achieved by a subject of experience when she knows that by using 'I' in I-sentences she is making reference to herself as a subject which is identifiable as a corporeal object among

\textsuperscript{12} Loc.cit.
\textsuperscript{13} Cf. loc.cit.
\textsuperscript{14} Op.cit., p.110.
corporeal objects in the world, as a 'man among men'. Only then can the subject be said to be ascribing an experience to something: in this case, to a particular corporeal object in the world which is also a person. If we consider that I-sentences are the expression of self-ascriptions of experiences, Strawson makes the same point by stating that 'our personal pronouns, the pronoun “I” included, have an empirical reference; and in some way such a reference must be secured if the general notion of ascribing experiences to a subject of them is to make sense'\(^\text{15}\).

'I' must, then, have an empirical reference, and this reference must be secured. It is secured, according to Strawson, because the subject who uses 'I' is an object we can identify as one corporeal object in the world among others. Quassim Cassam makes the point on Strawson's behalf as follows:

Successive perceptions are conceptualizable as perceptions of the very same thing precisely because of the possibility of ascribing those perceptions to a numerically identical subject whose route through the world anchors those perceptions to one particular object.

The identity claim needs to be anchored, and... it is the numerical identity of the subject of diverse experiences which provides the anchoring. And for the numerical identity of the subject to play this part, those experiences must be thought of as having a subject.

Diverse experiences may be conceptualized as elements of a single, subjective, experiential route because of the possibility of appealing to the notion of an enduring subject of those experiences, a physical thing located in space\(^\text{16}\).

I am now in a position to state how my view about self-consciousness in Kant differs from Strawson's. Whereas Strawson requires that 'I', to be meaningful, must be used by subjects as involving a claim to a type of knowledge - viz. of herself as an empirical subject of experience, which involves empirical criteria of identity of an 'enduring subject', a 'physical thing located in space' - I do not require that. 'I' is meaningful independently of this knowledge. What the meaning of 'I' is, in an account which respects this feature of independence just pointed out, is a task for when we get to deal with Kant's view in chapter 5, section 3. The point I want to make here is that transcendental self-consciousness is a faculty in Kant which does not involve the notion of a 'subject of experience' in the way


\(^{16}\) Cassam (1989), p.92 for all the quotes.
Strawson envisages, viz. as a subject which in I-thoughts knows herself as a self-ascripter of experiences insofar as she knows she is an corporeal object among others, a physical thing in space.

For one to state that the human faculty of self-consciousness is self-sufficient for Kant in a clear and justified way, one needs to make clear a set of notions that, as in a web, are interconnected with self-consciousness in Kant's thinking. To establish what the meanings and purposes of these notions are is the main task of the present chapter. Thus, I could say that the purpose of this chapter is, put in general terms, to examine the context of Kant's view of self-consciousness. I will show that the context in which Kant sets his analysis of this crucial conception of his allows him to present constraints on the task of producing a philosophical clarification of self-consciousness. Let me state how I plan to clarify the context and the constraints Kant sets to his investigation.

In section 2, I aim to point out that Kant's central question in the Deduction is that concerning the conditions of representation. The section does the negative job of precluding a tempting - at first approximation - interpretation of Kant's concerns with conditions of possibility: Kant would be engaged in some sort of psychological investigation about how the understanding conditions our knowledge of objects. The section leaves the way open to a different and more fruitful approach, or so I will claim.

Section 3 has as its goal to offer a construal of Kant's problem in the Deduction as one about the intentionality of thought. I explain, in the first place, what this problem is; then, in the second place, I show why Kant when properly understood is addressing himself in the crucial text of the Deduction to this problem as well. I will hope in this section to defend further this approach to Kant with a discussion of his 'critical problem'.

Section 4 deals with a principle Kant says constraints the whole enquiry in the Deduction: categories must be recognised to be a priori conditions of experience. Given this recognition, it must be explained how the categories do manage to be these conditions. This constraint upon the enquiry has a justification in Kant: categories are such conditions because meaning is autonomous. To the explanation and justification of this tenet section 4 is chiefly dedicated. The result of this section is, however, in the main negative, but it will allow me, at least, to point at the right direction in my search for an argument to the effect that self-consciousness is self-sufficient. More about the relation of these two topics will be said in section 3 of the next chapter.
Section 5 aims to show that what Kant understands as 'transcendental synthesis' is what makes possible an intentional thought or representation. In Kant’s terms: a judgement. This is Kant’s positive view characterized in an abstract form. The issue of 'synthesis' in Kant will, however, return in the ascription of externalism to his philosophical position in more detailed a manner in the next chapter (section 6).

Section 6 starts with the discussion of Kant’s idealism. Such a discussion is necessary because the way Kant envisages synthesis to solve the problem of how representing is possible is linked to the issue of a wider philosophical position: in Kant’s case, his Transcendental Idealism. In this section, I offer a rational reconstruction of Kant’s general philosophical outlook and thus far engage in some Kantian scholarly debate. From my reconstruction I then derive a view of the content of representations which I call the 'transcendental view'. Its main feature is that it does not need to postulate any intermediary between the representing subject and the represented object.

Section 7 concludes the chapter with an illustration of how a transcendental view of content offers to solve in contemporary debate the troubles with content and intentionality. I should here state that what I will designate as ‘a transcendental view of content in contemporary debate’ is in spirit and in many of its elements a Wittgensteinian view. So, my reconstruction of Kant’s positions is in my own eyes Wittgensteinian. I shall hope that it appears clearly in the text.

2. The Question about the Conditions of Representation

What is the purpose of the Deduction of the categories? In Kant’s own terms, it is to show that the categories have a priori objective validity, that is, that they are the fundamental concepts without which we could not think, and a fortiori know, any object (cf. A111). In the A-Deduction’s ‘systematic’ presentation of it, it becomes clear that with this purpose the Deduction constitutes an argument about the understanding, since it attempts to clarify the necessary connection with, or relation in which the understanding stands to, appearances by means of the categories (cf. A119, also A128)^17.

---

^17. At A85/B117 the purpose is explained through what is duly entitled the transcendental Deduction of the categories, viz. 'the explanation of the manner in which concepts [the categories] can thus relate a priori to objects'.
These statements already give rise to many queries. I would like to point out three, and organize my discussion of Kant’s intentions with the Deduction by addressing them in turns. In the present section, I will deal only with one of them (the second). I aim to make clear that Kant’s problem concerns the conditions of representation in a pervasive way: it should not be taken as the worry about how the understanding conceived as a cognitive faculty contributes in an \textit{a priori} way to the production of a mental entity, a cognition, which could be called ‘experience’. Such a view I will designate the ‘official view’ of Kant’s intentions with the Deduction: it construes Kant’s concern with ‘conditions’ psychologically.

Firstly, is there any special significance connected with the expression ‘objective validity’ in the claim that the categories have \textit{a priori} objective validity? How should this expression be understood?

Secondly, when I say that the categories will be shown to be concepts without which we could not think \textit{and a fortiori} know any object, this could be taken as suggesting that these two human capacities are not to be distinguished in Kant. It would be remarked that Kant makes the distinction in several places (cf. B146 and ff.; B165-6, including note a; A155-6/B194-5). I accept that Kant makes such a distinction. What I do not accept is how the ‘official view’ construes the distinction and through it attempts to explain the goal of the Deduction. According to this view, ‘to know’ requires, whereas ‘to think’ does not, the givenness of the object through intuitions which \textit{qua} data coming from the object to the mind have to correspond to the object’s concept for ‘knowledge’ to be achieved. With ‘thinking’ the only data our mind needs to deal with are concepts, which are representations in it. This view of the distinction can also be phrased as follows: ‘to know’ is a ‘cognitive task’ whereas ‘to think’ is a ‘subtask’ of it. Knowledge or ‘cognition’ involves intuitions, whereas thinking, as merely representing, does not\textsuperscript{18}. Accordingly, the Deduction is concerned with

---

\textsuperscript{18} Even though Patricia Kitcher adopts a specific cognitivist approach which tries to extract from Kant ‘psychological analyses of empirical capacities’, I take her as a representative of the psychological approach I am calling ‘official’. The above presentation uses her terms (cf.1990, pp.62-6). Kitcher’s views, pace her attempt to rescue \textit{transcendental} psychology, go like this: ‘As the Transcendental Aesthetic examined sensibility to discover \textit{a priorio} [not derived from the senses, \textit{i.e.} of non-sensational origin] elements in cognition, the central purpose of the Deduction is to reveal the \textit{a priorio} contributions of the understanding. It does this by analyzing very basic empirical capacities... in order to determine what faculties we must possess and what processes we must carry out, given that we are able to perform these tasks’ (p.62). ‘To understand how we are able to construct representations of objects... we need to look at both the patterns of our sensory stimulations and the ways the mind uses those patterns to construct representations’ (p.80). Kitcher
the thinking part of the whole process which results in knowledge. I will argue against the 'official view' after presenting query three.

Thirdly, where the Deduction is said above to be about the understanding, what has to be made clear is the necessary connection or relation in which it stands to appearances. We should also find an answer to why there should exist such a connection or relation at all. Furthermore, it is a relation of the understanding to *appearances*: what does this term mean, what is its role here?

We can start answering query two with the unanalysed pair 'representation-object' and ask with Kant which *produces* which (cf.A92-3/B124-6). Kant analyses the two possibilities about the way representation and object can be related. I will deal with them in order. (1) When the object produces (in the sense of causes) the representation, then we are dealing with an *empirical relation* which cannot account for the *a priori* possibility of the representation according to Kant (cf.A92/B125). This negation strikes one at this stage as unsupported. Because Kant is just assuming that the correct kind of approach to the question has to be in certain terms, viz. it has itself to be *a priori*. But then one wants to ask: why should the representation be possible *a priori*? What kind of constraint does Kant's understanding of the correct account of the categories put on any attempt to explain how representations come about? Why should this be so at all, that is, why should there be a correct account of the categories with such implications anyway, if all we want is to explain which - representation and object - produces which?¹⁹

Kant explained before having put this question about 'production' that he would be dealing with the principles of any transcendental deduction, and that he would be concerned with a 'question of right' (a *quid juris* question) that is about the use of *some* concepts we in fact possess (I will deal with this factual claim, and Kant's argument about this issue - the metaphysical deduction - in chapter 6). About the 'question of right' it is crucial to point out that it is a question of legitimacy, and therefore cannot in Kant's eyes be answered empirically, or by the explanation of a *de facto* origination, that is, *a posteriori*. As Kant

---

¹⁹ Compare this with Hume's approach at the beginning of the *Treatise*, where he inquires which produces which among impressions and ideas.

praises Robert P. Wolff for his type of approach to Kant's project (cf. *op.cit.*, p.231, note 5). And one can see that the latter's is a psychological one by checking his account of *mental* propensities in Hume and their equation with Kant's categories (*sic*) (cf. 1968, pp.100, 127-8).
puts it, empirical proofs do not suffice to justify the pure *a priori* employment which the categories are affirmed, rather dogmatically, to have. We should, then, be struck by the way the position is qualified by Kant at this juncture: the *aprioricity* of the categories is a fact which is in need of philosophical clarification. So we can say that the ‘recognised’ *aprioricity* of the categories constrains Kant’s approach in any possible transcendental deduction. This claim determines at the outset how the task - ‘the whole inquiry’ - is to be undertaken: according to a principle that requires that the categories be ‘recognised’ as *a priori* conditions of the possibility of experiences (cf.A94/B126). This is a very important point in Kant and it will receive separate treatment in section 4 below.

The second possibility Kant analyses is the following: (2) The representation makes the object possible. Kant immediately blocks a tempting but rather out-of-context interpretation of this possibility: we are not dealing with how the will produces or causes (and this is an obscure notion here) its object (which is an obviously ambiguous term in this context). What Kant wants to preclude with this point are cases in which, say, I form the intention to play football next Saturday and, as Saturday comes, I go to play football. This seems to fit Kant’s model of the causality of the representation (*sic*), which is exerted by means of the will (which alleviates things a bit, given that we are not in this case creating any normal physical object), whereby the *existence* of the object (my action) is produced. What Kant seems prepared to exclude as irrelevant here is a *willing* relation between representation and object, which could have as a model the formation of intentions and the carrying out of them²⁰. He wants, in this case, to restrict himself to the *representational* relation between a pair composed of representations and ordinary objects. So, concerning the representational relation between representations and objects, Kant makes an initial claim: we are not considering how the representation *produces* its object - as it were *representationally* - insofar as its *existence* is concerned. Such a view, of representations *representationally* (through the means of the process, if it is one, of representing) producing the *existence* of their objects would require the conception of an intellect with the greatest and strangest of powers. I think Kant’s philosophical fiction of an *intuitive* intellect or understanding (cf.B135, 138-9) can help us to appreciate what he was driving at.

---

The intuitive intellect or understanding is a faculty which by thinking alone creates or brings into existence the objects it is thinking of. This faculty gives itself its manifold. According to Kant, the human intellect or understanding is not like that, however; in contrast to the former, it is a limited faculty, and its manifold has to be given. This amounts to saying that our intellect or understanding is irremediably discursive, which means that the human faculty by thinking alone can only be dealing with concepts, that is, thinking for us is a conceptual activity. Humans, by thinking, can only be related to objects mediately: by means of the features that objects have in common and which are present in thinking as characteristic marks of a concept (merkmale). In thinking, then, we do not perceive any object; what we do is to use concepts ‘to judge by means of them’ (cf.A68/B93). Thus Kant’s view can be summarised as follows: in thinking we can only exercise our understanding as a faculty of judgement (cf.A68/B93 and ff.).

These considerations lead Kant finally to present the option he wants to consider: ‘how representation is a priori determinant of the object’ (A92/B125). This is the case when ‘only through the representation is it possible to know [N.B.] anything as an object’ (ibidem). Kant also puts the question he wants to answer mainly in the Deduction like this: ‘How is nature possible in the formal sense, as the totality of the rules under which all appearances must come in order to be thought as connected in an experience?’ And he completes: ‘The answer must be this: it is only possible by means of the constitution of our understanding, according to which all those representations of sensibility are necessarily referred to a consciousness, and by which the peculiar way in which we think (viz., by rule) and hence also experience [N.B.] are possible, but must be clearly distinguished from an insight into the objects in themselves’

These are quite general presentations of what Kant intends, but ones which nonetheless, I think, allow me to say how one should interpret passages where Kant states his views a bit differently, passages from which, however, interpreters normally derive the ‘official view’ of thinking and knowing in the context of the Deduction. My view is that thinking and knowing involve a distinction which is extremely important in a rather different way.

First of all, in passages where we could expect Kant to limit his demands to the way representation is determinant of how we think the object, he presents his point with the generalized claim about the possibility of knowing anything as an object. Given that this knowing must obviously include intuiting, this formulation implies that the way representations condition objects is more far reaching. Secondly, in the very place where he connects the categories with the conditions under which something is ‘if not intuited, yet thought’ as object, he follows this up with the qualification that ‘all empirical knowledge of objects would necessarily conform to such concepts [if we can deduce the categories], because only as thus presupposing them is anything possible as object of experience’ (A93/B125-6). This should convince one that Kant is concerned with concepts - in their roles as conditions - in relation to all empirical knowledge of objects, concepts which when shown to be these conditions would certainly figure in Kant’s account much more importantly than the ‘official view’ of the categories would allow for.

Further evidence to Kant’s real point concerning ‘to think’ and ‘to know’, especially with the background of the Deduction in mind, can be gathered from another text, which I believe further undermines the ‘official view’. The text appears in the preparatory remarks to the argument. ‘The objective validity of the categories as a priori concepts rests, therefore, on the fact that, so far as the form of thought is concerned, through them alone does experience become possible’ (A93/B126). The reference to the ‘form of thought’ should not be understood here as a remark about, again, the psychological element of thinking supposedly demanded if experience, as a kind of mental entity, should be produced. A reference to form in Kant always has a link with what is necessary and a priori, and not with the product of an eventual empirical generalization or necessity22.

This crucial meaning of ‘thought’ in Kant, which relates so importantly to the purposes of the Deduction, has its best explanation, I believe, in the Paralogisms B. There Kant clarifies one meaning of ‘thought’ as that which relates to things: that which can be

22. Cf. Allison (1983), pp.3-4, 10-3, especially p.11; also pp.26 and ff. Most of what Allison means by his ‘epistemic conditions’ would capture what I want to convey with ‘form’ here. I consider to be very important to argue this point, because it amounts to a refutation of the view that Kant would be interested in psychological conditions as the ‘official view’ has it. The theme of the relation of ‘form’ with apriority and necessity, especially with an eye on the nature of the latter, will remain the focal point of interest. Allison’s views in this respect are in my opinion on the right track, cf. also op.cit., p.7.
thought is 'a being that can be thought in general, in every relation, and therefore also as it may be given in intuition' (B411). In this sense, what can be 'thought' in general designates 'an object in general and therefore... an object as it may be given in intuition' (B411n). Kant is here talking of things as we normally understand this notion, which he claims cannot be represented otherwise than as subjects (cf. ibidem).

So, we should recognise that Kant is concerned with the problem of representing in general, and not with the restrictive, psychologistic notion of thinking as the giving of shape to a mental item received through the senses. This is important for the appreciation of what Kant can be trying to achieve with the Deduction. This should not imply, however, that the understanding is not responsible for thought with us humans.

3. The Problem as Concerning Intentionality

We have already had indications that Kant approached his task in the Deduction in a special way. He in fact adopts a principle to constrain every possible attempt at a deduction: 'The transcendent deduction of all a priori concepts has thus a principle according to which the whole enquiry must be directed, namely, that they must be recognised as a priori conditions of the possibility of experience, whether of intuition which is to be met within it or of the thought' (A94/B126)\(^{23}\). Why is it that Kant adopts a principle such that we recognise the categories are necessarily the conditions for the possibility of experience? This point obviously bears on the necessary relation Kant says he is exploring in the Deduction between the understanding and appearances, which was my third query in the previous section. I shall propose in the present section an understanding of what Kant's problem about representing amounts to and show how this appears in Kant's texts.

In reading Kant's statements in the preparatory parts of the Deduction (mainly sections 1 and 2 in the Deduction, but not only there), we should realize that Kant's problem is the problem of the intentionality of thought\(^{24}\). I shall explain what I mean by this expression, then why I believe this is Kant's problem.

---

23. Cf. Allison (1983), pp.10-1 for the argument in favour of the broader notion of 'epistemic condition', which captures the generality I have been advocating and which I emphasized in this quote. Cf. ibidem also for subjectivism as an interpretation of the conditions of representing in Kant.

24. Cf. Kitcher (1990), pp.65-6, for the same statement of Kant's problem; her account of Kant's solution is, as pointed out, different from mine, being functionalist. I believe Guyer also formulates just - Kant's problem in this way: 'the problem of how representations which must somehow be
First of all, why is intentionality a problem?

We can think of perception as the result of the process by which objects in the world present themselves to us. As Kant says in the Critique - when talking of 'Logic in General'- about 'intuitions': through them an object is given to us. Now, when this is intended to apply to empirical cases, we are dealing, according to Kant's unqualified introductory remarks, with 'sensations' 'which presuppose the actual presence of the object' (A50/B74). So, we can state on Kant's behalf a simple view of perception which intends to clarify how we manage to represent things which present themselves to us: things in space and time are given to us insofar as we have perceptions of them, which are representations accompanied by sensations. Thus far we manage to have empirical representations of objects. (cf. B147; these remarks on perception do not intend to capture Kant's critical position, rather they try to present a common manner of thinking about the matter using some Kantian terminology, a manner which is taken in its simplicity as non-problematic).

Now, it also seems perfectly acceptable that we represent in thought objects which are not at the moment present to us, as was the case of perception above. This, however, may trouble us when we come to think of how this is possible at all. We quite normally take thoughts to represent or stand for objects and situations in the world. One way of reflecting about this and finding it troubling is to reason that as I am here, in this part of the world, and what I am thinking about is somewhere else, perhaps very far away indeed, how do I manage to think of or represent something that is, say, on the other side of the world?

Another way of finding complications in what at first looks very straightforward is to think of representations and their relations to truth and falsity. Suppose I think or utter 'Brazil will be the next football World Champions'. Now suppose, when the 1998 Championships in France arrive, England are the champions. What I thought then is false, that is, what I thought did not become the case (I am not worried about Future Contingents and Fatalism here). But how can my thought represent a state of affairs which does not obtain, represent something which does not exist? It seems that meaningfulness as such requires that a representation should represent something, what is it if what it says is, or will be, the case is not, or will not, be the case? If it is nothing at all, then it seems the

connected with the nature of the cognitive subject itself, in order to be known a priori, can also provide insight into the objects which exist independently of this subject of knowledge' (1987, p.11).
representation must be meaningless. However, it is quite obvious that even false representations are meaningful, and that their meaningfulness is not imperilled by mere falsehood. Indeed, how else could they be false? But how is the desirable neat connection between these notions of meaningfulness and falsehood to be established?

Still another reflection concerning intentionality which may puzzle us puts the issue in very general, but important, terms: how can thought 'reach right up to reality'? This is a problem, for 'if my thought is true, what I think must be identical with what is the case, but if it is false, it cannot be; yet the content of my thought is the same in both cases'.

This is the general problem of how representations work qua representations, how is it that they really are about objects in the world, in such a way that we can say what is a fact and what is not. This is a formulation of the problem of how thought relates to reality. Now, I believe that this is Kant's problem as well.

At A104 Kant worries about the meaning of 'an object of representations'. This issue is presented by him through the perspicuous question: 'What, then, is to be understood when we speak of an object corresponding to, and consequently also distinct from, our knowledge?' (loc. cit.). This way of presenting the problem stresses the real difficulty, that an object which is distinct from ourselves and our representations shall nonetheless correspond to them deserves a philosophical account. It is this distinctness and correspondence of representations that I tried to make palpable in the reasoning I suggested may trouble us when we think about representing what is not present or does not even exist.

25. This is Wittgenstein's approach in the *Tractatus*, cf. propositions 2.1511 and ff..


27. Cf. Allison's view that it is in virtue of 'epistemic conditions' (which include the categories) that our representations relate to objects (1983, p.10). Cf. also pp.29, 72, 342n.

28. Cf. A129: 'And if we derived them [the *a priori* concepts] from the self, that which is merely in us could not determine the character of an object distinct from our representations, that is, *could not be a ground why a thing should exist characterised by that which we have in our thought*, and why such a representation should not, rather, be altogether empty' (my italics: AK). This, for sure, expresses the problem of how I manage to think of a particular something in a particular manner: why should a particular something be characterized somehow by thought? Why should a thing exist characterized like this? If it can not, then our representations are altogether empty: we are not thinking of anything, nor thinking anything, which are not forms of thinking at all. The clearest statement is perhaps, with the enunciation of Kant's solution, A197/B242-3.
Kant is to my mind quite clear about his problem being what I have called the intentionality of thought in his ‘critical’ letter to Marcus Herz [21/February/1772]. The problem, however, which I passed over in silence [in the non-critical Inaugural Dissertation of 1770] is how, then, a representation which is related to an object can otherwise possibly exist, without being affected by it in some way. I had said: the sensible representations represent things as they appear; the intellectual representations represent things as they are. But by what means are these things given to us, if it is not by the way they affect us; and if such intellectual representations rest on our inner activity, whence comes the agreement which they are supposed to have with objects, the objects not being originated by this activity; and whence is it that the axioms of pure reason concerning these objects agree with them, without this agreement being permitted to derive assistance from experience.

...The question is, how is the understanding to construct for itself entirely a priori concepts of things, with which the things are necessarily in agreement; how is the understanding to draw up real fundamental principles about their possibility, with which experience is necessarily in faithful agreement, and yet are independent of it; - this question always leaves behind an obscurity with respect to the faculty of our understanding: whence comes the agreement with things.

In this quote we find, I believe, Kant couching the problem of intentionality in his terms. If representations are to ‘reach right up to reality’, to represent things as they are, ‘whence comes the agreement’ which they are supposed to have with things, given that these things (the reality) are not being originated with the representation. Reaching up to reality is an achievement on the representation’s part which Kant characterizes as an agreement with things, and he asks about it whence it comes from, specially if we bear in mind that ‘the object [is not] the cause of the intellectual representations in the mind (in sensu reali). What is particular to Kant is that he formulates this problem of the intentionality of thought most forcefully in terms of the categories alone. That is, Kant is prima facie only concerned with this feature of his ‘categorial concepts’. However, this is not exactly how it is, given that, paraphrasing Kant, the major point is in principle applicable to other concepts (as in the Prolegomena, the point applies with universality in the mathematical sense: ‘for that which is sufficient for all cases’). So, Kant’s central issue is not only, as it were, the

---

29. Cf. op.cit. in Kant (1968), pp.112-3, Ak X125-6.

30. Cf. for this rendering of Kant’s text Arnulf Zweig (Kant, Philosophical Correspondence, 1759-99), pp.71-2.

intentionality of the categories, but of concepts which have a special role (categorial) in
general, even though he does not state it like this and says things which suggest a more
circumscribed interest. This point of principle can be brought out with a brief discussion of
Kant's 'critical problem'.

In the Introduction to the Critique (cf.A10; B19) and in the Preamble to the
Prolegomena (cf.AkIV276) Kant presents his problem in philosophy to be 'how are a priori
synthetic judgements possible?' This presentation of the problem is potentially misleading
insofar as it suggests that the empirical judgement, the synthetic a posteriori judgement,
presents no problem at all for a critical philosophy, as if the real question concerned
exclusively the justification of the synthetic a priori judgement.

Allison goes some way in clarifying the import of Kant's favoured formulation of
the problem. He states that 'the “critical problem” of the synthetic a priori is really
equivalent to the problem of the relation between pure concepts of the understanding and
objects as Kant presented it in the famous letter to Marcus Herz of February 21, 1772'.
Now, how far Allison really goes depends on how much he thinks the formulation of the
letter of Herz encompasses, and this I will not address here.

Now, someone who does understand Kant's 'critical problem' in all its complexity
is, I believe, Buchdahl. This author thinks that Kant's favoured formulation of the problem
is far too restrictive, and that, as the critical solution implies, the problem is actually much
broader.

Buchdahl sees Kant's problem as the ubiquitous problem of modern philosophy, i.e.
to explain the connection between the terms of propositions, that is, 'the propositional link'.
The problem is the a priori link that Kant, inheriting the discussion mainly from Leibniz,
believes must 'in some sense' be built into any proposition whatsoever, be it a general or
particular, necessary or contingent one. According to Buchdahl, Kant faces this problem
'critically', i.e. displaying the link as 'a required ingredient of the analytic framework of
experience'. 'So Kant's problem becomes the elucidation of the propositional connective...
in terms of the transcendental framework of experience; and in particular, of the principal
categorial concepts (e.g. substance, causation, interaction) which thus come to be explicated
with equal generality, as conditions of the possibility of something being an object, or of

Chapter 4

being objective, part of a public language, *i.e.* of empirical, or contingent propositions, be they general or particular, contingent or law-like. Buchdahl then adds a remark which puts Kant's problem into its proper context:

That this is the logical situation, Kant occasionally registers explicitly. Thus in a crucial passage of the transcendental deduction (#19) he says that the 'necessary unity' of the elements of the judgement is postulated as holding "even if the judgement is itself empirical, and therefore contingent" [B142]. This fact is concealed from the reader of the "Introduction" to the *Critique*, since Kant there talks as though the empirical judgement (which is synthetic *a posteriori*) posed no problems, contained no 'necessary connection'; as though only the synthetic *a priori* judgements require a special defence.

Buchdahl is right when he stresses this 'broader' interest on Kant's part. Kant himself implies this when he states that 'the explanation of the possibility of synthetic judgements is a problem with which general logic has nothing to do. It need not even so much as know the problem by name. But in transcendental logic it is the most important of all questions; and indeed, if in treating of the possibility of synthetic *a priori* judgements we also take account of the conditions and scope of their validity, it is the only question with which it is concerned' (A154/B193).

The relation of this understanding of the 'critical problem' with the intentionality of thought can be clarified as follows. Whether a judgement can be said to be knowledge depends on whether the world is or not as the judgement states. Thus far, knowledge depends on how things are, on what is a fact or what obtains. However, judgements must be

33 . (1969), pp.474-5, my italics: AK. Cf. op. cit. for all the other points. In order: pp.11-2 and chapters 1 and 2; pp.76-8; p.474. Cf. also Russell's connection of the issue of how *a priori* knowledge is possible with relations, that is, links in propositions (1980, pp.50-1): in 'I am in my room', 'in' has meaning and stands for a relation which does not, however, exist as I and my room exist.

34 . Op. cit. p.475n2. Cf. also Allison's remark in the Introduction to *On a Discovery*, p.54: 'The crucial question underlying the whole analytic-synthesis distinction is... the material or transcendental question as to whether or not the predicate stands in a real relation to the object. ...Real relations... concern precisely the content of thought. They hold between things thought about, or between concepts and things'.

35 . Cf. Russell's account of the 'critical problem': 'The thing to be accounted for is our certainty that the facts must always conform to Logic and Arithmetic'. Thus far Russell is right. He misunderstands Kant though, as when in the sequence he says: 'To say that Logic and Arithmetic are contributed by us does not account for this. Our nature is as much a fact of the existing world as anything, and there can be no certainty that it will remain constant' (1980, p.49).
about something, have an object. A belief is always a belief that something is the case, or is not the case. Similarly, an expectation is always an expectation that something will be the case, or will not be the case. Now, we can only know this something or the object if we can represent it. This is not meant to preclude that whether a belief is true or an expectation fulfilled is determined by how things were or are, the world was or is, or how things or the world will be.

Kant’s question is consequently this: how do we manage to represent this something or the object? And this representing is necessarily also present in a judgement which is synthetic a posteriori. Thus Kant asks: ‘What, then, is to be understood when we speak of an object corresponding to, and consequently also distinct from, our knowledge?’ (A104). And adds: ‘Now we find that our thought of the relation of all knowledge to its object carries with it an element of necessity; the object is viewed as that which prevents our modes of knowledge from being haphazard or arbitrary, and which determines them a priori in some definite fashion. For insofar as they are to relate to an object, they must necessarily agree with one another, that is, must possess that unity which constitutes the concept of an object’ (A104-5).

We can also phrase this question in light of Kant’s discussion in paragraph 19 of Deduction B: how do judgements have objective validity, which is not knowledge yet, but which has to pertain even to empirical or contingent judgements? This concerns the necessary relation judgements have to objects or to something which they represent.

4. The Autonomy of Meaning
To the Deduction is given central stage in the Critique (cf.Axvi). This is so also because there, in the Deduction, a crucial aspect of Kant’s philosophical outlook is brought to the fore: the autonomy of meaning. I intend to explain this notion through the discussion of an important passage in the Deduction (in the first edition version). This discussion should illuminate a point already touched upon: the requirement that the enquiry of the Deduction should be conducted under the recognition that the categories are a priori conditions of the possibility of experience (cf.A94/B126).

That a concept, although itself neither contained in the concept of possible experience nor consisting of elements of a possible experience, should be produced completely a priori and should relate to an object, is altogether contradictory and impossible. For it would then have no content, since no intuition corresponds to it; and intuitions in general, through which objects can be given to us, constitute the
field, the whole object, of possible experience. An \textit{a priori} concept which did not relate to experience would be only the logical form of a concept, not the concept itself through which something is thought (A95).

To begin with, an \textit{a priori} concept, a category, is not the logical form of a concept, it is rather a sort of concept which relates to experience, it relates to an object, through it something is thought. Now, an account of the possibility of such concepts which does not aim to connect them either to the concept of a possible experience or to the elements in a possible experience must be a wholly impossible and contradictory account according to Kant. Thus he explains further: ‘A concept which contains a synthesis is to be regarded as empty and as not related to any object, if this synthesis does not belong to experience either as being derived from it, in which case it is an \textit{empirical concept}, or as being an \textit{a priori} condition upon which experience in general in its formal aspect rests, in which case it is a \textit{pure concept}. In the later case it still belongs to experience, inasmuch as its object is to be met with only in experience’ (A220/B267).

As categories are \textit{a priori} concepts which relate to experience, to an object, concepts which are not empty of content (they relate to intuitions which give us the objects of experience and are the whole object of a possible experience), Kant has to account for the categories in one of the two possible ways mentioned by him. But \textit{a priori} concepts as the categories are not empirical concepts and can therefore not derive from experience. Categories do not consist of elements of a possible experience, they are rather the \textit{a priori} conditions upon which experience rests formally. And Kant remarks, making full justice to the content aspect of these concepts, that the categories do belong to experience, precisely because of this role as conditions. These two claims are two connected aspects of Kant’s view that meaning is autonomous, a view which is enshrined in the way the categories are described to work, and which we must recognise according to him. The categories, the \textit{a priori} concepts, which express what is a substance, a causal relation, an so on, do not consist of elements of a possible experience. This is the claim that the categories are autonomous in relation to such elements of experience, that is, these elements can not \textit{justify} the categories. Let’s see this aspect first.

The point should not strike anyone as surprising. To understand the purpose of a transcendental deduction of the categories just requires the recognition that an empirical derivation of them, an attempt to justify their objective validity by appeal to the elements of a possible experience, a \textit{de facto} exposition of origination, is not adequate (cf. A84/B116...
and ff.; B167). There are many facets for why this is so in Kant. Some are linked to Kant's Transcendental Idealism, which will be dealt with later on. In the context of what is presently under scrutiny, I will point out one of Kant's reasons.

Categories are concepts singled out for pure a priori employment and empirical proofs which appeal to how things are in experience are not sufficient to justify such type of employment (Cf.A85/B117). This employment of the categories is explained by Kant to involve, as with 'cause' for example, 'a strict demand that something, A, should be such that something else, B, follows from it necessarily and in accordance with an absolute universal rule' (A91/B124). Thus categories are marked out as involving necessity, which is a special 'dignity' that 'cannot be empirically expressed' and is the 'strict universality of the rule' which is not a characteristic of empirical rules or principles obtained, say, through induction (cf.loc.cit., also A92/B124, A112, A114, and A196/B241). What Kant is saying is that the categories, because of their feature of necessity, can not be justified by reference to reality, to the elements in a possible experience, to how objects are in the world, even with their characteristics that do not change from one situation to another, that is, are always constant.

A full justification for this negative claim has to be found in Kant's own way of facing the question of how a priori concepts relate to their object or to experience, viz. through being the conditions of the object or of the experience. However, what we have with the above strictures on Kant's part is the espousing of a philosophical tenet full of consequences for the whole of his philosophy: necessity is never given empirically. Put in other terms: 'experience yields us no knowledge of necessity' (A353)\(^{36}\). This amounts to a

\(^{36}\) Cf. A91-2/B123-4 and Kant's disagreement with Locke and Hume in the continuation; A112-3; A196/B241. Cf. also Allison (1983), p.78 for the disappointingly short and simple remark that this is Kant's 'fundamental assumption'. However, I take the most striking statement by Kant to be A106: 'All necessity, without exception, is grounded in a transcendental condition'. If we take into account the constitutive role of the conditions Kant is investigating, this Kantian approach then differs radically from others in Metaphysics. Cf. Davidson: 'In sharing a language... we share a picture of the world that must, in its large features, be true. It follows that in making manifest the large features of our language, we manifest the large features of reality. One way of pursuing metaphysics is therefore to study the general structure of our language' (1977, p.199). Cf. Kant's views on the distinction of levels between the physically real and the logical. For example, in the Attempt to Introduce the Concept of Negative Magnitudes into Philosophy, Kant strives to distinguish logical contradiction and physical opposition through the notions of logical addition and algebraic addition. In the Metaphysical Foundations of Natural Science (1786) Kants puts his view in a nutshell: 'But the principle of contradiction does not repel any matter that advances in order to penetrate into a space in which another matter is to be found' (p.42, AkIV498). Cf. also what Kant says about the transcendental truth of the categories, a notion Kant explains thus: 'I can represent to myself diverse things (substances), which are so constituted that the state of the one carries with it some
philosophical position which is so fundamental that, were it different, Kant’s philosophy as a whole would be of a completely different character. I will not dwell on this issue apart from saying that whatever philosophical defense of such a radical view is possible, it has to show itself in the far-reaching account of the nature of philosophy itself, in the philosophically suitable explanation of, for example, intentionality, and of the issues we are presently examining.

Before taking up the other possibility of accounting for the content of the categories and for their relation to the concept of possible experience, in fact Kant’s option, I would like to put out of the way at this juncture of the discussion a view of Kant’s position which could be levelled at him at the same time as a criticism: innatism.

Given that Kant is not deriving the categories from the elements of experience; and given that their possession is considered to be some sort of fact (cf.A85/B117); Kant would be committed to innatism with respect to the categories. This view claims that the problematic concepts that necessarily agree with things would, as a fact, be innate. I will not discuss the shortcomings of innatism as such. I will just assume that it is a weak option. Indeed, Kant’s interest in the philosophical question of the necessary agreement between the understanding and things can only be understood against a background according to which innatism is an inappropriate answer. At B167-8, Kant discards implanted subjective dispositions of thought. His main reason, interestingly enough, is that even if they were a fact (an empirical fact) this would still not explain the necessity which belongs to the very conception of the categories. Arguing that one is constituted just so as to think this and that as a matter of fact would for Kant play in the hands of the sceptic, who could then claim that objective validity is ultimately an illusion. But to adopt this position would - in a very interesting way - not be philosophically illuminating at all. ‘Certainly a man cannot dispute with anyone regarding that which depends merely on the mode in which he is himself organised’ (loc.cit.).

consequence in the state of the other, and this reciprocally; but I can never determine from these concepts which contain a merely arbitrary synthesis, whether a relation of this kind can belong to any [possible] things. Only through the fact that these concepts express *a priori* the relation of perceptions in every experience, do we know their objective reality, that is, their transcendental truth, and this, indeed, independently of experience, though not independently of all relation to the form of an experience in general, and to the synthetic unity in which alone objects can be empirically known’ (A221/B269).
Kant’s most unequivocal denial of innatism is, however, the following:

The *Critique* admits absolutely no divinely implanted or innate representations. It regards them all, whether they belong to intuition or to concepts of the understanding, as *acquired*. There is, however, an original acquisition (as the teachers of natural right formulate it), consequently also of that which previously did not exist, and therefore did not pertain to anything before the act 37 .

Kant accepts nonetheless talk of an innate basis, but only of the ‘ground’ of the *a priori* conditions he is investigating:

The ground of the possibility of sensible intuition is... the merely particular *receptivity* of the mind, whereby it receives representations in accordance with its subjective constitution... Only this first formal ground, *e.g.*, the possibility of a representation of space, is *innate*, not the spatial representation itself. ...Thus, the formal *intuition*, which is called space, emerges as an originally acquired representation (the form of outer objects in general), the ground of which (as mere receptivity) is nevertheless innate and the acquisition of which long precedes determinate *concepts* of things that are in accordance with this form. The acquisition of these concepts is an *acquisitio derivative*, as it already presupposes universal transcendental concepts of the understanding. These likewise are acquired and not innate, but their *acquisitio*, like that of space, is *originaria* and presupposes nothing innate except the subjective conditions of the spontaneity of thought (in accordance with the unity of apperception)^38 .

Now, one of the commonest ways of interpreting these Kantian claims is to say that time, space and the categories are not ready at the time of birth, but at that moment they have the opportunity of being developed, that is, when experience starts. However, as Jens Saugstad remarks, this would mean that they are still innate, although not fully developed, and their acquisition, which Kant calls ‘original’, would only be the development of what in fact existed in the mind, and not as Kant claims ‘the original acquisition... of that which previously did not exist’^39 . This author interprets Kant’s original acquisition as being an *externalism*: ‘externalism explains how representations can be both *a priori* and originally acquired: they are neither abstracted from the senses nor developed from intellectual “seeds”, but formed with the acquisition of behavioural skills’^40 . We will see in the next chapter how externalism accounts for the central of Kant’s claims.

37 . *On a Discovery*, p.135, AkVIII221.


How does Kant intend to give content to the categories then? As he has discarded the appeal to the elements of experience, to the objects in the world, Kant has to take the other possible option: categories, \textit{a priori} concepts, have to be contained in the concept of possible experience. I propose to approach this point from one particular direction in the following. In the next section, this point will be related to the issue of 'synthesis' in Kant.

Given Kant's account of analytic judgements in the 'Introduction' to the \textit{Critique}, it is almost impossible to resist the construal of Kant's option as stating that the concept of experience and, say, the \textit{a priori} concept of substance are connected analytically. According to such a view, the principles of Metaphysics would then be analytical, they would directly employ the term 'experience', as for example 'experience is of substances'.

But Kant does not think that this is the correct point of view on this matter (cf.A6-10/B10-4). We should not expect to achieve success in the present attempt by appealing to pre-formed concepts as the source of the content of the categories, so that the latter would thereby be guaranteed to have the correct meaning: 'Concepts of the understanding are also thought \textit{a priori} antecedently to experience and for the sake of experience, but... they are not preceded by any \textit{a priori} concepts of objects from which they could be inferred' (A310/B366-7).

We can also draw on some of Kant's metaphilosophical reflections to get Kant's position concerning the relation of the categories to the concept of a possible experience (these reflections will be further explored in chapter 5). In philosophy one should not hope to find any \textit{synthetic judgements} directly derived from concepts. One must, rather, look at \textit{possible experience}, which is the only 'third thing' able to connect the terms in the synthetic judgements of philosophy, a 'third thing' which is nonetheless contingent (cf.A736-7/B764-5; also A783/B811, A733/B761). Now, when we come to the judgements in philosophy, which are synthetic \textit{a priori} for Kant, we are dealing with \textit{principles} (not \textit{theorems}) which can not exhibit \textit{a priori} any of their concepts (as can mathematics through the \textit{construction} of concepts). So, principles of philosophy, of metaphysics, can only contain a synthesis of possible intuitions which are not given \textit{a priori} (cf.A719/B747). In philosophy's case, we use \textit{a priori} concepts 'in forming synthetic \textit{a priori} judgements, but only discursively in accordance with concepts, never intuitively through the construction of the concept' (A720/B748). In metaphysics, then, we can only \textit{exhibit} the categories, which appear in transcendental synthetic principles, \textit{a posteriori}, 'by means of experience, which itself is
possible only in conformity with these principles' (A721/B749). These principles have then the 'peculiar character' of making possible 'the very experience which is [their] own ground of proof' (A737/B765). ‘Save through their [the categories’s] original relation [necessary and a priori] to possible experience, in which all objects of knowledge are found, their relation to any one object would be quite incomprehensible’ (A94/B126-7).

We can, therefore, allow Kant the view that the categories acquire their content through their connection to the concept of a possible experience without this being expressible in an ‘analytic judgement’. ‘Every event has a cause’ is then synthetic a priori, a true principle of metaphysics, because (1) there is a non-empirical link between the concept of an event and the concept of being caused (however, this link is not that of being ‘contained in’) and (2) this link is forged through the concept of a possible experience: to experience something as an event is to regard it as having been brought about by a cause, even if the latter is unknown. This seems to be a clear case of a ‘principle’ which is established ‘discursively’, not ‘intuitively’, for the sake of the possibility of experience.

The point about why on Kant’s eyes the principles of metaphysics are synthetic a priori, rather than analytic, with its introduction of the concern with the possibility of experience, allows me to explain the second aspect of the view that meaning is autonomous.

When Kant proposes to explain how the categories have content, how they relate to an object, by claiming that it is because they express the conditions of experience and therefore the conditions of the object of experience (cf.A783/B811), his appeal to the ‘third thing’ - the concept of a possible experience - must be understood as follows. Let me continue to use the example of an event or of what happens.

It is not the case for Kant that an event is always caused because if we observe the world we would find out that events are always caused, as if it were true that every event has a cause because it is a fact that events are caused. This could not be the justification of a transcendental principle because the appeal to what is said to be a fact in this case does itself presuppose that the transcendental principle it purports to justify holds already (which would make the attempt circular). This is why Kant, on the contrary, says that such a principle ‘makes possible the very experience which is its own ground of proof, and that in this experience it must always itself be presupposed’ (A737/B765). Thus experience is a ground of proof, but what is offered to human knowledge in experience - as facts - does always presuppose what makes experience itself possible: the transcendental principles. The same
argumentative move is present in Kants' pointing out of a circle in a reasoning concerning truth:

Truth, one says, consists in the agreement of cognitions with the object. According to this mere verbal explanation, my cognition, then, in order to pass as true, shall agree with the object. Now I can, however, compare the object with my cognition only by cognizing it. My cognition thus shall confirm itself, which is yet far from sufficient for truth. For since the object is outside me and the cognition in me, I can judge only whether my cognition of the object agrees with my cognition of the object. Such a circle is explanation was called by the ancients diallelus.\(^{41}\)

The positive side of Kant's account of the categories - how they condition experience - will be further explored in the next section, because the issue of the autonomy of meaning in Kant can not be wholly separated from his views on 'synthesis' and the philosophical position called Transcendental Idealism.\(^{42}\)

---

\(^{41}\) Logic, p.55. Kant there agrees about this constituting a mistake, but asserts that the solution along the lines proposed is completely impossible: 'The charge was well founded indeed; but the solution of the task in question is completely impossible for anyone' (loc.cit.). It is important to notice about this argumentative move that it is presented in this occasion as if it were an epistemological problem only. I think it is fundamental to Kant's claims about the transcendental conditions he is investigating. Cf. also Prolegomena, #36 (p.AkS 18-20).

\(^{42}\) I would like to mention at this point a different view of what in Kant allows me, I believe, ascribe to him the account of meaning as autonomous. This view has to do with Strawson's imputation to Kant of a 'transcendental psychology' which is said to be an 'imaginary subject': it would describe how our mind's activity is the source of concepts in the process of creating nature (cf. Strawson (1966), pp.93-7). Thus, Derek Bolton (1982) maintains that the transcendental conditions of experience are independent from the experience - 'autonomous' as I am calling them - insofar as the ego is the transcendental origin of nature. That is, the ego is responsible for the non-empirical characteristics of our language and experience insofar as it is related to the world from the outside. I gather that Bolton's understanding of 'transcendental', in relation to our language and experience, as that which originates from outside the world we experience, comes from Bernard Williams' misconception of 'transcendental' in his discussion of Wittgenstein's supposed 'transcendental idealism', cf. (1974). More to the point about Wittgenstein and about what 'transcendental' can mean from a truly critical philosophical perspective is, I believe, Anscombe's suggestion that in Wittgenstein we find a 'linguistic idealism': 'linguistic, because he [Wittgenstein] describes concepts in terms of linguistic practices. And he also wrote [in a way which Anscombe intends to be explanatory of 'idealism']: "Essence is expressed by grammar" (PI, I #371)’ (1976, p.112). For me two features of Anscombe's suggestion are very important. Firstly, Wittgenstein's position is actually stronger: ""linguistic idealism" would go further and say "essence is created by grammar". For the "essential" is the "mark of a concept, not the property of an object" (ibidem; the latter quotation is from Wittgenstein’s Remarks on the Foundations of Mathematics 1, 73). Secondly, Anscombe qualifies the idealism. According to her, the idealism Wittgenstein embraced is committed to a 'realism without empiricism', i.e. to the view that asserts: 'that one knows something is not guaranteed by the language-game' (1976, p.133). Wittgenstein's position would be committed to the view that it is 'always by favour of Nature that one knows something' (On Certainty, #505). Anscombe's Wittgenstein adopts a philosophical position which is akin to Kant's in many ways.
Now I have to admit that the text which I used as the basis for my present discussion (A95) does seem to be concerned with the particular problem of the possibility of synthetic \textit{a priori} representations, and not with the problem of the possibility of representation in general, the problem of intentionality as I have presented it. Thus far my way of presenting Kant’s problem of the Deduction as that of the intentionality of thought certainly involves a terminological step. However, I believe it is important to put Kant’s concern in the terms I have used - the possibility of representation in general, the intentionality of representations - because for Kant synthetic \textit{a priori} judgements, which are synthetic (apply to experience) and \textit{a priori} (are independent from experience) insofar as they express the preconditions of the possibility of experience (the set of synthetic \textit{a posteriori} representations), are at the same time and precisely in this capacity the preconditions of the \textit{object} of experience. As Kant explains:

\begin{quote}
Synthetic \textit{a priori} judgements are thus possible when we relate the formal conditions of \textit{a priori} intuition, the synthesis of imagination and the necessary unity of this synthesis in a transcendental apperception, to a possible empirical knowledge in general. We then assert that the conditions of the \textit{possibility of experience} in general are likewise conditions of the \textit{possibility of the objects of experience}, and that for this reason they have objective validity in a synthetic \textit{a priori} judgement (A158/B197).
\end{quote}

So, what I would like to point out is that Kant’s concern with the possibility of experience, with his use of terms like concepts, intuitions and judgements, is being characterized by me with a different terminology: as the post-\textit{Tractatus} concern with the possibility of representation in the sense in which thoughts are deemed to depict reality (the phenomenon of intentionality).

\section*{5. The Beginning of Kant’s Solution to the Problem of Intentionality: Synthesis}

To get the right picture of Kant’s solution to the problem of the intentionality of thought we should start with Kant’s account of \textit{synthesis}. The present section aims to show that what Kant understands by ‘transcendental synthesis’ is what makes possible an intentional thought or representation, in Kant’s terms: a judgement. Thus far, synthesis is shown to make possible concepts with \textit{content}, as the categories have to be (not mere forms without relation to objects). Now, how Kant articulates his views in a wider philosophical position - his Transcendental Idealism - will be discussed in section 6.
Synthesis is a topic in Kant normally viewed with suspicion. The commonest approach is to link it with infamous views about the productive activity of mind creating nature, which are deemed to give the core of the conception of transcendental idealism. This view of synthesis, as being Kant's way of indulging in 'the imaginary subject of transcendental psychology', consequently fits well with assessments of Kant's aims in the Deduction as concerning only an analytical argument for the objectivity of experience, assessments which are then able to eschew completely transcendental idealism. I do not think this does justice to Kant's intentions, first of all concerning the topic of synthesis.

There certainly are texts in Kant which seem to point at a particular direction. In A99, Kant talks of the synthesis of apprehension in intuition in a way that seems to set the standard for what one should expect even a transcendental synthesis could achieve. There, each representation is said to be "contained in a single moment, [and] can never be anything but absolute unity" (A99). What occurs is the reviewing of and retaining together of the representations, so that a unity can be obtained. This certainly appears to involve a kind of 'production' of a representational item from a series of discrete sensory inputs. But even in such contexts, what is actually important is a rather different kind of concern, i.e. a concern with the representation of a manifold as a manifold; that is, the fundamental concern is about a representational ability, not about a psychological process of production of a mental entity of some sort. What I take to be crucial in Kant's claims about synthesis are a couple of logical points about the conditions of representing, and this has nothing to do with psychical entities, functional states, etc.; rather, they are made in view of what makes a judgement possible, what is able to give it content. Let me explain.

43. This is obviously Strawson's approach, already quoted. Cassam follows Strawson's footsteps in (1987), with the result that he believes that 'Kant's appeal to the notion of "transcendental synthesis" is quite unsuccessful, even in its own terms' (p.371). The basic line of reasoning in Cassam (and Strawson) concerns the view of content or matter he embraces. He deems Kant to hold that the content or matter of experiences has, in a very literal sense, to be given in sense experience, which immediately makes any Kantian claims about a productive synthesis concerning the content or matter of experience unintelligible. For an unashamedly psychological account of synthesis and its role in Kant's arguments, directed against Strawson, see Kitcher (1984), pp.114-21; (1990) chapter 3, pp.61-90.

44. Cf. Allison (1983), pp.140-4 for the view that the topic of synthesis is connected with the having of a discursive intellect; furthermore, that 'the problem is to explain how such a mind can represent to itself its data as combined' (p.141).
In Kant’s statements about synthesis what should strike one at first is not any psychological side of the story, but rather the logical problem of the ‘grasping of what is manifold [in representations] in one [act of] knowledge’ (A77/B103). One should note that Kant’s central concern with the doctrine of synthesis is with the content of concepts, that is, with what is represented in them, i.e. objects.

Kant starts by making the obvious point that, for an analysis of concepts to reveal their content, this content has to be forged by synthesis beforehand (cf. ibidem). Synthesis, then, in Kant is better seen as the forming of a concept from other concepts that are in the former as characteristic marks. As an illustration, ‘Book’ is such a concept, comprising as its characteristic marks the concepts appearing in ‘a written or printed work consisting of pages glued or sewn together along one side and bound in covers’ (The Concise Oxford Dictionary). Now Kant’s point about analysis coming second in what concerns content can be explained through this example: the dictionary’s entry represents the proper analysis of the term, on the condition of our having this concept in the first place.

Kant does not present his points in this pedestrian way (and the process I described above he seems to attribute to general Logic, cf. A78/B104). He makes us believe that the issue concerns pure synthesis only, that it concerns Transcendental Logic exclusively. So our concepts of the understanding are required to give unity to representations in judgements and to the synthesis of representation in an intuition (cf. A79/B104-5). However, the issue of representing is the central point here in all its generality. The appeal to categories and pure synthesis reveals more about Kant’s solving of the problem than about the concern in all its diverse implications. In fact, the topic of a pure synthesis will also appear later on when we come to deal with what is the ultimate basis for the whole phenomenon of concepts having contents or judgements being objectively possible, i.e. what it is that makes rules possible, and this has to do with ‘acts’ and ‘activities’ as I will claim, but they will not be conceived psychologically. This, however, has to wait for the next chapter’s account of Kant’s externalism. Now the issue is circumscribed to how rules enter Kant’s view of the intentionality of judgements. We will presently see how in Kant concepts are rules which express a synthesis.

Kant certainly does not believe that we can solve the problem of how we represent merely by pointing out that concepts are ‘analytical unities’, that is, by stating that all concepts are of the nature of mediate representations; what has to be recognised with this
fact is that the understanding through synthesis - by the forming of concepts through the encompassing of partial concepts - injects 'transcendental content' into the concepts. This has, in Kant's wording of it, the status of an argument in Transcendental Logic: we can posit pure concepts of the understanding which apply a priori to objects, a point which General Logic would never be able to establish (cf.A79/B105).

Again, Kant's wording should not preclude us from recognising that his argument concerns representing in general. Pure concepts of synthesis are possessed by our understanding, and by them alone can we understand anything in the manifold of intuition, that is, can we think an object of intuition (cf.A80/B106, also #22 Deduction B). So, the 'transcendental content' of certain concepts is due to nothing else than the relation of the concept to an object in such a manner that the concept represents the object. Now, because this relation has conditions, Kant approaches the issue through Transcendental Logic, so that a priori considerations become important. Most notably, the relation becomes relation to an object, which confers upon our representations (concepts) a 'dignity' that results - and this is the absolutely crucial element of Kant's position - from their being subjected to a rule, which is here just assumed to have the power of necessitating us to connect representations (concepts) in a determined manner. This 'subjection to a rule' is what the a priori part of Kant's story really means (cf. quote infra). The important conclusion about synthesis here is that even the transcendental concepts contained in the pure understanding are given their content through the synthetic activities of our de facto discursive understanding, for categories 'are merely rules for an understanding whose whole power consists in thought' (B145) (so much should also be clear from Kant's refusal of innatism). More can be said about the a priori though. And this should also help with Kant's claims about appearances displaying order and regularity in accordance with our understanding necessarily (cf.A125).

The pure synthesis Kant insists on is said to be that which 'rests upon a basis of a priori synthetic unity' (A78/B104). What we are dealing with here is then, according to Kant, a synthetic unity of a manifold which is 'rendered necessary' (cf. ibidem). The aprioricity and necessity of the synthetic unity are the necessity linked with the conception of the content of a certain concept. In my example of the book, the a priori necessity of the synthetic unity forged in that concept is the necessity that is present in the explanation that 'book' means 'a written or printed work consisting of...'. Kant, however, implies that this type of account should be restricted to his categorial concepts. The a priori necessity linked
with the synthetic unity in the concept ‘substance’ is then the necessity connected with the claim that “substance” is that which is permanent. My point is that the account is in principle extendable to other concepts as well. (One must obviously admit that ‘book’ is an empirical concept with content which plays a modest role in our lives, nothing compared to the role of ‘substance’ in our conceptual scheme).

At this juncture it is important to see what follows from this view of synthesis to the account of the content of the categories, especially in view of the belief that, in an empiricist fashion, meanings are determined by nothing else than by the elements of experience. Let us take the following quote, and start by noting the role of rules in Kant’s account.

How, then, does it come about that we posit an object for these representations, and... ascribe to them some mysterious kind of objective reality. Objective meaning cannot consist in the relation to another representation (of that which we desire to entitle object), for in that case the question arises, how this latter representation goes out beyond itself, acquiring objective meaning in addition to the subjective meaning which belongs to it as determination of the mental state. If we enquire what new character relation to an object confers upon our representations, what dignity they thereby acquire, we find that it results only in subjecting the representations to a rule, and so in necessitating us to connect them in some one specific manner (A197/B242).

In this quote Kant makes the point which is crucial to his account: how content can result from subjection to a rule. We should be clear about the account of content implied by Kant’s claims. There is in Kant a view of the content of concepts (I will discuss later the content of thoughts and beliefs) radically different from the one according to which the elements of experience by themselves determine what is represented. Kant sometimes calls content, as we have seen, the ‘transcendental content’ of the pure concepts of the understanding (cf. A79/B105)\(^45\). On other occasions, this view of content I want to bring out equates content with logical matter: ‘In any judgement we can call the given concepts logical matter (i.e. matter for the judgement), and their relation (by means of the copula) the form of the judgement’ (A266/B322). We should realize, however, that the ‘matter for the judgement’ (as a concept) stands, notwithstanding, for something in the object. In the continuation, Kant writes: ‘In every being the constituent elements of it (essentialia) are the

---

\(^{45}\) However, at A575/B603 Kant generalizes the view and talks of what are empirical concepts as having material or transcendental content. Cf. Allison’s Introduction to On a Discovery, where he claims that Kant is interested in the predicate’s ‘real’ relations to objects, and that these are the content of thought. Real predicates are affirmed to give the determination of things in the sense of their real essence (cf.p.54 and ff.).
matter...’ (ibidem). This point about content or matter being what in judgements are concepts, which nonetheless stand for elements in the objects or for the objects themselves, is I believe fundamental. In the Jäsche Logic the matter (or content) of a concept is the object.\(^\text{46}\)

Now, it is because concepts are representations which are mediate (cf. Kant’s remarks on the impossibility of a conceptus infimus or a species infima, A655-6/B683-4) that we have the need of a synthesis. In fact, there would be no concepts with content without synthesis. To have a concept with content is, then, for Kant to have a synthetic unity represented in a concept. Thus, content as what is represented can be attributed to concepts only insofar as there is in them a synthetic unity of concepts which are constituents of the original concept (partial concepts of it). And this, Kant asserts, is only achieved through the subjection of representations to rules, which is the way of necessitating us to connect the constituent concepts of a concept in one specific manner. This is the way we have concepts at all, which is at the same time how Kant envisages concepts having a relation to an object (as quoted supra).

Transcendental synthesis is, then, the general designation Kant gives to the process of uniting representations necessarily so as to give them content, objective meaning. The process is successful, however, precisely insofar as there is subjection of representations to rules, which is what allows one prior and independently of experience to say that such and such features are combined in the object. Kant, therefore, speaks of representations necessarily belonging to one another (cf. #19 of the B Deduction).

Kant also writes: ‘The combination (conjunctio) of a manifold in general’ is ‘an act of spontaneity of the faculty of representation’, which is the understanding. All combination is then ‘an act of the understanding’ which receives the name ‘synthesis’. As an act of the understanding, synthesis intends to indicate ‘that we cannot represent to ourselves anything as combined in the object which we have not ourselves previously combined’. Synthesis is then ‘an act of the self-activity of the subject, it cannot be executed save by the subject itself’ (a topic which will become important later on). But what we thus represent as combined in the object is the result of a unity which is in fact a ‘unity of diverse concepts in judgement’ (cf.B129-31 for these points). Synthesis is, therefore, that, which in making

\(^{46}\) Cf. op.cit. p.96.
representations necessarily belonging to one another, at the same time makes judgements possible. 'The possibility of experience is, then, what gives objective reality to all our a priori modes of knowledge. Experience, however, rests on the synthetic unity of appearances, that is, on a synthesis according to concepts of an object of appearances in general' (A156/B195).

Reverting the discussion to the problem of intentionality, Kant's crucial notion of 'subjection to rules' is what makes possible an intentional thought. Firstly, that I mean such and such an object is not dependent on anything mental (processes or states). The general reason for this is that no process or state could have 'the consequences of meaning' (Kant remarks that objective meaning cannot consist in the relation to another representation understood as a determination of a mental state, because this would not 'go out beyond itself', it would only have 'subjective meaning', it would not constitute a 'relation to an object', cf.A197/B242, quoted supra). That I mean such and such an object depends on my utterance of the thought counting as one about such and such an object. This is a normative connection, one which can only depend on the subjection to a rule, not on the description of a mental process or state (this is Kant's 'new character', the 'dignity' acquired by representations when they are subjected to a rule: they can then relate to an object because this subjection is what makes us connect representations in specific manners, cf. ibidem^47.

Secondly, as a further example, it depends upon a rule my having, say, the expectation that p with exactly this content. Because this is the correct specification of my expectation only insofar as the expectation in question is the expectation which will be fulfilled if p. So, these are two ways of referring to the same expectation, ways which are connected by a rule of meaning: they mean the same (obviously this concerns only the identification of an expectation, what it is an expectation of, not the fulfilment of it: we can say what will necessarily fulfil it, not whether it will be fulfilled at all)^48. Thus, it is through the subjection to a rule that I can have thoughts with intentionality, that is, thoughts which are about certain objects. It is precisely because, say, a belief that p is such a belief only by being the belief that is made true by the fact that p. This is a connection between two

47. Cf. Glock (1996), pp.180-1 for the points about intentionality made in the text, including Wittgenstein's remark about what can not have the 'consequences of meaning'.

48. Cf. op.cit., pp.186-9 for the points about expectations and their contents and the connection of these with the future.
descriptions, established by a rule, which allows one to believe something, that is, one manages to have a belief about something independent of oneself.

6. Kant’s Idealism and the Transcendental View of Content

Kant’s views on the topic of synthesis, with the result that representations have to be subjected to rules, combined necessarily, if we want to represent to ourselves anything as combined in the object, are set by Kant himself in a wider picture which is his Transcendental Idealism. I propose to examine in this section this wider picture and its central tenets. I aim to show that Kant’s solution to the problem of intentionality through synthesis appeals to a transcendental view of content according to which representations are not mental intermediaries between the subject and the object. I am going to claim that Kant was not a representationalist: as far as ‘appearances’ are concerned there are no intermediaries to block the access of judgements to their objects. Concerning the overall philosophical position - Kant’s Transcendental Idealism - I will be offering a reconstruction of it which stresses its idealist as well as its realist elements (my account of Kant’s idealism differs significantly from Strawson’s: I allow Kant an idealism while retaining a realism. I have already indicated the direction of my account in section 4 above).

The basic motivation for an examination of the larger philosophical position is that the point about an a priori necessity being present in synthetic unities (concepts) is linked with the view that ‘appearances’ display order and regularity in accordance with our understanding necessarily. So, according to Kant, the understanding with its representations is linked with ‘appearances’ in a necessary fashion. Kant tackles the problem of explaining what this means with the help of the revolutionary notion of object presented by Transcendental Idealism.

Kant’s idealism is called, by the author himself, formal (cf.B519) in an explicit attempt to distinguish his objectives from the objectives of those who, by doubting or even denying ‘the existence of outer things themselves’, embrace a material idealism. Kant’s goal is different. He is interested in how knowledge in general, including for this matter knowledge of outer things, is possible. Accordingly, he calls transcendental the knowledge

49 Cf. op.cit., p.187 for the point that subjection to a rule of language is what can ‘dissolve the puzzle about how a thought, something I have here and now, can be about something which is far away and in the future or past’.
of this possibility insofar as it is necessarily attainable a priori (cf.A11-2/B25, this explains something of Kant’s conception of the task and scope of philosophy). Furthermore, it is important to understand that Kant’s idealism is critical, because the position it defends in regard to this possibility - that we can only know how things appear and not how they are in themselves (cf.A30/B45; A42/B60; A128-9) - is based on a philosophical reflection on the conditions of possibility, and consequently on the limits of human knowledge. This reflection is not envisaged as a psychological investigation, nor as a reflection on the ultimate constituents of a reality that has an essence in itself. This is why the Kantian inquiry does not result in a doctrine, but rather in a transcendental critique, whose aim is not to extend the human knowledge; but instead, to offer a touchstone for the value of a priori claims (cf.A12/B26).

So, the core claim Kant makes with Transcendental Idealism is that we can only know ‘appearances’, not ‘things in themselves’. Now this very claim often gets understood as meaning a straightforward psychologistic thesis, i.e. that we can only really know what is in our minds, the result of how things strike us. Kant in many places in the Critique seems to endorse such a view when he equates ‘appearances’ with ‘mere representations’ (cf. especially A104; A119-2; B207; A490-1/B518-9). But this is not, in my view, what Kant intends.

‘Appearances’ in Kant should be equated with the transcendentally idealistic conception of object. This ‘object’ for Kant is the object of possible experience, and the main claim is that it is transcendentally ideal, that is, it is defined by necessary conditions of experience. This amounts to claiming that the concept of the object is dependent on the

50. Cf. Allison (1983), pp.26 and ff. for how Kant’s imprecise wording is responsible for what he calls the ‘standard picture’: the view of Kant’s Transcendental Idealism as a metaphysical theory that affirms the unknowability of real things (he would have called them ‘things in themselves’) and consequently limits knowledge to the purely subjective realm of representations (‘appearances’) (cf. op.cit. pp.3-4). Allison shows, however, in what is again an example of how Kant’s text deserves a very careful approach, that in the very passage where ‘appearances’ are said to be ‘mere representations’ what really matters is the manner in which objects are represented. ‘The claim, therefore, is not that objects have no independent existence...; it is rather that such existence cannot be attributed to them “in the manner in which they are represented”’ (op.cit., p.27). I take this kind of statement to imply an idealism with regard to our manner of representation, an idealism which results from a transcendental perspective on these matters. This will be discussed in the continuation. I would like to add only that my view of Transcendental Idealism is very much influenced by the Allisonian formal, anti-psychologistic (unfortunately not wholly anti-empiricist), two aspects interpretation of Kant’s idealism.
notion of a judging subject, and on the concepts used by this subject to judge how things are in the world. Kant is not claiming that the existence of the object is dependent on the existence of the judging subject, as we have already seen, nor is he claiming that the object is not real insofar as its concept depends on us (cf. infra). The Kantian point is that ‘object’ is a notion which, by being applied to objects in the world, is required to conform to the conditions under which we can represent objects as objects, that is, ‘object’ is required to conform to the conditions of our thinking of objects, to our concepts of object. This view, according to Allison, constitutes Kant’s ‘radically new conception of an object’. ‘An object is now to be understood as whatever conforms to the mind’s conditions... for the representation of it as an object’. Allison insists on the essential reference to mind in Kant’s new conception of object, with the implication that ‘object’ should be understood as something which is by its very nature represented. The definition of ‘object’ has built into it a reference to the mind and its powers of representation. In a comparison, one could say that ‘object’ in Kant’s new conception has become - from the point of view of Wittgenstein’s concerns in the Tractatus - an element, an instrument, that belongs to the ‘method of representation’; it has become that which is used in order to represent how things are in the world. Kant’s notion of ‘object’ is a transcendental idealistic notion insofar as it is the core of the view that there is an internal relation that connects ‘object’ and its meaning.

51. Cf. (1983), p.30 for all these points. Cf. Allison’s discussion on the notion of ‘object’ in the Deduction (1983, pp.133-6). Cf. Förster’s objections (1985), including the appeal to the Felix Meiner edition of the Critique to settle a textual matter. Concerning the point that in the notion of ‘object’ a reference to mind is included, I would like to add this comment: ‘...What we clarify in [conceptual] investigations is not the empirical landscape before us but our point of view, not what we apprehend but our vision’ (Hacker, 1986, p.206). ‘Our vision’ is clearly not conceived as something that concerns our empirical or psychological minds here.

52. Cf. Hacker’s account of the dismantling of the notion of ‘object’ in the Tractatus (1986, pp.113-20). It is said to change from the ultimate constituents of reality to an element, or an instrument, of the method of representation, of the symbolism. The related notion of samples also changes, so that they now become part of the method of representation (cf. also Baker & Hacker, 1980, pp.81-118). Cf. Buchdahl (1969, p.630) for the view that Kant ‘peels the object off its foundation in the ontologically conceived “sum-total of all possibility” (as in the old doctrine of The Only Possible Ground for a Demonstration of the Existence of God [1763]) and places it within experience’. Cf. Buchdahl, op.cit., pp.484-90 as well.

53. Cf. Allison’s remark: ‘The position [Kant’s] is idealistic because... it grants to these conditions [the formal or a priori conditions of human experience] the function of defining the meaning of “object” or, equivalently, of determining what can count as “objective” for the human mind’ (1983,
There are in Kant many issues related to this idealistic account of object. Concerning the related notion of ‘thing in itself’, does it designate a special kind of entity, if one at all? The ‘transcendental object’ of the Deduction A: is it really a non-empirical and also non-intuitable object that can nonetheless be named \(X\) (cf.A108-11)? There is also the problem of transcendental affection: is the claim one about a non-empirical causation of our representations by the above mentioned \(X\)?

From what I have been saying I hope it is clear why I side with Allison’s account of these matters, of which the main features are: (1) ‘Things in themselves’ makes a reference to a consideration of things in transcendental reflection, it does not name a special being or entity, nor does it imply a special feature of things, a kind of meta-physical quality of ordinary objects. These terms, and correlates (‘appearances’), ‘function adverbially to characterize how we consider things in transcendental reflection, not substantivally to characterize what it is that is being considered or reflected upon’. So ‘in itself’ only implies that the thing considered in transcendental reflection is thought of independently of the existing conditions for representing objects. (2) ‘Transcendental object’ refers, again in the transcendental reflection about these matters, to ‘things in themselves’ taken collectively as the ‘cause’ (ground in fact) or ‘material condition’ of our representations. (3) Accordingly, ‘affection’ is neither a causal notion in the empirical sense, nor the notion of a super-physical or literally meta-physical causation. (So the double-affection theory - according to which we are first struck by the transcendental object, which results in the creation of appearances which then strike the empirical self producing representations - is mistaken). ‘Affection’ concerns only the transcendental reflection that human experience, in what concerns its material condition (the spatio-temporal items of the world), is passive. We do not create through the power of our representational abilities the empirical objects of our experience, that is, the world\(^{54}\).

\(^{54}\) Cf. Allison (1983), especially chapter 11. The points I made can be found, respectively, in the following places: (1) pp.241 (the quote) and p.254; (2) p.250; (3) pp.247-54.
This last point fills in a blank I left above. I said that there was a special place in Kant’s picture for his insistence that objects should be, not only thought of, but also intuited. Empirical objects, then, when considered in transcendental reflection, bring into this perspective the matter of experience, what is given in experience. So, the point Kant is making is that his idealism is a form of realism, it is an empirical realism. Concerning this point it is worth reading Buchdahl’s remarks: “‘Experiencing’ is not inventing, is not a fictional construction; facts, objects, simply “exist”, are given - as an ultimate matter of fact; an ontological ultimacy which Kant expresses by maintaining that their givenness is not due to some causal action on the part of a divine source."

It may appear to some that, even without claiming that we create the objects (in regard to their existence), Kant is nonetheless committed to an intolerable kind of idealism. He would be claiming that our minds must attribute properties to objects which are really of our own making, or that we contribute to the process of knowing the properties of the objects. As with Locke’s secondary qualities, our minds would be the real cause of an object being spatial, temporal, coloured, solid, etc. The object itself would not possess these qualities, it would at most have a role in the production of the final experience of the object. Kant here is close to having his idealism wholly assimilated with Berkeley’s. But this is most definitely not Kant’s position.

What we would ordinarily call objects (material objects) are for Kant most definitely real objects, and they are not by any means the result of our mind’s creative faculties. Kant is emphatic about this:

55. Cf. (1992), chapter 6, p.162. This chapter contains also Buchdahl’s criticism of the ‘double affection’ view, as well as the basics of his ‘two aspect’ approach, which I take to be similar to Allison’s. Buchdahl obviously uses a different model of presentation: it involves ‘reduction’ and ‘realization’ aspects, cf. pp.153-65. Cf. also Buchdahl (1969) p.638, where he states that ‘the givenness of objects is a fact of experience’.

56. This is again, in the main, Strawson’s view (cf.1996, pp.38-42, his reference to Berkeley is at p.22). This view could be said to imply that what Kant says of Descartes’ idealism (that it is an empirical idealism, cf.A368-9) is actually true of his own type of idealism.


58. ‘At the empirical level, the “objects” of Kant’s world are the straightforward things of commonsense, their grammar is not indebted to any scientific-theoretical components’ (Buchdahl, 1969, pp.638-9n4).
All outer perception, therefore, yields immediate proof of something real in space, or rather is the real itself. In this sense empirical realism is beyond question; that is, there corresponds to our outer intuitions something real in space. Space itself, with all its appearances, as representations, is, indeed, only in me, but nevertheless the real, that is, the material of all objects of outer intuition, is actually given in this space, independently of all imaginative invention (A375).

In spite of possible misleading overtones in this quote, Kant holds the position which he calls 'empirical realism', that is, he maintains that our outer intuitions have corresponding to them something real in space, which is his manner of saying that there are objects in the world as we normally conceive of them. 'Representations' which are 'only in me' are those representations of objects we use, and of these representations it is said that they stand to something real in space, in the world. Again, this is the point that chairs, planets and sub-atomic particles exist in the world and that we represent them through these concepts. The point is that if it is a fact that the chair is half a metre to my left then exactly that is represented by 'the chair is half a metre to my left'. When we talk of objects it is of objects we are talking. Our representations reach out to reality, and reality is not something invented by us. One sits on chairs, bumps into them as well (don't you?), sees the planets (almost all with the telescope) and has proofs of the existence of sub-atomic particles (so the scientists tell us) and represent all these objects. This is the realist element of Kant's position.

Kant, however, qualifies his claims about how we represent and in so doing reveals the idealist element in them. Kant's qualification is an attempt at challenging a view of our contribution in the process of representing, it does not doubt reality's role. Kant's main goal is to get the right picture of our involvement in the representing of objects, and his major task is twofold. Firstly, he challenges empirical idealism, which is the view that what we actually know directly are only what is 'in our minds': external objects have an uncertain status (cf. A367 and ff.). We have seen how, against such a view, Kant insists upon the reality of the objects in the world: they are known directly by us. The purported intermediate mental term, the tertium quid, of the representing relation is written off by Kant. Now, secondly, the other side of the empirical idealist position is encapsulated by the tenet that what we have, what is 'within us', is what we know directly, and Kant's Transcendental

Idealism is also aimed against the latter conception. So what Kant wants is to break the link that leads the transcendental realist to become the empirical idealist.

It is in fact, this transcendental realist who afterwards plays the part of empirical idealist. After wrongly supposing that objects of the senses, if they are to be external, must have an existence by themselves [be known in themselves], and independently of the senses, he finds that, judged from this point of view, all our sensuous representations are inadequate to establish their reality (A369).

Kant, from his transcendental perspective (which is contrasted with what could be called a mentalistic perspective: empirical idealism plus transcendental realism), identifies his idealism with an empirical realism: 'the transcendental idealist is, therefore, an empirical realist, and allows to matter, as appearance, a reality which does not permit of being inferred, but is immediately perceived' (A371). In this statement 'as appearance' should be understood as referring to objects in the world, which are necessarily represented as objects by the only means we have of doing so: our concepts\(^60\). This argument introduces Transcendental Idealism.

**By transcendental idealism** I mean the doctrine that appearances [objects in the world] are to be regarded as being, one and all, representations only [things represented as objects], not things in themselves [things with an essential nature of their own independent of our representations], and that time and space are therefore only sensible forms of our intuition, not determinants given as existing by themselves, nor conditions of objects viewed as things in themselves (A369; my additions: AK).

-------------------

\(^60\) Kant's idealism should not be understood as the simple view that the concepts we have are our concepts, as if 'red' could be a concept fixed upon by us, as if chairs were things we chose to group under 'chair', as if perhaps, at least in some cases, we could have done otherwise. This view is a psychologism with a linguistic conventionalism: our minds would be responsible for the whole course of the phenomenological world. As Walker puts it: 'The groupings of things together which we naturally make, however much they may seem to be forced upon us, are all the product of our own activity' (1985, p.237). I believe that if we add to this view the proviso that, in truth, things in themselves have properties which in a way interact with our ordering of the world, we have Walker's own approach (cf. op.cit. passim). What is important to notice about this approach is that it functions precisely under the assumptions Kant wants to challenge, it does not realize that Kant's attack is on both empirical idealism and transcendental realism, that is, on their common mentalistic presuppositions. Kant is not interested in - as Walker thinks - 'relating the mind's contribution to the contribution made by an independent reality' (op.cit. p. 241). Concerning Walker, the Kantian dictum seems to be a curse: he who thinks that he knows the properties of things in themselves is bound to be an empirical idealist. Thus Walker duly writes: 'However similar two items, e.g. two red pillar boxes, may strike us a being, that is no guarantee that they are similar intrinsically' (op.cit. p.238). What 'intrinsically' is intended to capture here is a mystery. But, I cannot help thinking that we do in fact have the guarantee that can be asked for two pillar boxes being both, for example, red: would anyone in England think he does not really know that any two red pillar boxes are similar with respect to colour?
As Allison puts it, the independent existence of objects cannot be attributed, ascribed, to them 'in the manner in which they are represented'. This should not be understood as implying that I could not explain to someone that, say, a physical object is something independent of my person. Allison's remark intends to be a point pertaining to the transcendental reflection on how we represent: concerning the essential (necessary) nature of a physical object, this nature - when it involves the determination, for example, as something independent of my person, which I can give as explanation to someone - cannot be captured independently of how we represent objects. Let me illustrate the point with how one particular category works according to Transcendental Idealism.

Suppose I use Kant's notion of substance (cf. A143/B183) and explain to someone that "substance" is that which is permanent. What I am doing is explaining the essential nature of substance, in a way which is not independent of our way of representing objects. Substances for us just have as their essence being things which are permanent. Because of this, if I wanted to explain 'substance' as 'a property of things' I would not be giving the essential nature of substances: but not because I would be failing to represent reality as it really is - as it is 'in itself' - rather I would be failing to explain what we understand by 'substance' (cf. Kant's remark that 'we shall never dream of seeking to inform ourselves about the objects of our senses as they are in themselves, that is, out of all relation to the senses', A380). So, by parity of reasoning, if I say that red is a substance I am not committing a super-physical or meta-physical mistake, I am only failing to explain what we mean by the word 'substance'. Kant would say that from the transcendental perspective it does not make sense, but this is not to say that what is being ruled out as senseless is a possible state of affairs in a noumenal world of things in themselves inaccessible to our

61. 'The character of the "given" as "objective", i.e. as other-than-self will likewise have to be constructed as a contribution made a priori. The character of the object as being "other-than-self" will have to be located, not in what is received, but in what is spontaneously contributed by the cognitive subject in the context of judging. We might describe this by saying that the "self" is here regarded as the seat of the a priori, and of what is "other-than-self": objectivity being located within the framework of possible experience' (Buchdahl, 1969, p.629). "The real in the appearance must always be given, it can only thus be given "in the context of a possible experience"" (Buchdahl, 1992, p.164).

human understanding. I believe this would be to fall victim to an illusion, actually to a transistor illusion: one which consists ‘in treating the subjective condition of thinking as being knowledge of the object’ (A396). Kant’s point is that what we take to be the essential, necessary, nature of things is of our own making, it does not respond to how things are in themselves. This is the point that meaning, as our creation, is autonomous. Essence does not answer to how things are in themselves; and, at the same time, it does not have any application beyond the scope of what the principles of possible experience state, because meaning is also immanent (cf.A295-6/B352-3, where ‘transcendental’ is contrasted with ‘transcendent’ and equated with ‘immanent’; cf. also Opus Postumum, pp.244-6, for the view that transcendental claims are autonomous).

These points combine in the view that essence claims can only be justified transcendentally, that is, by an appeal to what we make to be the essence - to how we constitute the object in view of the possibility of experience. They are in fact constitutive of what are our ways of representing the world, so they are autonomous.

The point that essence claims can only be justified transcendentally and what I tried to capture by ‘immanence’ are put by Buchdahl into the following remarks about ‘transcendental subjectivity’:

[The] “sharpening” of Kantian meanings is graphically illustrated in putting Kant’s main point by saying that the “objectivity” of the phenomenal object is not “objectively given” but wrought “subjectively”. This at once makes it clear that the second “objectivity” in this account no longer has a normal sense. Nor is this surprising. For, figuratively speaking, one cannot characterize the limits of language

63. Cf. Kant’s comment about his idealism in the Prolegomena (p.Ak374): it does not claim to know things in themselves, nor that space and time are their qualities. It is useless to claim that substances are permanent because this is a feature of them as they are in themselves.

64. In the Prolegomena, p.Ak373n, Kant asserts that an immanent use is a use restricted to experience. It is important to note what the spirit of Kant’s work is: ‘High towers and metaphysically great men resembling them, round both of which there is commonly much wind, are not for me. My place is the fruitful bathos of experience’ (ibidem).

65. Cf. Buchdahl (1969, p.618) for the ‘constitutive’ force of the categories in regard to an object qua appearance. Cf. op.cit., pp.628-41, for the position that Kant’s view on the accidental and the necessary make the critical philosophy have ‘autonomy’.
and of our world by contemplating them from without, through the use of linguistic
tools taken from this very world, without radically altering their sense\textsuperscript{66}.

We should bear in mind, however, that this is admittedly figurative speech. Our
‘logical situation’ in fact is: we alter their sense to convey what we want - the transcendental
story - but we do not ‘leave’ our language. Perhaps it would be clearer to change Buchdahl’s
image of the ‘from without’ to one of the ‘from within’. The point put as follows helps to
illuminate better our situation:

Grammatical understanding and grammatical concepts, even the most familiar ones
like sentence, verb, noun, are not so straightforward and down-to-earth a matter of
plain physical realities as I believe people sometimes suppose. The concept of a
noun, for example, is far less of a physical concept than that of a coin; for someone
might be trained to recognize coins with fair success though he knew nothing of
money, but no one could be trained to recognize nouns without a great familiarity
with language; and yet the concept of a noun is not one which he will automatically
have through that familiarity, as he will have that of a coin if he operates with coined
money\textsuperscript{67}.

So, essence claims get their justification by appeals to what the meaning of the term
we want to elucidate is, but this meaning is of our own making and is not depended on
something extraneous to our way of representing objects, as in the Kantian view of
‘substance’. Consequently, in stating the essence claim we do not go outside what the
constitutive principles of possible experience are, we do not go outside the proper domain of
the transcendental\textsuperscript{68}. Everything we need is our concept of substance.

‘Appearances’ in Kant is then a determination of ordinary objects insofar as they
have their essence determined by us. In saying that what we know are only ‘appearances’,
Kant does not want to preclude our knowledge of ordinary objects by interposing between
them and us an intermediary. Thus far, the transcendental view of the content of our
representations states that their content is what is represented (with obvious qualifications
about existence, cf. for this next section). Now, it is because what is represented is subjected
to rules, as when we say that X is a substance, that we manage to think of objects.


\textsuperscript{67} Anscombe (1965), p.8.

\textsuperscript{68} In Wittgenstein’s approach, this point is made as follows: ‘The connection between “language
and reality” is made by definitions of words, and these belong to grammar, so that language remains
self-contained and autonomous’ (Philosophical Grammar, #55). That is, the term to be explained
and the explanation pertain to the same domain.
Again, Kant’s categorial concepts relate *a priori* to objects because through them we state what is meant by ‘object’. We should not forget that these concepts, in stating the conditions of experience, state at the same time the possibility of the objects of experience. Kant’s transcendental idealistic account reveals then the *internal* relation between these concepts and their meanings in such a way that they can *go out beyond themselves*. This has a consequence for how content - including the issue of ‘mental content’ - has to be conceived from the transcendental perspective. The content of a representation, of a thought, will be the object it represents, the object it is a thought of, not something in us which is a representative, a surrogate of the object, a true *tertium quid* functioning as an intermediary between the subject and the object. I can, thus, say that Kant does not hold a *simulacra* conception of content and intentionality. Let me argue this point in the following.

Kant’s transcendentally idealistic notion of object could - if one is still persuaded by the psychologistic conception of Kant’s views - be thought of as constituting a *simulacra* theory of content and intentionality. Objects would be really known only insofar as they produced *simulacra* of themselves: their properties would somehow migrated into our minds. Consequently, the identity of, say, a thought would be given by the *simulacrum* of the object, which would be in our possession. This would explain for example how we know what we think and how we know that what we think is the thought of such-and-such an object. So it may seem that the *simulacra* conception has to be true, otherwise we would not know what we think, which we normally take to be necessary if we are to be thinking something at all. What is not required is that the object of the thought exists, but the *simulacrum* of it (in other words, what is called the content of the thought) must be in our possession, otherwise thoughts would not have a content, which is unacceptable.

From what I have been saying about Kant’s approach, it is not plausible to believe that Kant holds a *simulacra* conception. Kant is explicit in his repudiation of such a viewpoint.

...For I can only know what is contained in the object in itself if it is present and given to me. It is indeed even then inconceivable how the intuition of a present thing should make me know this thing as it is in itself, as its properties cannot migrate into my faculty of representation.


70 . Prolegomena, p.26, AkIV282.
Kant is suggesting that I manage to think unerringly of an object not because its features flow from it into my mind, as if they migrated into my mind producing there a simulacrum of the object. Kant is here in fact writing off one term of what is usually conceived as a relation between an object and ourselves, Kant is dispensing with the mental representation in us. Because what has to be explained - how I can think of an object in terms of its essential features - cannot appeal to something coming from the object which lodges itself in our minds and then guarantees that our thought has that object. The account Kant provides claims rather that objective meaning is given to ‘mental states’ - enabling them to go out beyond themselves - by subjection to rules which necessitate us to connect ‘mental states’ in specific manners. One of our ‘mental states’ will then be a thought of X because it is necessarily linked - through a rule - to other ‘mental states’ which concern X: my thought is of X if my expression of it is in terms of ‘X’.

7. An Illustration of the Transcendental View of Content in Contemporary Debate

The transcendental view of content which I constructed out of Kant avoids the representationalism present in many of the positions in the contemporary discussion of the content of, for example, beliefs in ordinary belief-ascriptions. However, representationalism has also been attacked forcefully from Wittgensteinian quarters in recent times. I would like to present this attack and the positive views it enshrines in outline in this section as an illustration of the transcendental view of content.

It is easy to think that in an ascription of belief like ‘A believes that p’ the that p states the content of the belief so ascribed. But then the point gets construed as the dictum that the content is a proposition. This last move is wrong.

Firstly, we should keep the following apart. It is one thing to have a belief (with content), and another is to believe a proposition. To believe a proposition is to believe, say,

71. Cf. Kitcher, loc.cit., for a wholly different interpretation of this passage. She thinks Kant is dealing with the psychological problem of perception and that his point is that we can only derive information about objects by the effects of their various sensible properties on our sensory organs.

72. The account that follows - which I believe also represents Kant’s considered position, revealed in the rational reconstruction presented in the previous section - is drawn in toto, in this formulation, from the works of Anscombe (1965), White (1972), Malcolm (1991) and Hacker (1992), who in turn acknowledges his debt to White. My presentation follows closely the points made especially by Hacker, cf. op.cit. pp.256-61.
Davidson's proposition that mental events are identical with physical events. The condition that has to be satisfied for someone to believe Davidson's proposition is that Davidson did in fact put it forward; as when you believe gossip or stories, they have to have been told. One may, in similar fashion, believe that a certain proposition is true, as one can expect, hope, fear, it to be true. But then - and we should be able to see this - the content of one's belief, expectation, hope, fear, is clearly not the proposition that \( p \), but that the proposition that \( p \) is true.

Secondly, another important distinction is that between a belief and its content when we have the following kind of considerations in mind. One has a belief, but one does not have a content. A content is what is believed, and this is not possessed by anyone. Furthermore, as what is believed, the content cannot be fervent, passionate, firm or tentative, as possibly some of one's beliefs are. The content of one's beliefs, expectations, etc, may instead be certain, probable or possible; but not those things a belief can be. I will return to this.

Now the kind of ownership implied in the having of beliefs, expectations, hopes, fears, etc, should not be misrepresented. The ownership is not a literal ownership - it is in truth not an ownership at all. It does not have the model of, say, owning a car or a football kit. That is, it is not the ownership of a thing, an entity. What is crucial here is that what comes first is not the 'owning' of 'mental entities' (beliefs, expectations, etc), it is rather what a subject, a person, does. What comes first, in the proper logical assessment of the issues, is what can be put in the form of a general sentence: 'A Vs that \( p \)'. What happens then is that we form a noun from the verb V. We go from 'A thinks that \( p \)' to 'A has the thought that \( p \)'. But we should not be taken in by this, we are not hereby referring to an entity, a thing, which is supposed to be mental, and all the other properties normally associated with it. It is more illuminating to say that beliefs, thoughts, expectations, etc, are logical constructions. This should not be understood as a denial of the existence of beliefs, thoughts, etc, it is meant simply to deny that they are kinds of things\textsuperscript{73}.

\textsuperscript{73} The claims made here have the same purpose as those presented in chapter 3 against Strawson's view of experiences as particulars with numerical identity. When it was said that sensations and other contents of consciousness have only general identity the point being made was exactly the present one against the literal owning of a thing or entity (cf. infra the discussion about the individuation of beliefs, etc). There the manifesting of pain was brought to the fore, in place of the having of something. Strawson's position was then attacked using arguments concerning more specifically sensations and what can have phenomenological features. In the present context,
When we realize that what comes first is that a subject or person Vs, we can also see that the adverbial mode of Ving - 'passionately', 'firmly' - transfers as it were to the belief, thought, expectation, etc. So we can say that A's belief is fervent, passionate, firm, etc. Here, however, we should be warned against another possible source of confusion.

We use 'thought', 'belief', 'expectation' to designate what one - with the provisos made about ownership - has when one Vs that $p$. But we also use those terms to designate what is Ved, that is, what is thought, believed, expected, etc. We say in a possibly misguided way: 'That $p$ is my belief (thought, expectation, etc)'. But it would be an error to think now that because beliefs are fervent, passionate, firm, that $that$ $p$ could also have these features. What happens in this case is that we characterize what is Ved by reference to the Ving of it, or we could also say qua Ved. For example, we characterize that $it$ is raining by the believing or thinking of it, or qua believed or thought. However, it still is the case that what can be said of the belief, thought, etc, even qua what is believed, thought, etc, cannot be said of what we 'have' when we believe, think, etc, that $p$. Furthermore, in spite of our way of speaking - i.e. as if $that$ $p$ were one's belief, thought, etc (which should be understood as meaning simply 'A has the belief, thought, etc. that $p$') - it makes no sense to say, as we say of beliefs, thoughts, etc, that 'A has that $p$'. We should also realize that when we call what is believed, thought, feared, etc, our 'belief', 'thought', 'fear', etc, - as we also do in relation to what one 'has' when one believes, thinks, fears, etc, that $p$ - we do not want to be taken as implying that we believe our beliefs, think our thoughts, fear our fears. We certainly do not want to be taken as implying that our beliefs, thoughts, fears, etc, fail to reach out to reality.

So, although the distinction between beliefs and their contents is very important, there is one respect in which it is of no use. What one 'has' when one Vs that $p$ is individuated by reference to what one Vs, i.e. that $p$. In this vein, as we saw, we say that that $p$ is A's belief, thought, etc. But we also say that that $p$ is what A believes, thinks, etc, and this should not be taken to imply, as I just reminded, that A believes her beliefs, thinks her thoughts, etc. When the question about what A 'has' is our viewpoint, this latter way of speaking or expressing ourselves - 'A believes that $p$', or 'What A believes is that $p$', about everything that will be discussed can be brought against Strawson's view of thoughts as particulars (this is a necessary consequence of the view of 'I think that $p$' as a cognitive self-ascription), with the charge being that it makes the intentionality of thought unmanageable. Allison's account of the mind's representations as 'subjective objects' that belong to me faces the same kinds of difficulties (cf.1983, pp.258-63). I shall return to this point in Allison in the last two chapters.
which we said that it implies that our beliefs go outside what is 'had' - can suggest a problem of individuation which seems to be too quickly dealt with in our other formulation: ‘That \( p \) is A’s belief, thought, etc’. The pressing questions then are: what individuates what one has; and if it is the content of what one has, what is a content and how does it achieve this?

That beliefs, thoughts, etc, are individuated by their contents should not surprise us. If we really take A Vs that \( p \) as what comes first, then what A has derives from what A Vs, and that is that \( p \). This that \( p \) is the content of what A ‘has’, so the content individuates what A ‘has’, or what A Vs before we nominalize it to what A ‘has’. As an example we could think of A’s intention to go to the beach next weekend. What A ‘has’ is individuated by what he intends to do, which is to go to the beach next weekend.

In trying to answer the question of what a content is it is useful to see how content individuates beliefs, etc. It is what A ‘has’ when A Vs that \( p \) which has a content, and not the Ving of it. The belief that \( p \) has the content that \( p \), but not the believing of it. The believing has psychological features, whereas the belief has logical ones.

What I said above about beliefs, thoughts, expectations, etc, being ‘logical constructions’ was a qualification with their content in mind. Contents are not contained in A’s belief, thoughts, etc. This is because, firstly, a belief or a thought is not a thing which could ‘contain’ a content. Secondly, contents are not themselves contained in anything, they - in this respect, as beliefs, etc - are not things: ‘that \( p \)’ is not the name of a thing. Content is rather what is given in answer to ‘what is A’s belief (thought, etc)?’, and that is that \( p \). Again we should insist that what can be put into the words ‘that \( p \)’ is not something which the subject literally has and of which we ought to wonder how it is connected with reality, as if we had to connect metaphysically or otherwise A’s thinking and the world. It is A’s thought - not his thinking - which ‘has’ a content, and is in fact individuated by such a content, but what A thinks is that \( p \), not a content, as A does not believe her beliefs, fear her fears, etc.

What individuates a belief, thought, etc, is then the expression in words of the belief, thought, etc\(^\text{74}\). This is ‘that \( p \)’, which we should by now be convinced does not introduce any other entity over and above what is referred to in ‘A thinks that \( p \)’. This is

\(^{74}\) Cf. Anscombe (1965, pp.8-9).
exactly in the same vein as when we insisted on what comes first: no new entity is introduced by saying that A has a thought, that is, A thinks that p.

From this we should be able to see that although a subject can believe the proposition that p, in the sense in which one can believe gossip and stories, it is an error to take this to be the same as the subjects’s belief that p. When A believes that p, then that p, and not the proposition that p, is A’s belief. For in the sense required the proposition that p cannot be A’s belief, in the same manner in which the gossip or the story that p cannot be A’s belief. Furthermore, A’s belief that p may be exactly what B doubts, fears, etc, and to doubt, fear, etc, that p is not to doubt, fear, etc, the proposition that p.

These considerations should be enough to preclude one from maintaining that the content of a belief is a proposition, one important representationalist tenet. The source of the temptation to hold them to be the same can be traced to a failure to differentiate the accusatives of ‘A believes...’. There are object- and intentional-accusatives fitting this sentential form75. The denial that the proposition that p can be the content of the belief that p hinges on the recognition that to believe the proposition p is to believe an object-accusative, that is, something which must exist if anyone is to believe it. This is also the situation with gossip and stories. If we are to believe them, or to believe in them (as with the case of a person), they have to exist. Now when the sentential form gets filled by ‘that p’ we are dealing with an intentional-accusative. This accusative also gets different verbal expressions, as when we say that A believes in the Devil, praises English football skills or diagnoses an Ebola infection. The point of this distinction is to show that the content of a belief is in this case its intentional-accusative. So the content of one’s belief, what individuates what one ‘has’, is in truth determined by the intentional-accusative of the verb ‘believe’, and this is that p, which is not an object-accusative, and it is not of the nature of a proposition, a piece of gossip or a story. Therefore, when we say that someone believes a proposition, a piece of gossip, a story, or even a person, we should not be misled by this way of speaking or expressing ourselves. The contents of those beliefs are not a proposition, the gossip, the story or the person, but rather that the proposition, gossip, story, and what the person said are true.

75. This is one of the major contributions by White to this discussion, cf. (1972), pp.70. The point is also made by Anscombe, cf. (1965), pp.9-11.
This is now the right place to discuss intentionality. The conception of content I have been combating, that which maintains that one literally has *something* when one has a belief (that obviously has a content), offers an answer to the problem of intentionality along well-known representationalist lines. Even if what one believes to be the case is actually not true, one can still believe what one does, for what one believes can be said to have an 'intentional existence' as the content of one's belief. However, from what we have said it should be clear how unsatisfactory this answer is. The position is unavoidably committed to maintaining that, because what we believe may not be the case in reality, what we believe is the content of one's belief. We have already seen how unsatisfactory it is to be committed to the view that in a certain situation in which our belief is in fact true what we really believe is a content-entity, which would at best stand for something in reality, and so correspond to what is the case. This position does not, therefore, provide a plausible account of how we actually manage to believe or think what is the case. For what we believe, in this case, is only the 'intentional object' of our belief. This position, by appealing to an 'intentional existence' of the 'intentional object', is definitely committed to an intermediate between us and reality, given that if one believed that \( p \), and it were not a fact that \( p \), one would still be able to identify one's belief as *that* \( p \), which has to be the same 'intentional object' of our belief when it is a fact that \( p \). In this way our belief does not exactly reach up to reality. How we manage to do that remains a huge difficulty for such a position.

This is why approaching the problem of intentionality through asking how *something* we have ('inside us') can be about something in the world ('outside us') is misguided. We should be able to see by now that a thought is not a kind of entity with things inside it and others outside to which it must be in some sort of relation, not yet wholly explained. When we say that A has the thought that \( p \) we are not implying that A has a thing, a piece of reality (mental reality), which must stand in some material relation to a thing which is physical (extra-mental).

Sometimes it is thought that such problems with the notion of an intentional object can be easily disregarded if we take this object to be the object a belief, thought, fear, etc, is *about*. If the object does not exist, then it is simply a case where the referential part of the proposition (which is the content of the belief, thought, etc) does not manage to refer. This
argument is deemed to be ontologically economical, insofar as the intentional object will just be any normal object in the world, with no special status at all\textsuperscript{76}.

This attempt to distinguish content (as a proposition) and intentional object does not succeed because it works exactly under the same assumptions as the conception of the intentional object as a content-entity. The attempt has to concede that the intentional object, the ordinary object, may not exist; but it cannot make do without ascribing contents to beliefs, that is, without - for reasons parallel to the Brentano type of account - being committed to the existence of the proposition; otherwise there would be nothing for the subject to believe. This is exactly the same situation as with the view of the 'intentional existence' of the intentional object. Furthermore, the attempt contains the same kind of shortcomings. For when a subject believes that \( p \), the content of her belief is the proposition that \( p \) and it is this proposition which is believed, given that the intentional object may not exist. This involves the same problem as before: we end up not believing what is the case even when one's belief is true. In Searle this shortcoming is quite clearly present in his view that the mental state stands in a relation of fitting with the states of affairs which are in the world and satisfy the beliefs, fears, etc\textsuperscript{77}. This is mistaken, for when I believe that it is Monday then what I believe is what is in fact the case if it is Monday, certainly not something that, as an intermediary, a tertium quid, stands in a relation of 'fitting' with what is the case.

An account of intentionality which does not fall prey to these criticisms is one that stresses that when a subject thinks that \( p \), the verbal expression of the subject's thought - the sentence '\( p \)' - connects with reality because of what the constituent parts of the sentence signify or refer to. The subject's thinking - and succeeding in so thinking - of a certain object or situation is dependent solely on what the parts of the verbal or symbolic expression of the thought manage to be about. Now we should be able to appreciate that this is so because 'that \( p \)' is not the name of a thing, but is rather the verbal expression of a belief or thought

\textsuperscript{76} This is the view held by Searle (cf. 1983, p.16). As stated in the presentation of his position, Searle does in fact hold that the content of a belief, thought, etc (mental states for him), that \( p \) is the proposition that \( p \): mental states have propositional content (cf. op.cit. p.17). Searle, then, thinks he is advancing matters by sustaining that the content of the states is not their intentional object, which for him is the object they are about.

using ‘p’. So when I think that Salman Rushdie is safe somewhere in England I think of
Salman Rushdie even without knowing where he is, not because my thought somehow
‘represents’ him, it is rather because the expression of my thought would involve exactly the
words I used, where ‘Salman Rushdie’ refers to the British citizen threatened by the Fatwa
of the late Ayatollah Khomeini. By the same token it should be clear why that p does not
‘represent’ an object or complex of objects in the world. It is because we have to provide for
the possibility that it may not be the case that p. Intentionality, then, resides with us and with
language, and it does not depend on, as it were, touching the objects and the states of affairs
in the world (leaving aside whether states of affairs are in the world at all). It depends solely
on our beliefs, thoughts, etc, and their expressions, which among other things involve
referential terms. This is possible because - as Kant would have put it - expressions are
‘connected by rules in one specific manner’ (cf. A197/B242).
 CHAPTER V

KANT ON SELF-CONSCIOUSNESS: SELF-SUFFICIENCY AND EXTERNALISM

1. Introduction
Kant’s thesis that ‘the numerical unity of this [transcendental] apperception is thus the a priori ground of all concepts’ (A107)¹ expresses what for him is fundamental in the investigations he is carrying out in the Analytic of the first Critique. In this chapter I will analyze this claim by Kant and argue that the unity of transcendental apperception Kant is interested in is a transcendental ground which should be conceived in an externalist manner. This ground, according to externalism - and this is the import I want to give to this term - is the transcendental subject conceived of as a corporeal agent that exercises its abilities to perform overt actions which it is conscious of itself as the subject of. I will argue for externalism in section 6 of the present chapter. In section 7, I will seek to determine, as a continuation of what came before, the place of Kant’s view of consciousness of self in his arguments in the Deduction according to externalism. These two sections constitute the clarification of one aspect of Kant’s claims about transcendental apperception: the concern with the basis for the whole phenomenon of concepts and judgements having objectivity, which is the possibility of determining how things are independently of ourselves. There is, however, another aspect to these claims which has to do with the characterization of consciousness of self or self-consciousness as a faculty of thinking subjects. The self-consciousness expressible in the use of ‘T’ is a faculty with distinctive features for Kant. Concerning this latter aspect I will argue that its central characteristic is self-sufficiency (this is the purpose of section 3). For this, the results of the previous chapter will prove important. How the two aspects combine in Kant will also be the topic of section 7. Let me now state what will take place in the remaining sections of this chapter.

¹ Kant’s thesis appears at A118 in relation to synthesis, and at A127 in relation to the necessary conformity to law of all appearances. Cf. also B134na for the view that the synthetic unity of apperception is the ‘highest point’ to which all employment of the understanding has to be ascribed.
At the beginning of the last chapter I pointed out the differences between Strawson's and my view of self-consciousness in Kant. I would like to start this chapter with a closing of my discussion of Strawson on self-consciousness in this thesis. In section 2, I would like to examine Strawson's position on its own. This has the goal of preparing by contrast what I will present as Kant's position about the faculty of self-consciousness. Here is the place, however, to state quite clearly that my objections to Strawson do not concern his bringing in of persons (corporeal human beings), but rather the route by which he brings them in.

Sections 4 and 5 examine some of the merits and shortcomings of the views of two very important Kant scholars - Allison and Henrich - on the issues of my interest. The main point of these two sections is to help to clear my way from some common obstacles in getting Kant's intentions right.

2. Strawson's Account of Self-Consciousness: the Lack of Self-Sufficiency

Self-consciousness is not self-sufficient in Strawson because it depends for its successful exercise on the satisfaction of conditions which are extraneous to what is available to the subject's awareness. Strawson sees as crucial a conceptual relation between the criterionless self-ascription of experiences through the referential use of 'I' and the empirical criteria for the individuation of persons. This is an aspect of Strawson's account which is captured by his tenet that the possibility of self-ascription presupposes other-ascription². What is problematic about this tenet is not that the use of 'I' is linked with the use of 'you', 'he', 'she', 'they'; but rather that the use of 'I' is linked with empirical criteria for the identification of persons which in the case of first person present-tense indicative mode of psychological utterances (what Strawson calls self-ascription of experiences) are appealed to but not in fact employed. As we saw in chapter 3, Strawson thinks that self-ascriptions of experiences are criterionless. It is easy to agree with Strawson that it does not make sense to ask 'this inner experience is occurring, but is it occurring to me?' The problem lies with the philosophical assessment of this situation, with the assessment of the relation between self-ascription on the basis of a referential use of a word and the criterionlessness of the use of exactly the word in question: 'I'. As R.A. Noë remarks about Strawson: '[He] is correct [about the nonsense of that question], although some reservations may be expressed as to

whether he appreciates the full import of his argument. As I will argue, self-consciousness
is made dependent by Strawson on an unwarranted feature of 'I' - viz. that in the relevant 'I'-
sentences the pronoun is referential in spite of the non-employment of empirical criteria or
rules for the identification of a subject of experience.

Strawson's general approach to these matters about the use of 'I' is informed by a
desire to guarantee that the 'this' which designates an experience has something to latch on
to. In chapter 3, I challenged Strawson's view of experiences as particulars. There I
indicated that Strawson's view that 'I', despite its criterionless use, is referential requires a
general rule that should be used even in self-ascriptions of experiences: even in these cases
we use a clause present in the concept of a subject of experience, i.e. such a subject has
identity as an object of outer sense. It is the satisfaction of this particular clause which
guarantees the successful use of the concept in the first place. It is the necessity of such a
rule which will come under scrutiny in the ensuing discussion. Is it really for first-personal
use, or is it only for the use of others? Do we use 'I' possibly observing the import of such a
concept of a subject of experience? What is the point of saying that we use 'I' successfully
(referentially) because of criteria which we do not in fact employ (never in certain types of
sentences), but others may?

Strawson's approach involves the view that 'I experience such-and-such' is always
cognitive. It is not clear, however, how this approach intends to guarantee that I, in such
utterances, self-ascribe experiences. Without bothering himself with such a question,
Strawson moves on as follows. It is because experiences are dependent particulars that the
ascription of them will only succeed if 'I', in what are deemed to be self-ascriptions, actually
refers to something objective. This means, refers to something in the world, as an object of
outer sense. This is the central tenet. Strawson's account of persons develops from this
point. Persons are just such kind of things because persons have bodies. The troublesome

pp.277-83.

4. Compare this with Tugendhat's distinction between an identification of myself and a reference to
myself: with 'I', he says, I do not identify myself, but I do make reference to myself as an individual
identifiable by other singular terms which may substitute 'I' (cf.1993, p.18). However, with
Strawson reference is identification and 'I' does referringly identify a subject of experience even to
myself.

component of this outlook, however, is the claim that ‘I’ refers exactly because there are empirical criteria for the identification of subjects of experience which nonetheless I do not use when I refer to myself. This is the basic issue here. I will try to show that his view requires that my competent use of ‘I’ is dependent upon a rule which I, in fact, can not appeal to to guide my use of the word, and that this notion of dependency is not defensible.

Let us first see how Strawson envisages his account should work. On page 167 of the *Bounds of Sense* he insists that "for any empirical (i.e. legitimate) use to be made of the concept of a subject of experiences it is required that there should be such empirically applicable criteria of subject-identity as are supplied by our ordinary concept of a person as something which, *inter alia*, is an object of outer sense*". Thus to use the concept of the subject of experiences in an empirical judgement which intends to have truth-value, that is, be objective, we have to be able to identify this subject, at least in part, as an object in the world, as having a location. The astonishing move by Strawson comes now. In the sequence of the text, he asserts that "this rule is general*. By this he means that the concept of a subject of experiences, with the requirements about the identity criteria built into it for the sake of its very intelligibility, is used (1) from the third person perspective, when we refer to such a subject and ascribe an experience to her, and (2) from the first person perspective, when I refer to myself in the process of self-ascribing an experience. This is what is meant by the claim that the rule is general. The assertion is intended to bring the first personal case under the scope of application of the rule about the objective use of the concept of a subject of experience. Strawson is emphatic about it: ‘Not even the use of the concept of a subject which is made in ascription of immediate experiences to oneself (in consciousness of oneself as being in such-and-such a state) would be possible unless this requirement were satisfied". This means that when I say ‘I experience such-and-such’ the criteria of identity have to be satisfied, and this has to appear in the claim I make about myself as having an experience, otherwise I would not be claiming something objective at all. But according to Strawson in such cases I do not myself employ the empirical criteria of identification. So, I

---


have to use a concept to make reference to myself, and in using the concept I succeed in making the reference because of empirical identity criteria whose use I myself do not however make. Can Strawson accommodate the pressures coming from these opposite sides? Can this view be defended?

Strawson would claim that he is dealing with the following simple idea. A subject of experiences in the self-ascription of them uses the word 'I' to make reference to herself in accord with criteria which allow others to individuate the subject in question. I would like to point out at this juncture that it is exactly this perspective - that of others - which has the primacy in Strawson's approach. Strawson thinks that by giving primacy to this perspective he is doing justice to the required public component in every use of an expression. But instead of construing the public component in a suitable way he is forcing upon the character of 'I'-sentences a feature only appropriate to cases where we describe or attribute experiences to others.

Strawson's way of solving the problem for his approach is to claim that there is a public language whose structure allows certain predicates to be ascribed in both first and third person cases while remaining univocal. In the former kind of uses observation is irrelevant, but in the latter it is crucial to observe the behaviour of the subject to which the predicates are to be applied. Now, Strawson claims that to learn how to use these predicates is to learn how to use both ways in which they are applied. Otherwise we would not have this type of predicate, or we would not understand the particular instances of the type in question. This type subsumes what Strawson calls P-predicates, which are predicates applicable only to persons. The task now is to see whether Strawson's conception of the structure of the public language concerning P-predicates can really be such as to involve the pre-eminence he gives to the observations from the third person perspective. I will leave aside, however, the P-predicates as such and turn back to the origin of them: the concept of a subject of experience. I will examine Strawson's implicit claim that the use of this concept by each of us when we say 'I' manages to be referential just because of the structure he believes our language possesses.

The central element of Strawson's position was presented to be that, in the use of 'I', a subject that self-ascribes experiences is in accord with criteria which allow the

---

identification of her as the subject of the experiences. But what about what the subject herself must know to use 'I' successfully? Surely the subject must know the meaning of 'I'. This, according to Strawson, involves the concept of a subject of experiences which must be applied in *general* as involving criteria (rules) for the empirical identification of persons. Now, if it is required that the subject uses 'I' in accordance with rules that state how subjects may be empirically identified because otherwise she would not be using 'I' meaningfully, then when the subject learns the meaning of 'I' she also learns how to follow those rules. However, if learning the meaning of 'I' is learning how to follow those rules for the empirical identity of the subject, then one is in fact *using* those rules, *following* them, when one uses 'I' meaningfully. But this can not be in accordance with Strawson's view, since he is explicitly committed to the claim that the subject does *not use* the criteria or rules for the empirical identification of a subject in self-ascriptions of experiences.

One idea, which attempts to help Strawson on these counts, is to say that in spite of not *following* the criteria or rules the subject does in fact act in *accord* with just this type of criteria or rules. But this, as will become clear, is not a defensible claim. This cannot justify the view that the subject is using 'I' to refer to herself.

If the subject is said to be referring to herself because of the existence of a rule with which she is in accord but which she is not following, then in fact no reference is being achieved by the subject herself. To be in accord with a rule is not to follow a rule. Let me bring this distinction and its point to the fore through the means of an example.

Suppose you play a game of football in which it was agreed that the off-side rule would not be observed. There will be attacks the teams make in which the strikers will receive the ball without being beyond the imaginary line drawn at the point where the last defender is (that is, attacks where the strikers would not be off-side in normal games). In these cases it would be intelligible to say that the attacks are in accord with the off-side rule, but this would be accidental in the supposed game, given that the rule is not being used (we can imagine someone trying to observe the rule alone, say to play better when it comes to be observed, but this is a different case, and this person is certainly not going to play well the game as it was agreed it would be played). So, it would also be intelligible to say that other

attacks are in disagreement with the rule in question, and this would also be accidental. The strikers in both kinds of attacks would be according or disagreeing with the off-side rule, but they would not be following, or even trying to follow, the rule. That is, they would not be observing, trying to observe or, on the contrary, infringing wilfully the rule. They, in principle, could not be observing (etc) it, given that it was agreed that the rule would not be applied, which implies that nothing they did would count as observing or infringing the rule. This is equivalent to saying that in this game there is no off-side rule. This example shows that simply to accord with a rule in cases where nothing you do counts for you and others as complying with or infringing a rule is not in fact following or failing to follow it. In the case of your decision about what to do in the game as it was agreed it would be played, this rule with which your action only accords or is in disagreement with is as good as nothing.

To claim, then, that it does make a difference for your action to be in accord with a rule, without you being in principle able to follow it, is not intelligible. But, this is the nature of Strawson’s claim about the use of ‘I’.

If the subject were to follow the rule about the referential use of ‘I’ then she would be identifying herself by the use of the empirical criteria normally used to identify any subject of experience. But, then, and this directly contradicts Strawson’s other claim, self-ascriptions could in principle be doubted, given that empirical identifications following rules sometimes fail. However, as Strawson most definitely wants to cling to the impossibility of this doubt, his true position must be that which I have been criticizing. Therefore, Strawson is positing a rule with a certain content that must be accorded with, but which cannot be followed, given that nothing counts as failing to follow this rule, not even the claim by someone who, deluded about his identity, says ‘I, Ludwig Wittgenstein, wrote the

11. Even if Strawson rejects the move on his behalf, i.e. that of distinguishing the accord with and the following of a rule, and claims that the subject is following the rule of reference although not employing it, this would not be good enough. Because, as the example tried to bring out, nothing a subject could do would count as following or failing to follow the rule in question. So, there is here no following which would be a non-employing of the rule. I suspect that it is because of this point concerning a rule of reference that Hacker remarks that to concede in Strawson’s case the criterionlessness of self-ascription of current or recalled experiences does not suffice to remedy what are the flaws in the account. ‘For the possibility, the legitimacy, of such criterionless employment of these concepts, their use without a justification but with right..., calls for further elucidation’ (1996, p.316 n121).
Tractatus\(^{12}\). The conclusion of this must be that ‘it is incorrect to speak of a rule of self-reference here at all\(^{13}\).

It should not be doubted that Strawson’s intentions are to present a tenable alternative to what he takes to be the doomed position of Cartesianism. He wants to account for self-ascriptions without any appeal to soul-substances. Furthermore, Strawson sees his referential view of ‘I’ and the linked view of cognitive self-ascriptions as representing a clear statement in epistemology of the pre-eminence of the public over the Cartesian special inner and private self-awareness, which in spite of their independence from the objective world were seen as the only secure basis to construct any knowledge about the world. However, Strawson’s view about self-consciousness makes it wrongly dependent on something on which it cannot depend. The main point is that in saying ‘I think such-and-such’, I do not refer to a subject to which I ascribe a particular thinking experience. What is wrong with this is that the use of ‘I’ is made dependent on something on which it cannot depend, and it need not, as we will see. To claim that it does is not to recognize that the use of ‘I’ is self-sufficient, that all that it needs is the following of the rule which gives its uses their meaning. What occupies a special place in relation to self-consciousness is the irrelevance of the consideration of anything empirical for the use of ‘I’, especially of myself as an object which I would referringly identify.

3. Kant’s View of Self-consciousness as Self-Sufficient

From the previous discussion of Strawson it should be clear that a subject who uses ‘I’-sentences could not be referring to herself as a being in the world just because she is using ‘I’ in accord with a purported rule of reference to subjects of experiences conceived as such beings. The following of such a rule was rejected because nothing the subject could do in using ‘I’ would qualify for a following or failing to follow the referential rule.

The following of a rule is not available for an account like Strawson’s because of the content of the rule his self-ascriber of experiences should be following: she should be referringly identifying an object in the world because she wants to ascribe particular


experiences she is aware of to their subject. The point I want to make is that in using 'I' I am using rules, but not of the Strawsonian kind. What is important here is that the subject has the mastery of the use of 'I'. And this requires, pace Strawson, rules the subject can follow. Now, it should also be clear in this respect that 'it is the potentialities of an agent for using a rule in normative activities that is crucial for his following this rule, not the actual occurrence of a rule-formulation in his visible activities'\textsuperscript{14}.

In the case of 'I', however, the rule that has to be followed is not one which includes the knowing or believing by a subject that with a certain person or with a certain thing in the world things are such-and-such, or so I will argue. This is an aspect of the use of 'I' which is easily overlooked. The meaning of the expression is normally thought to be exhausted by the following rule of reference: 'Its reference is to the speaker, and [the rules governing its use] leave no latitude to the speaker's intentions in the determination of its reference'\textsuperscript{15}. But this cannot be so: there can be no rule determining behaviour which is totally independent from whatever intentions a speaker may have in uttering something. I tried to bring this out by insisting on the requirement that the rules to be followed by a subject must necessarily be available to her as guidance for his behaviour, and by showing that this does not happen with Strawson's type of rule of reference. Whether the subject is to have success in using 'I', to use it correctly, this depends only on his mastering the rules for the use of the word. However, if in certain contexts he uses it successfully without the possibility of him being using it due to a rule that has a determinate content, then he can not be using it because he mastered this particular rule. And this is the case with Strawson in my view.

So, the availability to the subject of a rule - in order that there is the possibility of the rule making a difference to the subject's behaviour and so of informing it - is also required in the present case: the rule for the use of 'I' must be specified so that it can make a difference for the behaviour of the subject when it guides the subject. But, then, concerning 'I' there are two well known immunities to failures in using the word. These are presented as the immunity to lack of reference and the immunity to mistaken reference. For reasons already indicated the issue can not concern reference. But the data advanced as the basis for

\textsuperscript{14} Baker & Hacker (1992), p.158.

\textsuperscript{15} Shoemaker (1968), p.84.
these immunities are well collected: 'I' is used by a subject and 'I think so-and-so, but is it really I who think so-and-so?' is nonsensical. Kant, one could say, reaches a different conclusion from this type of data.

According to Kant, the competent use of 'I' by a subject does not depend on any attempt on the subject's part of making a referringly identification of herself as an object in the world, object to which experiences would then be ascribed exactly because of this character of hers. This is what it means to say that self-consciousness is self-sufficient for Kant. This fundamental feature of this human faculty, brought out in what is needed for its successful use, is stressed by Kant with the centrality he gives to the perspective of a self-conscious subject. It is from this perspective that we can appreciate how and why 'I' is used self-sufficiently for Kant.

'I' is used self-sufficiently,  
with an a priori knowledge of its meaning by its user,  
because it is the presentation of the subject in thought or speech as subject only. This is in fact the only content which Kant admits is conveyed with the use of 'I'. So, the question - 'Who asked for water?' - posed to an amnesiac about his personal identity or even about his being an object in the world is answered properly, with a competent use of 'I' - the answer being obviously 'I' - as far as she is aware that she does present herself thereby as subject. In this context, as the subject who wants water and asked for it, in independence of any other belief about what and how she is.\(^\text{16}\)

The autonomy of meaning is then what explains that a use of 'I' might be competent with exactly the content according to which in using 'I' the user is presenting or positing herself as subject only. I will say more about what Kant takes 'I' to convey in the last section of this chapter, now I want to examine the self-sufficiency of 'I' in connection to what Kant says about the perspective of a self-conscious subject.\(^\text{17}\)

Returning then to the perspective of the self-conscious subject, we should notice that it is discussed by Kant in relation with the categories. I would like to discuss what is the role

\(^{16}\) This example is presented by Malcolm (1986), pp.251-2. On the meaning of 'I', Malcolm illuminatingly remarks that even in 'an "I"-sentence where no possibility of error has been provided for, "I" [does not] mean a thing called "the subject"' (op.cit.,p.249). Malcolm's view is that 'nowhere does "I" have the role of designating or meaning a distinctly conceived object or subject' (op.cit.,p.252, my emphasis on 'a distinctly conceived':AK).

\(^{17}\) Cf. what Kant says on 'I' at A355-6 and on self-consciousness at A378 and A400.
of this perspective in Kant with the help of Dieter Henrich’s proposal of three criteria for the Transcendental Deduction of the Categories. From this discussion, a notion of certainty about the exercise of self-consciousness will be revealed to be important, and thus far it will become clearer why ‘I’ is only used in the presentation of the subject as subject.

In presenting the criteria for a successful transcendental deduction in Kant, Henrich makes some important points. Firstly, and this constitutes Henrich’s first criterion, it is not enough to show that the categories are conditions of experiences by showing that they, as principles, represent the indispensable activity for the accomplishment of knowledge. Henrich claims that what is necessary for the success of the deduction is the demonstration that the person claimed to have experiences actually uses the categories as principles for her experience, which means that she uses them as a priori concepts. ‘In other words the person who has experiences must consciously make use of the categories as such a priori concepts in his knowledge’. This requirement of a conscious use of the categories is the first important point made by Henrich.

Reminding Kant’s interpreters of this requirement leads Henrich directly to a second criterion, which he calls the ‘Rousseauian’ one: ‘The conscious use of the categories must be such that it can be attributed to every rational being with respect to its ordinary empirical knowledge’. By this Henrich means that using the categories in the conscious way required is not something which involves philosophical analysis, nor does it assume that one has as a goal the achievement of scientific knowledge: ‘[The categories] are basic concepts of the experience which arises from the natural use of reason’.

Now, keeping these two criteria in mind and taking into consideration that Kant’s envisaged deduction of the categories is to be related to the possibility of self-consciousness - in fact it is ‘in relation to the possibility of self-consciousness [that] the categories can be regarded as conditions of the possibility of experience’ - we are led in Henrich’s discussion to a very interesting consequence, which represents his third criterion: ‘the deduction cannot confine itself to proving, from the distance of an external analyst, that self-consciousness could not come about without the use of the categories. For this would not be to show the internal connection between the consciousness which is self-consciousness and the natural

---

18. Henrich (1989), pp.252-4. Quotations below are to be found in these pages, unless indicated otherwise.
consciousness of the validity and functioning of the categories. It must be shown from the perspective of the person who is conscious of himself, that his knowledge can not be separated from the knowledge of the categories as a priori concepts and of their actual use.\(^\text{19}\)

The point that the categories have to be used in a conscious way - which, according to Henrich, must bring to the discussion the perspective of a self-conscious subject, because it is only to the subject as far as it is self-conscious that the categories can be a priori rules - takes us to the feature of self-consciousness which I believe Kant wanted to stress. This feature is affirmed, by Henrich, to be present insofar as self-consciousness is disclosed 'with Cartesian certainty'. This particular formulation has been criticized. However, I do not believe that Henrich means more by this, which constitutes his third criterion, than 'with a certainty which is independent of all experience'. To call this a Cartesian certainty, however, precludes the proper understanding of this feature of self-consciousness in Kant Henrich has latched on to.

The self-consciousness about which the argument of the deduction is concerned is a self-consciousness which, according to Henrich, must itself have the status of an a priori certainty. He puts this as follows: 'Whatever the properties of self-consciousness are to which the Deduction refers, they must be known a priori in a self-consciousness which, according to Henrich's use of 'knowledge' in this context, since it cannot have its normal import. Failure to notice this makes one lose the whole point here. Cf. as well pp.256-7: 'But such elements or acts [particular causes in a posited mechanism] would then be part of a causal nexus which cannot be determined a priori and of which one is not conscious in self-consciousness and hence does not control. No conditions of self-consciousness which cannot like itself be clarified from the self-conscious being's own perspective can be given a transcendental grounding and hence be categories, for these are only internal conditions of the unity of an experience accompanied with possible self-consciousness' (my italics: AK). The view is actually that 'the categories, which make possible the form of unity of all thoughts of what is given, would have to be used by the self-conscious subject in such a way that that subject is conscious of their unifying function; hence in such a way that they stand "before its eyes" when it becomes conscious of itself as a principle of unity. And its consciousness of itself consists in nothing else' (op.cit. p.255).


\(^{21}\) Cf. Henrich (1989), p.270, cf. as well p.274 for the view that the Cartesian certainty regarding the ability to use 'I think' goes only as far as the 'I' is related to the contents of thoughts which are 'specified by what is represented in them', by what is 'objective' in them.
although related to experience, is not dependent as regards its structure and mode of consciousness on experience'. Let me state what is Henrich’s point here.

Henrich states that the *a priori* certainty of self-consciousness - self-consciousness can not wait to see if it is met with success, as it would have to if it should referringly identify a subject of experiences in the world - is what requires the independence from experience of self-consciousness.

Self-consciousness is *a priori* certain. The question then is how to defend such a position, *i.e.* how exactly is it supposed to be possible. It is this question which can receive a satisfactory answer when we appreciate Kant’s point about the autonomy of meaning. It was shown earlier that from this tenet Kant derives not only a solution to the problem of intentionality, through an appeal to rules, but also a view of content which I have called transcendental. In order to self-consciousness, which as a faculty is exercised through the use of the representation 'T', to be *a priori* certain, all that is needed is that the content of be the representation be available to the subject aprioristically. So, because self-consciousness is *a priori* certain it has to be self-sufficient.

Self-consciousness as the ability to think, intend, want, etc, by using 'T' is self-sufficient insofar as it does not depend, formally speaking, upon experience or the knowledge of the empirical criteria for the identity of a subject of experiences as an item in the world; it depends rather solely on the grasp of *rules* for the expression of thoughts, intentions, wants, on the part of a self-conscious subject. Otherwise, materially speaking, it depends on the having of thoughts, intentions, etc (cf. B423na for how the 'T' needs representations to supply the 'material': they 'supply the material for thought, [without which] the *actus*, 'I think', would not indeed take place; but the empirical is only the condition of application, or of the employment, of the pure intellectual faculty'. Cf. also *Reflection* 5453, AkXVIII186).

Now, because rules are conceived as what necessitate us to connect representations in specific manners, we are able to know with certainty what this manners are if we know the rules. No causal necessitation could provide us with the possibility of knowing, of having consciousness of the import of the specific manners we have of connecting representations, or *relating them to objects*. Furthermore, it is rules which, as we saw at the end of chapter 4, guarantee the objectivity of the intentionality of thought, that thoughts have a content expressible in a clause as *that p*. The rules also individuate the thoughts,
intentions, wants, one ‘has’, by determining the intentional accusatives of those ‘mental states’. Let us see how this general point is connected in Kant with the use of ‘I’.

To begin with, the use of ‘I’ by us humans is for Kant, as with the possession of categories (concepts with a priori meaning), a fact. In What Real Progress has Metaphysics made in Germany since the Time of Leibniz and Wolff? Kant affirms that the fact that I am conscious of myself indicates a capacity which is elevated above the capacity to receive sensuous intuitions, that it is the basis for the possibility of understanding. Kant’s appreciation of what is occurring here makes him state that ‘it has the effect of separating us from all animals, to which we have no reason for ascribing the ability to say I to themselves’\(^\text{22}\). In the same passage this ability is said to result in an infinity of self-constituted representations and concepts, and this is possible because of the ‘self’-aspect, that is, of the conscious first-personal aspect or the aspect of self-consciousness on the part of a subject.

We find here that Kant is committed to the view that the ability to say I to ourselves is accompanied in its successful exercise by an a priori certainty, so that an infinity of representations and concepts may result from it. Again, the certainty of self-consciousness is something that makes itself visible in the following of rules in the process of conjoining representations. But, then, this a priori certainty about the use of exactly ‘I’ is what makes it a self-sufficient faculty for Kant. It is in this capacity that it constitutes the perspective of a self-conscious subject - conscious then only of being a subject - which is required for the synthesis of all appearances according to concept or rules.

The issues being discussed, the certainty pertaining to self-consciousness and the demand that the perspective of a self-conscious subject has to be present if the account of objectivity is to work, are brought together by Kant in A108.

This transcendental unity of apperception forms out of all possible appearances, which can stand alongside one another in one experience, a connection of all these representations according to laws. For this unity of consciousness would be impossible if the mind in knowledge of the manifold could not become conscious of

\(^{22}\) Op.cit.p.Ak270, my italics: AK. Noteworthy is the claim that we are separated from other animals by the effect of possessing a special ability, that of saying I to ourselves. In the Anthropology Kant writes in the same vein: ‘The fact that men can have the idea “I” raises him infinitely above all the other living beings living on earth... So any language must think “I” when it speaks in the first person, even if it has no special word to express it. For this power (the ability to think) is understanding’ (AkVII127). So, when a child learns to speak in terms of ‘I’ she first comes to think herself (cf. ibidem). Cf. also Kemp Smith (1984), p.LI and ff.
the identity of function whereby it synthetically combines it in one knowledge. The original and necessary consciousness of the identity of the self is thus at the same time a consciousness of an equally necessary unity of the synthesis of all appearances according to concepts, that is, according to rules, which not only make them necessarily reproducible but also in so doing determine an object for the intuition, that is, the concept of something wherein they are necessarily interconnected. For the mind could never think its identity in the manifoldness of its representations, and indeed think this identity a priori, if it did not have before its eyes the identity of its act, whereby it subordinates all synthesis of apprehension (which is empirical) to a transcendental unity, thereby rendering possible their interconnection according to a priori rules (A108, my italics: AK).

What Kant is claiming here is that the categories, which the mind itself introduces, operate as rules, and that the mind has to be able to be conscious of the ‘identity of its act’ in following rules. But what is required for self-consciousness to be possible is that in the consciousness of ‘identity of function’ (of the rules) the subject thinks ‘I’ to itself, an ability Kant also registers, as with the possession of concepts that are rules, as a fact in the account of philosophy. This means that ‘I think so-and-so’ would not be possible without the consciousness that I am following rules in combining the representations involved. And in so doing I use ‘I’ and in that I present myself only as the subject of these representations.

Just how important the following of rules is - including in the expression of the awareness of our doing that - Kant stresses again in terms of its being the condition of the possibility of all meaning: ‘We are conscious a priori of the complete identity of the self in respect of all representations which can ever belong to our knowledge, as being a necessary condition of the possibility of all representations’ (A116).

In the B-Deduction the self-conscious aspect on the subject’s part is stressed several times. There we find Kant saying that synthesis is ‘an act of the self-activity of the subject, it cannot be executed save by the subject itself’ (B130), and ‘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’ (B131-2). In the same context Kant also claims that the ‘I think’ representation is an act of spontaneity (cf. B132).

Kant’s view links closely the possession of self-consciousness, with its identity throughout its manifestation through the self-sufficient use of ‘I’ in connection with representations in general, and the consciousness of the rules used in connecting
representations. 'This thoroughgoing identity of the apperception of a manifold which is given in intuition contains a synthesis of representations, and is possible only through the consciousness of this synthesis' (B133).

However, here we find the indication to something else. The consciousness of the synthesis, of the rules which synthesize representations, is said to be dependent on my conjoining one representation with another (cf.B133). Now, the conjoining, and the related issue of the basis of the synthesis which is present in every concept or representation of which we can be conscious, will receive its due attention when I come to explain later Kant's thesis that the unity of apperception is the transcendental ground for the necessary conformity to law of all appearances, and that it is the a priori condition of all combination (cf.A118, 127).

In this section I have tried to show what it means to say that self-consciousness as a faculty for Kant is self-sufficient: in Henrich's words, it is independent from experience in what concerns its structure and mode and this independence expresses itself in the a priori certainty which accompanies the use of 'I'. In my way of putting it: self-consciousness is self-sufficient because it does not depend for its successful exercise on knowledge of the empirical criteria for the identity of a subject of experience conceived as an object in the world of which we would have experiences. Concepts as rules were shown to be fundamental also in relation to 'I', but then the rule that mattered is one in which the user of 'I' presents itself merely as subject.

In the sequence, we have to address that other aspect of Kant's view of self-consciousness: that the transcendental apperception as a unity of apperception is the a priori ground of all concepts. It is worth seeing how Allison and Henrich deal with this point. This will be the purpose of the next two sections.

4. Allison: Between Subjectivity and Objective Unity

Allison's view about the identity of the transcendental apperception suffers from an indefiniteness. On the one hand, he says things that favour an interpretation of that identity in terms of 'subjectivity'. I borrow this phrase from Henrich: 'Subjectivity means that all thoughts in relation to which the thought that "I" think them is possible do not have this property by virtue of the content that is thought in them. They are 'my' thoughts because I
think them, or can think them\textsuperscript{23}. They are mine because I think them, that is, independently of what is thought in them there is something objective that guarantees that the thoughts pertain to me. Henrich thinks of subjectivity as a form of putting thoughts together which presupposes a unity, as it were, of a subject of thoughts. This can eventually be explained by discovering the causes of this unity or, as he says, combination-elements or combination-acts which are constant in their form\textsuperscript{24}. This is precisely what Hurley understands by 'objectivity' in her presentation of the Strawsonian unity-requires-objectivity argument\textsuperscript{25}.

On the other hand, Allison's account of the identity of the transcendental self is in terms of the 'objective unity', which he conceives as a necessary condition for the identity of complex thoughts themselves. It is the identity of the subject insofar as an act of synthesis is necessary to explain even the analytic unity of apperception or the analytic unity of all general concepts. In this respect Allison reserves a special role for the notion of an act which 'contains' a synthesis and the correlated claim that apperception or self-consciousness is only possible through consciousness of this synthesis. It is in this connection that Allison's view of synthesis as an act receives its importance\textsuperscript{26}. I will make explicit the two lines of Allison's interpretation of Kant, and then see to what extent they are independent. My claim will be that the 'objective unity' option reduces to the 'subjectivity' option, or, in another jargon, the 'unity-requires-objectivity' type of view. Basically, because Allison believes that we have to be able to guarantee ascription of experiences. But, then, Allison does not take

\textsuperscript{23} Cf. Hurley (1994) for the view that togetherness or co-consciousness of beliefs or experiences can only be guaranteed 'objectively', that is, by appeal to persons in the world. Hurley comes up with this 'neo-Kantian' \textit{[sic]} approach, inspired in Williams (1978, pp.95-101), as an attempt to explain Kant's view that the unity of consciousness is dependent on self-consciousness. The interesting point of the paper is the accusation that Kant is a partisan of the myth of the \textit{giving}, the idea that there is a homuncular agent who is unlocated and has unexplained abilities to synthesize. This way of questioning Kant's position would also be reinforced by worries like Walker's: that when it comes to the self there is a 'breach' in the wall of Transcendental Idealism (cf.1975, pp.131-5). Walker relishes this possibility, especially in view of the power consequently ascribed to transcendental arguments, but Hurley sees here a deeply confused view on agency and spontaneity.

\textsuperscript{24} All these points are to be found in Allison, \textit{op.cit.} pp.142-4. Allison discusses these problems mainly in three places in his book. (a) In the Deduction chapter (7): pp.133-58; (b) in the chapter about the Apperception and the Paralogisms (13): pp.272-93; and (c) in part of the chapter on the Refutation of Idealism (14): pp.304-9.
exactly the Strawsonian line, he rather ends up locating the subject that should account for objectivity outside the realm of experiences. A true transcendent point of view in my opinion. I will start examining the ‘objective unity’ option.

Allison’s starting point is the claim that the representation of a manifold as a manifold is a single complex thought. Backing such a claim is Kant’s theory of concepts as analytic unities which are to be understood as synthetic unities of representations. The Concise Oxford Dictionary’s definition of ‘book’ provides an example of such a synthetic unity and also gives an idea of a concept as a synthesis in the sense of result or product, as we have seen in the last chapter.

Allison’s second step is to show that a single complex thought requires a single thinking subject. The argument for this requirement is normally taken by Kant interpreters from William James, and so does Allison: the thought of a whole as a whole requires that each element of the whole be thought by the same single thinking subject, otherwise we would not have guaranteed the thought of the whole itself and could end up with just a set of distinct thoughts of the elements of the whole. Now what James’ understanding of this claim was seems to be clear. It is plausible to think that he was trying to give a concrete illustration of Kant’s verse analogy. In the Second Paralogism, first edition, Kant argues for the simplicity, and not exactly for the singularity, of what thinks in the following way. ‘For representations (for instance, the single words of a verse), distributed among different beings, never make up a whole thought (a verse), and it is therefore impossible that a thought should inhere in what is essentially composite. It is therefore possible only in a single substance, which, not being an aggregate of many, is absolutely simple’(A352). Now it is true that Kant uses the phrase ‘single substance’, but this should be interpreted as substance in its logical meaning: as a logical subject (cf.A147/B186; B288; B300). So, the claim is that thought can only inhere in what is always thought as a subject, never as a predicate, and not that the thought can only inhere in the same single permanent substance. Nonetheless, as a matter of fact, Allison uses James to account for the ‘I’ of apperception - the logically simple subject - of his support text: B407[^27].

James, however, read the verse analogy in a different spirit, and his interpretation is not at all indicative of a logical subject. ‘Take a dozen words, and take twelve men and tell

to each one word. Then stand the men in a row or jam them in a bunch, and let each think of his word as intently as he will; nowhere will there be a consciousness of the whole sentence. So, what is a condition of the unity of a thought for James is that one person thinks the thought. This is the clearest presentation of the point about the necessary supposition of co-consciousness or togetherness of content.

Unfortunately, without paying attention to the distinction I have drawn, Allison's explanation of the crucial principle of the necessary unity of apperception is in fact unclear. He maintains that from the claim that a single complex thought requires a single thinking subject it follows that (1) a numerically identical 'I think' can be reflectively attached to each of the component representations taken individually, and (2) that the thinking subject can be aware of the numerical identity of the 'I think'. Now, what would be a numerically identical 'I think' being reflectively so attached? If it is (1) the claim that the 'I' in the 'I think' needs to refer to the same thinking subject, the same person or human being, then we have the typical Strawsonian picture. But could it be interpreted otherwise? Could it not be (2) the claim that in this context we are talking about the 'I think' as a type, and not about a token? But, then, does it make sense to talk about its numerical identity?

Let us examine the illustration Allison offers in support of his claims. If A, B and C are to be thought together in a single consciousness, then the I that thinks A must be identical with the I that thinks B, etc. Furthermore, consciousness of A and B and C as a single complex thought, as a unity, must make possible the consciousness of the identity of the subject in relation to these representations as a unity. How is this explanation supposed to be illuminating if what is doing the work from the start is the notion of a 'single consciousness'? If 'single consciousness' stands for what guarantees co-consciousness or togetherness from the start, then the necessary unity or identity of apperception is not proved at all through the requirements of a single complex thought. The argument is in fact very weak, for a single consciousness is what uniquely can guarantee unity of representations from the start, so the position itself presents matters, without implying consciousness of identity. That is, it guarantees unity even if I do not think 'I think A', 'I think B', etc. The thought 'I think' as a thought could not achieve it anyway. But, if it is the person, who


29. Cf. Peacocke (1983), p.108, about 'I' as a type, i.e. the type [self].
stands for the ‘I’, that explains the unity of a single complex thought, then it is not simply because of the use of ‘I’, or of a type, which is possible in connection with any representation, that we will have necessary numerical identity of consciousness. Because one is not aware of one’s numerical identity with the use of a same sounding ‘I’. More is required: the ‘I’ has to refer to the same subject. Allison seems to accept this.

But then Allison has to face a well known Kantian tenet, and, as a consequence, he qualifies his intentions: he is not trying to prove that we are aware a priori of our numerical identity, because this would certainly amount to having a Cartesian view of the self; rather we are aware of the ‘fact’ that this identity must be presupposed as a necessary condition of knowledge. How does Allison defend this claim, and does this now assert something different from what is contained in James’ point? We may expect an clarification coming from Allison’s account of the act of synthesis, but in his strategy the necessity of such a synthesis is itself derived from the necessary unity or identity of apperception. Nonetheless, he says that the act of synthesis is what makes that identity possible.

Much of the difficulty concerning the connection between the thoroughgoing identity of the apperception and the theory of synthesis relates to two Kantian theses which have one formulation in B133: (1) That identity ‘contains’ a synthesis and (2) that identity is possible only through the consciousness of the synthesis. Is it possible to learn something from Allison’s explanation of these theses?

The explanation of (1) reminds us, firstly, that apperception involves an actual consciousness of an identical ‘I think’. This means that the thought of the identity of the ‘I think’ ‘contains’ a synthesis, that is, through it a unification of representations is also brought about. Suppose we have ‘I think A’ and ‘I think B’ pertaining to a single consciousness. Allison’s point is that the subject of these thoughts can only become conscious of her identity in the thinking of these thoughts if it combines A and B. The claim then is that apperception of identity requires that a unity of representations has been brought about. This is why this act of becoming aware of identity ‘contains’ a synthesis.

The claim that awareness of identity involves awareness of synthesis or combination finds its defence as follows. The combination of A and B which makes the consciousness of identity possible is a ‘putting together’, that is, the awareness of identity is made possible through a combination or synthesis of A and B, but it is itself at the same time an awareness
of the result of their combination. So, one would have to think ‘I think A’, ‘I think B’, and then ‘I think A and B’.

But in this case, the objection is straightforward. The necessity of a conjunction of ‘I’-thoughts threatens an infinite regress. Why is it not required that one thinks those thoughts plus ‘I think A, B and A and B’?

It is at this juncture that Allison insists that ‘my consciousness (apperception) that both A and B are my representations is inseparable from my consciousness of the act of thinking them together in a single consciousness’ (claim 2 above)\(^{30}\). That is, the consciousness of the act of synthesizing is expected to do the trick, to avoid the regress. Now, Allison’s all important qualification of this claim is that it does not amount to introspective psychology nor to an idealistic ontological thesis about how the mind creates the phenomenal world. But do these negative points pre-empt the accusation that the position indulges in the myth of the giving?\(^{31}\)

Allison tries hard to connect the points he is making with a consciousness, not experience, of the form of thinking, that is, with the formal account of the objective unity (or analytic unity) of all general concepts. The decisive issue, then, concerns the claims that there is no intuition of spontaneity, that consciousness of spontaneity is intellectual, and that the consciousness of the act of thinking is a non-empirical consciousness\(^{32}\). The fundamental reasoning behind all these claims is that objectifying thinking, by making it a psychological object destroys its character as thinking. We need to accept that thinking would be already on the scene doing the objectifying if we tried this kind of approach. And it is precisely by making this type of point that Allison expects to convince. Allison seeks to clarify this consciousness of the activity of thinking, pertaining to Kant’s transcendental self, by comparing it with Wittgenstein’s metaphysical or philosophical self of the *Tractatus*. It must always be thought of as already on the scene doing the conceptualizing; since, as the eye of the visual field is not *in* the visual field, the subject of apperception is not an object in its conceptual field. Now, in Kant, according to Allison, this subject is responsible for the act of synthesis.


\(^{32}\) Cf. op.cit. pp.272-8.
But, with justification, philosophers cannot accept the idea of, as it were, an act without an agent or of acts which give themselves their content\textsuperscript{33}. Nor is it easy to accept the view, so do still other philosophers complain, that the ‘I’ when used by the subject does not refer to an empirical subject of experience.

From the discussion of Strawson it should be clear that I side with Allison in the effort to bring out the importance of Kant’s view that the consciousness of the act of synthesizing is not an experience, and that it is not an empirical consciousness of something which can have phenomenological properties. I also agree with Allison’s resistance to the objectifying of thinking. However, the appeal to the Tractarian metaphysical self as an attempt to solve the problems involving the notion of a consciousness of the act of synthesizing is not, in my view, convincing\textsuperscript{34}. I will attempt to produce an account of Kant’s founding role for the transcendental apperception which does not involve appeals either to psychological objects or states, or to the conception of a truly meta-physical self. I believe Kant follows a different path, which can be called ‘externalism’.

5. Henrich and the Deduction from the Identity of Self-Consciousness

Henrich proposed a program of interpretation of Kant’s Transcendental Deduction in terms of the identity of transcendental apperception in his paper entitled “The Identity of the Subject in the Transcendental Deduction”. What interests Henrich is the view that ‘numerical identity belongs to self-consciousness as such’\textsuperscript{35}. What can be the import of ‘as such’ here? I will discuss one possibility.

Numerical identity is normally contrasted with qualitative identity. Following Christian Wolff, by numerical identity is to be understood ‘the total identity of an individual thing’\textsuperscript{36}. Kant shares this view. At A263/B319, \textit{numerica identitas} is to be taken as the

\textsuperscript{33} Hurley’s kind of worry, cf. \textit{loc.cit.}.

\textsuperscript{34} Cf. Hacker (1986) chapters III and IV, pp.56-107 for the view of what is still psychologistic about Wittgenstein’s metaphysical self, \textit{viz.} connected with this self are ‘The Doctrine of the Linguistic Soul’ (‘it is the soul that is the fountainhead of language and representation’, p.75) and the thesis of ‘Transcendental Solipsism’ (‘the self whose representation the world is, is not among others’ p.102).


identity belonging to a thing when it is always one and the same, only one thing, not many. Kant gives us one particularly clear account of how we establish the numerical identity of an object. 'If we want to know through experience, the numerical identity of an external object, I shall pay heed to that permanent element in the appearance to which as subject everything else is related as determination, and note its identity throughout the time in which the determinations change' (A361-2).

If Kant really wants to attribute numerical identity to self-consciousness, then he has to face a problem on his own account. We would be talking of self-consciousness as a thing. This is a problem because from the restrictions of the Paralogisms it is clear that the identity of self-consciousness does not allow the inference to the identity of a thing. Henrich's solution to this problem involves qualifying the position on Kant's behalf with the mentioned 'as such'. According to him it is in this connection that Kant talks of the 'logical identity of the “I”', the 'I' as a thought only, etc. 'Logical', however, is to have a special meaning here.

Henrich holds a view that tries to accommodate numerical identity, with its strong implications about individuality, and the several claims about the reality of self-consciousness, to the 'I', insofar as it is a real thought; such as when, by the thought 'I think,' a reference (certainly unspecifiable) is made to something real.

The consciousness "I think" is undifferentiatedly "the same" in relation to all possible contents of thought. It can thus also be said of it that it is "permanent and abiding" in relation to these contents. And this can be interpreted harmlessly to mean that it is one and the same "I" that is certain of its numerical identity in relation to all these contents, when it forms the thought "I think" in relation to all these contents. This "I" is also a thought which is to be thought as the same in indefinitely many "I think"-acts. It is in being able to have this thought, however, that the entire essence of this self-consciousness consists.

One plausible understanding of this text is as follows. 'I' is a type, the same in relation to all possible thought-contents, as with Peacocke's type[self] mentioned a while ago. Putting aside the problem that we should be talking about types as things or objects, because of their numerical identity, this seems to satisfy some needs. The 'I' of apperception

sie, wie Aristoteles es erläutert, ein und dieselbe materielle Einheit sind, so haben wir es mit der Identität im engeren Sinn zu tun, die man als numerische Identität bezeichnet' (p.169).

would express the type, and would represent the form of consciousness. Furthermore, if 'same subject' means 'same type', then it is clear why this subject does not distinguish objects in the world, given that types do not do it, only tokens. This gives some content to the Kantian idea that 'I think' is a universal proposition valid for all thinking beings which does not distinguish objects in the world at all.

This view is not worthwhile pursuing here. Furthermore, as an attempt to capture Henrich's position it is off the mark, because he presents this view as the formalist view of the Neo-Kantians, whom he criticizes for turning this type (similar to the type [self]) into 'a merely formal property of thoughts which can itself never become an instance of consciousness'\(^{38}\). I think this complaint on Henrich's part has to do with the real act side of self-consciousness, which he wants to promote.

Henrich's position does not seem to be that in Kant we could be dealing with the identity of a type. He seems to point to a different approach: 'It is in being able to have this thought [of the numerically identical 'I'], however, that the entire essence of this self-consciousness [the same in relation to all content of thought] consists'\(^{39}\). But, then, Henrich's middle path is not clear to me: 'The identity implied in self-consciousness is neither formal-logical identity nor that which can be exhibited on the basis of criteria for the identity of a type of entities in the world'\(^{40}\). Henrich intends to show that Kant has a novel conception of identity, but I fail to understand what that might be.

### 6. Kant's Externalism

With this section I start my attempt at clarifying what Kant intended with the notion of a 'transcendental ground of the unity of consciousness'. This is Kant's notion for the ground of the necessity of the synthesis in the concept, as will become clear in the sequence. Kant states quite explicitly that this ground is the *transcendental apperception*. What has to be explained is how the necessity in concepts, that present in the way rules make us connect representations in determinate manners (which just is to subject representations to rules), is seen by Kant as grounded in transcendental apperception. So we have to deal with what

---


makes possible for rules to work as rules, *i.e.* with necessitating power, what it is that in this capacity is named by Kant in that way. I will propose a rational reconstruction of Kant's views on this matter, and this reconstruction will be called 'externalism'. The main tenet of externalism is that the transcendental ground Kant is investigating is the transcendental subject conceived of as a corporeal agent that exercises its ability to perform overt actions which it is conscious of itself as the subject of. So Kant's ground is the human being with the abilities to exercise techniques. In this respect I would like to remark again that my objection to Strawson is not that he brought in persons (corporeal beings) in the account of the conditions of experience, it concerns rather the route by which he brought them in. Subjects as corporeal persons will be shown to be important, but as subjects that exercise overt techniques, not as guarantors of privately owned particular experiences. My main goal in this section is to argue for the externalist tenet concerning the transcendental subject being a corporeal agent that develops and exercises techniques as overt actions. For this I will claim that what Kant intends with the 'productive synthesis of imagination' is to show that *overt* actions of what can only be conceived as a corporeal agent are the basis of his account of the possibility of knowledge. In the next section, consciousness of self will be accounted for in externalist terms and its position in Kant's argument will be determined.

The main argument of this section can be summarized as follows:

1. Human knowledge depends on conditions which involve a unified consciousness.

2. A unified consciousness presupposes *a priori* a subject that is conscious of itself as the subject which performs various overt actions.

From which it can be concluded:

3. Human knowledge presupposes *a priori* a subject that is conscious of itself as the subject which performs various overt actions.

Conclusion (3) can be taken as the externalist claim about the ultimate conditions of human knowledge, but of this claim the basis is the conception of a subject which it puts forward. Premise (2) is clearly the last argumentative layer on which, according to externalism, Kant must build his account. It is on the subject as qualified in (2) that we must look for the transcendental ground Kant says he wants to clarify. I will argue, therefore, for the externalist conception of the subject as it appears in (2) and (3).
I take premise (1) to be supported by all the analyses that I presented up to this point. Human knowledge should not be taken as a mental item of some sort. It is expressed in empirical judgements that state how the world or something is. Thus far it has conditions which, we saw, Kant designates as pure intuitions and categories. Now, we also saw that these conditions can only work as conditions insofar as they contain a unified consciousness or unity of consciousness which expresses a synthesis as necessary. It is thus that we get to the point of having to explain what accounts for the unity of consciousness of the unified consciousness. I would like to go over this again. Let me state what I take the unity of consciousness here to mean.

One way transcendental apperception comes into Kant’s argument in the Deduction is the following. After stating that all necessity without exception is grounded in a transcendental condition, Kant turns his attention to the question of what can be ‘the transcendental ground of the unity of consciousness in the synthesis of the manifold of all intuitions, and consequently also of the concepts of objects in general, and so of all objects of experience, a ground without which it would be impossible to think any object for our intuitions; for this object is no more than that something, the concept of which expresses such a necessity of synthesis’ (A106). So, the concept of an object expresses the necessity of a synthesis present in the concept itself (cf. also B137). This synthesis is said to have a unity of consciousness (‘the unity of consciousness in the synthesis’). This seems to be nothing other than a unity present in the concept. We should then take the unity of consciousness as making reference to the synthesis or unity in the concept (cf. B133-4 and note). If so, then the question about the transcendental ground of the unity of consciousness is the question about the ground of the necessity of the synthesis in the concept.

Now, this ground - the ‘original and transcendental condition’ - is, as I already pointed out, ‘no other than transcendental apperception’ (A106-7, cf. Deduction B, ##16-7). But Kant has much more to say about what this ground is beyond designating it thus. Far from being an inner sense or an empirical apperception, the transcendental apperception has necessarily to be represented as numerically identical. This means, as an initial approximation, that such an apperception cannot be so represented ‘through empirical data’ (cf.A107). What Kant wants to explain is the validity of this transcendental presupposition about transcendental apperception, and this amounts to explaining the claim that ‘there must be a condition which precedes all experience, and which makes experience itself possible’
Chapter 5

(A107). The condition Kant is talking about is none other than transcendental apperception: 'The numerical unity of this apperception [transcendental] is thus the a priori ground of all concepts' (A107).

Kant's major thesis appears at work in A108. Firstly, the unity of consciousness depends on the mind's capacity to become conscious of the identity of function whereby it combines the manifold as in one representation. This means, secondly, that the original and necessary consciousness of the identity of the self is present in the consciousness of a necessary unity of the synthesis of all appearances under concepts, or rules. So, rules would not be possible without the original and necessary consciousness of the identity of the self. This reveals Kant's reciprocal connection between the consciousness of self-identity and the consciousness of the necessity present in rules. We shall now concentrate on how the subject determines how concepts as rules have their necessity. This I called in the former chapter the ultimate basis for the whole phenomenon of concepts having content or judgements being objectively possible. What I propose to clarify is what Kant asserts at the end of A108: 'The mind could never think its identity in the manifoldness of its representations and indeed think this identity a priori if it did not have before its eyes the identity of its act, whereby it subordinates all synthesis as apprehension (which is empirical) to a transcendental unity, thereby rendering possible their interconnection according to a priori rules'. What is this act which subordinates what is given empirically to a transcendental unity, making then rules that interconnect representations possible? From an answer to this question we should also expect, according to Kant, what is required to 'determine more adequately our concept of an object in general' (A108). What Kant is dealing with here is the philosophical explanation of how we institute necessity and what is required for it.

Now, I propose to give an externalist account of the grounding part of Kant's argument, of the part which deals with what is the basis for the interconnection of representations according to rules. This is the account of the subject as it appeared in premise 2 above. I propose to show that this is Kant's fundamental conception of the subject by showing how it enters Kant's account of the transcendental synthesis of imagination (the figurative synthesis or synthesis speciosa, which plays such an important role in the Deduction B, cf.B151-2). Kant's view on how a productive imagination on the subject's part
is essential to our faculty of knowledge is fundamental and poses many problems, but an 
externalist interpretation promises to offer the correct understanding of this obscure claim41.

An externalist interpretation of Kant in general is an interpretation which gives pre­
eminence to action in Kant’s inquiries about the conditions of knowledge. It is an 
interpretation which gives the appeal to actions a distinctive import. The actions that will be 
crucial for Kant’s transcendental account are not the sort of inner acts appealed to by most 
of Kant’s interpreters. They are not going to be either ordinary mental acts or unconscious 
activities, which commonly have their explanation in a quasi-mechanical psychological or 
cognitivistic account, they will be shown not to be suitable for a mentalist interpretation42. 
According to Saugstad, the first proponent of externalism in relation to Kant, ‘Kant’s 
position is that human knowledge depends, ultimately, upon our ability to perform a fixed 
set of overt actions essentially involving the movement of the human body’43. My view of 
Kant is essentially Saugstad’s, but I render this statement about what externalism on Kant is 
thus (as I put it before): the transcendental ground of the necessity present in the synthesis in 
the concept is the transcendental subject considered as a corporeal agent that exercises its 
ability to perform overt actions which it is conscious of itself as the subject of. ‘Externalism’ 
is quite clearly chosen because of the emphasis on the overt character of the actions 
mentioned.

One of the most important texts for ascribing to Kant an externalist position, as I 
take it, is the paragraph that starts what is normally considered to be the second part of the 
Deduction B. i.e. #24. The role of this second part is to connect the understanding with the 
object of the senses, and for that Kant insists on the need of a figurative synthesis. However, 
the all-important notion in the argument is that of ‘motion, as an action of the subject’.

41. It is interesting to note that Henrich in his important paper (1989) does not make anything of the 
productive imagination in Kant.

42. Cf. Bennett (1966) chapter 8 for the standard mentalism that appeals to what unites mind-stages into minds or what is required for one to have a mental history. Cf. Kitcher (1990) for the quasi-
mechanical cognitivistic approach.

43. Saugstad (1992), p.382. I have referred to this seminal work on several occasions. I would like 
here to register my debt to Saugstad, not only for his view on the figurative synthesis, but also for 
the weight he places on specific Kantian texts. After covering much of the requisite ground also 
covered by Saugstad, I will bring the discussion to bear specifically on the topics of my concern. It 
is Saugstad who proposed the term ‘externalism’.
We cannot think a line without *drawing* it in thought, or a circle without *describing* it. We cannot represent the three dimensions of space save by *setting* three lines at right angles to one another from the same point. Even time itself we cannot represent, save in so far as we attend, in the *drawing* of a straight line (which has to serve as the outer figurative representation of time), merely to the action [*Handlung*] of the synthesis of the manifold whereby we successively determine inner sense, and in so doing attend to the succession of this determination in inner sense. Motion, as an action [*Handlung*] of the subject (not as a determination of an object), and therefore the synthesis of the manifold in space, first produces the concept of succession (B154-5)\(^44\).

Now, at least one of Kant's examples - the drawing of a line in *thought* - seems clearly to indicate a kind of mental act on the part of the subject. The last example - that of the representation of time - seems to point instead to an externalist position: time needs for its representation the *drawing* of a line as an *outer* figurative representation which involves motion\(^45\).

The mentalist reading, in terms of mental acts, faces, however, unsurmountable difficulties in its attempt to capture Kant's meaning. Take for example Kant's discussion of the synthetic unity of apperception as the supreme principle of all employment of the understanding in the Deduction B (§17). In this passage Kant uses again the example of the line, and claims that we have the need for an original synthetic unity of apperception which is - and this is the important point - 'completely independent of all conditions of sensible intuition' (B137). Kant, therefore, links the example of the line and the required *a priori* synthesis in the following remark: 'To know anything in space (for instance, a line), I must *draw* it, and thus synthetically bring into being a determinate combination of the given manifold, so that the unity of this act [*Handlung*] is at the same time the unity of consciousness (as in the concept of a line); and it is through this unity of consciousness that an object (a determinate space) is first known' (B137-8).

This case of the drawing of a line is a difficult text for a mentalist to interpret because it is clearly meant as an illustration of the requirement of a synthesis which is independent of what is given as an empirical element in experience. The point is that Kant is not willing to compromise the independence from sensibility, in fact, from anything not due

\(^{44}\) Kemp Smith renders *Handlung* 'act', I have preferred 'action'.

\(^{45}\) Cf. Kant's remark in "On Inner Sense": 'We can only represent time to ourselves in that we affect ourselves by describing space and grasping the manifold of its representations' (sheet I, lines 4-5).
to a pure understanding. Thus, Kant’s point does certainly not include a reference to empirically determined mental acts on the part of a subject, which in the end are nothing other than something that could be the object of some experience. Such a constraint of independence has prompted interpreters to take a wholly different path.

Kaulbach takes the ‘motion, as an action of the subject’ to be referring to the motion of a transcendental subject, more precisely, of his transcendental hand [sic]. The ‘description of a space’ (which is the type of motion contrasted with the motion as the determination of an object in B155n) would not be the ‘empirical work of the drawing hand: rather... it is as if a transcendental hand produces the transcendental motion of the describing, of the “drawing”, of the “drawing wrongly” [Verzeichnes]... The transcendental hand [has] the figuration [Figürlichkeit] of all figures... always prior to all description of single representations’\(^46\). Arguably, it is to such an account of what ‘motion, as an action of the subject’ can be, that Allison has to appeal. It is worthwhile remembering that for him the transcendental subject is the Wittgensteinian philosophical or metaphysical self that does not belong to the world. Supposedly, it is such a subject with his ‘transcendental hand’ that first gives meaning to notions such as line, circle, dimensions of space, and even time. In Allison’s case, this is the self with, neither a physical eye (he is not an object in the world), nor a geometrical eye, because it would rather have a super-physical or truly meta-physical eye, given that he is after all the knowing subject qua knower or the subject that thinks or entertains ideas\(^47\).

The externalist interpretation of ‘motion, as an action of the subject’ states, however, in Saugstad’s formulation, that this motion ‘refers to overt actions essentially involving the movement of our limbs’\(^48\). More specifically, Kant is, according to the externalist view, referring to, as an example, a bodily motion of the hand.

What we have to do now is justify this externalist interpretation. This will bring into the discussion the intelligible aspect of human action; the discussion will however be confined to Kant’s understanding of the role of the productive imagination. The manner in

\(^{46}\) Quoted by Saugstad (1992), p.386; the translation is mine: AK.


which motion as the synthesis of the manifold in space is crucial for Kant is our concern now.

The task of justifying the externalist interpretation can be done using three arguments, according to Saugstad\(^{49}\). These arguments concern the objectives and clarifications Kant himself offers in the note he provides on the phrase 'motion as an action of the subject (not as a determination of an object)'. The note is B155n:

Motion of an object in space does not belong to a pure science, and consequently not to geometry. For the fact that something is movable cannot be known \textit{a priori}, but only through experience. Motion, however, considered as the describing of a space, is a pure act [\textit{reiner Aktus}] of the successive synthesis of the manifold in outer intuition in general by means of the productive imagination, and belongs not only to geometry, but even to transcendental philosophy.

Following Saugstad, we should note, firstly, that Kant intends to stress a distinction here between the mere motion of an object in space and a motion which is a description of a space. If this distinction were not important for Kant he would not have bothered making it. Now, there is obviously one respect in which the motion of an object in space is directly comparable to a motion which describes a space. Suppose the description of a space by the drawing of a plane geometrical figure by the hand of any human being on a piece of paper. This drawing is judgeable empirically, and in this respect is at the same level as the motions of objects in general. The first argument is against mentalism, and its point is that if 'motion, as an action of the subject' were conceived in a mentalist way, as a mental act of some sort, the danger of confusion Kant is trying to avoid in the footnote would not materialize anyhow, given that nobody would easily confuse a mental act with the motion of a physical object in space. The externalist has to admit, however, that here Kant is only concerned with the pointing out of a distinction.

To reinforce the point made, Saugstad reminds us, secondly, that Kant would hardly have talked of mental acts as motions, given that he is strict about what motion can be, \textit{i.e.} alteration of place (cf.A32/B48). Furthermore, 'place' being a spatial notion, it is very difficult to see how the mental for Kant, which is linked to the inner, the non-spatial (cf.A33/B49-50), could be confused with motion. Again, apart from the intention to bring to our attention the distinction already mentioned, the footnote would lose its purpose.

\(^{49}\) I will here present the arguments for the externalist analysis as they appear in Saugstad (1992), pp.385-6.
Thirdly, whereas there is one respect in which the motion of a physical object in space and the bodily motion in the description of a space are at the same level, there is also a very important difference between the two. As explained, both can - as motions - be observed; but only the descriptive motion can be an action performed by a human agent. So, according to Saugstad, what must be meant by 'pure act' is precisely this active and human aspect, the doing of something, or the performance-aspect.

Furthermore, one should note that motion, as an action of the subject, is believed by Kant to be of fundamental importance and interest for philosophy itself. This is no small claim, and what Kant is trying to distinguish as important is motion taken in a certain respect, in spite of its remaining motion physically speaking. I think we can assume that the 'motion' side of the story is clear, it refers to, say, 'bodily motions involved in familiar, overt actions through which we produce physical figures'. The aspect of motion that is important, which distinguishes certain motions from the motion of physical things in general, may still be prone to misunderstandings. What we need is an account of the action component of Kant's approach.

The externalist's stress on the role of the body - because actions have to be motions, they have to be overt (possibly) - may seem to downgrade the transcendental sine qua non conditions Kant is investigating to mere physical happenings or events. But such a view can be rebutted by the proper understanding of action in this context.

Kant's appeal to a pure act is not the appeal to something outside the order of natural happenings, something truly meta-physical. It is rather an appeal to what externalism construes as the performance-aspect of a natural happening. An important claim connected with this is that this pure act is not what appears in the object or in the happening, that is, as that which is immediately available on a particular occasion as pertaining to the senses. Thus we find Kant claiming that 'whatever in an object of the senses is not itself appearance, I entitle intelligible.' (A538/B566). This is not meant as a claim to the effect that the intelligible is something behind or separate from what appears; it is meant rather as the view that from the empirical perspective physical movements are all at the same level, but from another perspective certain physical movements could not be more different from certain others. Certain physical movements are actions for Kant. In his comments on the

solution to the Third Antinomy Kant makes the general point that human movements, when considered as performances of human agents, are the result of an intelligible causality which has at its basis transcendental freedom. In this respect, human actions are the intelligible first causes of happenings in the world (cf. A532/B560-A537/B565). I call attention to this element of Kant's view of human action in order to point out solely that it coheres with what matters for Kant in regard to the motions of our concern: the performance-aspect. Kant sees actions in general as having reasons, as being grounded, as being connected to practical 'oughts', as being ascribed to the agent's 'intelligible character', all of which requires seeing physical happenings against a wholly different background, as embedded in certain specific circumstances: of people doing things for reasons. This is also in line, in a wider picture, with the way I presented Kant's Transcendental Idealism in the former chapter. So, to say of something that happens that it is an action is not to name a special being or entity or to designate a kind of meta-physical quality of what are otherwise ordinary happenings, it is rather to consider it as a performance of an agent. This is why Kant calls the describing of a space a pure act, and this is what can be termed the intelligible in what we see happening.

What Kant is driving at with the pure act, and its correlation with the intelligible, in his explanation of the describing of a space, can be gathered more perspicuously from the consideration of what works as a condition of representation in the case of the representation of something pertaining to outer sense. I will present, in this connection, one meaning Kant gives to 'empirical intuition'. And this meaning will be claimed to be very important. My intent is to get a firm grip on the action- or performance-side in relation to the indispensable synthesis of the manifold. This will amount to an explanation of how overt actions can relate to experience as its conditions, how practices as techniques lay at the basis of specific judgements of experience, how they condition objectivity.

51. For a worked-out interpretation of Kant's account of action along these lines cf. Allison (1990), specially part 1.

52. One should note that my main concern is not how Kant distinguishes motions that are actions from motions that are merely motions of an object in space, even though I said something to this effect in the text above. My point is that Kant is remarking that this distinction is made, from which he derives important consequences.

53. Saugstad follows Kant and gives the example of the familiar activity of measuring lengths as a synthesis (cf. 1992, p.388). Kant asserts that measurement is 'the successive synthesis of its [a whole's] parts' (B456n). I will use as illustration of the claim the measurement of weight instead.
Suppose I wanted to have a barbecue for some people and I had bought a big chunk of beef without knowing its weight. My first task would be to determine the weight of this piece of meat so as to decide how much to cook. To achieve this I could take a balance of some sort with some standard of weight, say, kilograms. In the case of the balance being one with two pans or scales and a central pivot, what I could do is to put the piece of beef on one scale and weights (of kilos and fractions of kilos probably) on the other so that the pointer on the pivot is in a perfectly perpendicular line with the plane the balance is standing on. I discover that I had to put two Kilos and five hundred grams on the weight scale to have the pointer at the right position. Probably someone much practised in the task of weighing could have taken my piece of meat in his hands and judged the weight by the 'feel' of it. What this presupposes however is that this person could do at least something similar to my weighing procedure, otherwise it does not make sense to take this person as estimating the weight of the beef.

What is noticeable in this example is that only material things were involved in the weighing all along (even the balance is perhaps dispensable, I could imagine some rudimentary weighing taking place through the holding in the hands of two weights to be compared): the piece of meat and the weights (kilos, etc). However, these material things do not enter in the whole procedure in the same way, they do not fulfil the same task. One is being weighed, which is a quite 'passive' participation, and the others are being used as standards of weight. One is being determined to have a certain weight, a result Kant would call a experience judgement, and the others are being the means through which that weight is determined. What should strike us as fundamental here is that one set of material things is in fact used as a means for a certain objective.

This example shows how an empirical determination of weight, an objective determination, is dependent on the procedures which involve the overt actions of comparing the object to be weighted with a material standard of weight.

The intelligible component in this example of weighing is the taking of a material object as a means of determining, through a comparison, the specific properties of other material things, the taking of a material thing as a standard through which to judge weights of objects according to rules.\(^{54}\) No object has this relation (of being a means of objective

\(^{54}\) Promises sometimes play this role. There was a time when currencies had to be backed by gold stocks.
determination according to rules) as an intrinsic property. Objects can serve this purpose only by our putting them to this use, by our submitting them to the procedures of weighing. Kant’s motion as a description of a space has exactly this nature. The important philosophical point is that in drawing a circle on a piece of paper I am describing a space which can be used to judge whether for example my cup has a circular mouth (cf. Kant’s observation that ‘the single figure which we draw is empirical [of a triangle on paper say], and yet it serves to express the concept [of a triangle] without impairing its universality’, quoted infra)\textsuperscript{55}.

The general moral to be drawn from this, according to externalism, is that, if we take ‘experience’ in Kant to stand for empirical judgements in general (cf. B147 and B166), like one about the weight of my piece of beef, we can see that experience is conditioned by a set of ‘acquired behavioural techniques’\textsuperscript{56}. This set of techniques works in relation to experience as a set of \textit{a priori} conditions, they are the basis for a rule-governed activity of, for instance, weighing. In general, they are the conditions of the objectivity of experience.

From this it can be concluded that motion, as an action of the subject, action which is a description of space, is what gives a basis for the idea of a synthesis of the manifold as a result of the productive imagination (synthesis which is required for the unified consciousness present in any concept), and is a notion which requires the conception of the subject of knowledge as a corporeal agent that exercises its ability to perform \textit{overt} actions (actions involving the motion of his body) which it is conscious of itself as the subject of. Therefore, I conclude that, because of the account of how the conception disclosed in premise (2) above enters Kant’s argument, I am vindicated in holding conclusion (3) of my argument true.

The externalist construal of Kant’s ‘motion as an action of the subject’, or of the motion as the describing of a space is well confirmed by how it allows us to interpret other claims by Kant. I propose to list some passages of Kant that get illuminated by externalism. However, as with all philosophical masterpieces, the first \textit{Critique} contains difficulties and recalcitrant passages to an externalist approach. I will also comment on some of these.

\footnote{55. Cf. Saugstad (1992), p.381.}

\footnote{56. Cf. \textit{op.cit.} p.389.}
Firstly, externalism gives a role to ‘sensible intuition’ as an intermediate between concepts and objects, as a particular which relates to the universality present in concepts and to the individuality that pertains to objects. This is exactly Kant’s point when he claims in the Aesthetic that the outer representation of time is itself an intuition, that is, a particular, not a general concept or a discursive representation. ‘From this fact also, that all the relations of time allow of being expressed in an outer intuition, it is evident that the representation is itself an intuition’ (A33/B50).

This role of the sensible intuition as an intermediate which, on the one hand, is related to concepts and, on the other hand, is directly connected to particular objects is very well explained by Kant in his account of proofs in Mathematics, i.e. through the construction of concepts (cf. A712-38/B740-66).

The single figure [the constructed triangle] which we draw is empirical, and yet it serves to express the concept, without impairing its universality. For in this empirical intuition we consider only the act whereby we construct the concept, and abstract from the many determinations (for instance, the magnitude of the sides and of the angles), which are quite indifferent, as not altering the concept ‘triangle’. This... mathematical knowledge [considers] the universal in the single instance, though still always a priori by means of reason. Accordingly, just as this single object is determined by certain universal conditions of construction, so the object of the concept, to which the single object corresponds merely as its schema, must likewise be thought as universally determined (A713-4/B741-2).

The last sentence of this quote points to how Kant sees the construction, in this case of a geometrical figure\(^57\), determining the object of the concept by being an intermediate schema of it. The interest of this lies in the determination the concept achieves through the use of the schema: the object of the concept is ‘universally determined’. This also gives a clue as to how Kant envisages the objective validity of mathematical judgements.

An a priori judgement has objective validity only in so far as its concept can be applied to objects given a posteriori (B147), i.e., in virtue of the behavioural techniques (synthesis) that make experience possible. “7+5=12” has objective validity because it is true by virtue of the technique which is also the condition for the possibility of a posteriori quantity-judgements. It is strictly universal by virtue of the universal applicability of its concepts to all sets of objects, and necessary since my action would not qualify as following the rules of addition unless I arrive at 12\(^58\).

---

57. Cf. Saugstad (1992), pp.390-3 for the externalist analysis of Kant’s arithmetical example ‘7+5=12’.

Secondly, the synthetic *a priori* in sense-perception also receives a plausible explanation from the externalist perspective. We are just too prone to think that the *meaning* of, say, ‘red’ comes from the red found in several objects, like for example a red curtain. We are just inclined to think that, say, a round plate in the world represents roundness *in itself*. But this is not so in Kant’s *transcendental* perspective.

Externalism does not preclude that how an object strikes one may be used in the determination of the properties of objects in general. But, even in this case, it is not the ‘striking’ which is put to a use, but rather the thing that strikes us as so-and-so. It is not the ‘mental red’, purportedly produced by seeing a red object, that can be used as a standard in a comparison which aims to determine which objects are red. What can be used in a comparison is rather a sample of red in a colour-chart, or some red object which is transformed into a sample by the claim that it displays the colour red\textsuperscript{59}. This reminder allows one to make a point: it is not things and their properties as such which determine the meaning of the expressions linked with them\textsuperscript{60}. In Kant it is always required that prior to the statement that an object is thus-and-so (an empirical judgement expressing knowledge) we have syntheses (in intuition and concept) that work as conditions of knowledge. I believe that externalism has shown how ‘material things’ work as ‘sensible intuitions’ which we manipulate in techniques we use to determine the properties of objects\textsuperscript{61}.

\textsuperscript{59} This is the point I take Kant to be making at B133-4n, when he speaks of red in general. There the connection of the property red with other properties is presented as a requirement. He says that for red to be a *conceptus communis*, to be an analytic unity, which is a property of all general concepts, it has to be beforehand a synthetic unity. I take this to mean: red has to be determined as pertaining to objects to which other properties pertain as well, but this in a standardized manner.

\textsuperscript{60} I assume Saugstad intends to make this kind of point in (1992) p.393. I think, however, that his discussion is not very clear. Cf. Hacker (1992a) and (1992b) for the intelligibility conditions concerning ostensive definitions. The point I was trying to bring home is that ‘the grammar of the definendum does not flow from the object pointed at’ (Hacker 1992b, p.370). Cf. also Hacker (1993), pp.37-57. In Kant, for someone to have a ‘determinate thought’ it is required combination, but ‘combination does not, however, lie in the objects, and cannot be borrowed from them, and so, through perception, first taken up into the understanding. On the contrary, it is an affair of the understanding alone, which itself is nothing but the faculty of combining *a priori*, and of bringing the manifold of given representations under the unity of apperception’ (B134-5).

\textsuperscript{61} Compare this restricted account of the synthetic *a priori* in sense-perception and the correlated account of ‘sensible intuition’ in the activities of measurement to the general account of what aprioricity means in the previous chapter. Cf. also A157-8/B196-7 where non-empirical synthesis is said to be *a priori* knowledge only because ‘it contains nothing save what it necessary to synthetic unity of experience in general’.
Thirdly, I would like just to indicate what direction externalism offers when one has to face Kant’s claims about the need of the participation of sensibility in experience. Related to this are Kant’s views about the ideality of space and time; but this will not be pursued here. The participation of sensibility which is required should be understood as the realization on Kant’s part that not all overt actions are synthesis engendered by the intellectual capacity of the human being. Some overt actions involve the human body in a different way (cf.B519). Kant’s theory of space, for example, depends upon the actions of pointing to the right and to the left, which are not syntheses, but rely rather on a relation to our body, to our right and left hands. There is no way of denying that externalism’s insistence on the overt actions of - and in this case on the relations to things in space of - a corporeal subject are in line with the most important claims of Kant’s transcendental idealism.

Fourthly, externalism is also able to construe Kant’s most important epistemological tenet in an interesting way, which is not rationalistic, nor empiricist, and is thus true to the critical spirit. In Kant, concepts and intuitions require each other. The famous Kantian dictum expresses precisely this correlation: ‘Thoughts without content are empty, intuitions without concepts are blind’ (A51/B75). What Kant means by that is clarified in what follows: ‘It is, therefore, just as necessary to make our concepts sensible, that is, to add the object to them in intuition, as to make our intuitions intelligible, that is, to bring them under concepts. These two powers or capacities cannot exchange their functions. The

62. Cf. What Real Progress, p.Ak267: ‘The form of objects, which is all that can be presented in an intuition a priori, is not based on the constitution of this object in itself but on the natural constitution of a subject capable of representing objects in intuition. And it is only this subjective [element] in the formal constitution of the senses... which makes possible a priori intuition, i.e., intuition preceding all perception. This, and the possibility of a priori synthetic judgements, is easily conceived from the point of view of intuition’ (my italics: AK). For this reason we can ‘determine our intuitions a priori with the consciousness of the necessity of the judgements that determine them... To determine means, however, to judge synthetically’ (op.cit. p.Ak268; my italics: AK).

63. Cf. Saugstad (1992), p.395n26 for these points. There is, however, a long way to go if one wants to explain exactly how and why Kant sees the phenomenon of incongruent counterparts for example, which is specially related to our orientation in space through our left and right sides, as an indication of the ideality of space, especially given his changes of opinion about the conclusions to be drawn from the phenomenon over the years. Allison is very sceptical about Kant’s success (cf.1983, pp.99-102). Buchdahl sees some hope coming exactly from the notion of pointing (cf.1969, pp.598-60). This is precisely what I want to bring out: pointing as an overt action.
understanding can intuit nothing, the senses can think nothing. Only through their union can knowledge arise. But that is no reason for confounding the contributions of either with that of the other’ (ibidem).

The requirement to make our intuitions intelligible is one that has been interpreted in various forms. However, Kant asserts that making our intuitions intelligible is as necessary as making our concepts sensible. The latter task is achieved by adding to them the object \textit{in intuition}.

Let me now spell out the externalist view of Kant’s \textit{critical} meaning-principle. The intuitions which have to be made intelligible are not the objects in the world as far as they are objects with intrinsic properties (what would it be to make objects \textit{intelligible}? nor are these intuitions the mental representation of objects (how could this help in making our concepts \textit{sensible}?). According to externalism the ‘in intuition’ demand refers to the critical model of subsumption used by Kant: ‘The intuition is neither the object, nor the awareness of the object, which is subsumed under the concept, but the external representation through which objects are subsumed under the concept’. The point here is that concepts for Kant, at least in theoretical use, must be made sensible, and this is to provide intuitions for them. Now, according to externalism, the proper way of doing this is by making an intuition intelligible. As intuition here does not mean an object insofar as it possesses intrinsic properties, an ‘intelligible’ intuition must be an object which, because it is external to the subject, functions as an \textit{external representation} that is used as means to judge whether other objects have the property exemplified in the object in a \textit{standard} manner.

---

64 Cf. Allison: ‘Kant does indeed regard a distinct concept as the product of a synthetic activity. As a result of this activity it stands in relation to intuition, and thus to an object’ (cf. \textit{On a Discovery}, Introduction, pp.62-3).

65 Cf. \textit{What Real Progress}, p.Ak273 for how reason extends to objects of the senses so that judgements concerning objects in general be cognitions. Cf. also Kant’s conception of \textit{formal intuition} as the representation of space as object which provides us with a unity of representation (B160-1na). Cf. still B161 for the most abstract formulation of what the unity of representation is: a given intuition in general in relation to the original consciousness according to categories.

66 Cf. Saugstad (1992), p.397n29 for a comparison of Kant’s claim about intuition in this context and Wittgenstein’s notion of paradigms. What should be clear from what has been said - in confirmation of views first presented in chapter 3 - is that Kant does not hold Strawson’s ‘principle of significance’. According to the later, Kant’s sensible intuitions would be particular objects to which concepts apply, but then they are also the awarenesses of particular objects. Walker, in (1989) p.43, exemplifies this empiricist tenet: ‘In its primary sense intuition is the immediate
As the last corroboration of externalism in relation to Kant I would like to mention Saugstad’s analysis of what is a schema in Kant’s transcendental philosophy. In the Schematism chapter, schemata (as the drawn circle on a piece of paper) is clearly a sensible intuition which has a mediating function in relation to, for example, the geometrical concept of circle and my cup’s mouth. Only externalism can give a plausible account of Kant’s demand for *schema* thus conceived.

I would like to close this section with three difficult passages for the externalist reading of Kant.

Firstly, even though in chapter 4 I claimed that externalism was well positioned to explain Kant’s notion of an *acquisition originaria* of the categories and the concepts of space and time, Kant in A86/B118 says that impressions are the first stimulus for experience being brought into existence. This suggests that the categories are given somehow prior to experience, awaiting then for the first stimulus of impressions.

Secondly, on synthesis Kant is in fact equivocal. In A99 synthesis seems to be presented in a way which is clearly inimical to externalism. Synthesis is there presented as a type of production of a representational item. Kant says that ‘each representation, in so far as it is contained in a single moment, can never be anything but absolute unity’. The mind has then to synthesize representations, and this is ‘to run through’ the manifold and ‘hold [it] together’.

 awareness of particulars; by transference the word becomes applied to the particulars we are immediately aware of’.

67 . Saugstad (1992), p.397. Cf. also Saugstad’s externalist analysis of Kant’s demand in the third *Critique* (pp.Ak342-3) of the ostensive display of the concepts of the understanding in analogy with the anatomist’s ostensive display of the parts of the human eye. The result is that, in parity with the anatomist’s ‘demonstration’ of the eye, sensible intuitions in Kant cannot be private representations of objects, they are rather ‘representations which can be shown to others’ (cf.p.394). This gives at least a plausible sense to the demand of making a concept *sensible*, at the same time in which it makes comprehensible the requirement to make a intuition *intelligible*.

68 . Cf. Saugstad’s analysis (1992, pp.397-8) of the plate example: plates are homogeneous with the geometrical concept of a circle. (What is remarkable about this analysis is that it restores a Kantian text purportedly adjusted by the great Kant-scholar Hans Vaihinger. Kant in fact wrote the contrary of what Kemp Smith following Vaihinger has as B176. It should be: ‘The roundness which is thought in the former [the empirical concept of a plate] can be intuited in the latter [the geometrical concept of a circle]’. Cf. the Felix Meiner *Kritik*, p.197).
Thirdly, as we will see in the next section, externalism by being keen on overt actions as techniques can include the speaking of a language among what is relevant for knowledge to be possible. In this respect it is not any passage in the Critique which is troublesome to externalism, but rather the lack of clear remarks on Kant's part on the role of language.

About these three recalcitrant reminders to externalism, I would like only to observe that (1) the first can, I believe, be accommodated: occasion of use is certainly not the same as a cause which puts to work a pre-existing mechanism of some sort; (2) the passage at A99 pertains to the first edition views on the 'synthesis of apprehension', which have to be taken with a pinch of salt anyway\(^6^9\); and (3) there are some passages where Kant speaks of language and actually insists on its fundamental role\(^7^0\).

7. **Self-Consciousness and Kant's Externalism**

In this section I intend to clarify the connection, which seems crucial to Kant's argument of the Deduction, between self-consciousness as a *principle* and the subject that exercises techniques which are the conditions of the possibility of objective knowledge. Kant says of the principle about self-consciousness: 'the principle of apperception is the highest principle in the whole sphere of human knowledge' (B135). There can be no doubt that this principle is, in Kant's own eyes, the fundamental principle to 'all employment of the understanding' (B134n, quoted *infra*). My goal in trying to clarify the connection mentioned is to get a better picture of what Kant understands is self-consciousness or transcendental apperception. We have already seen that for Kant self-consciousness is linked to the use of 'I', we also saw that for him 'I' is used in such a way that to this human faculty we can attribute self-sufficiency. I will in this section argue for the externalist view of what self-consciousness is for Kant beyond what we got so far: the *actus* 'I think' and all the remarks

\(^6^9\) But see #16 of the Deduction B (B131-2), where Kant talks of representing as something that happens 'in me', and to this he seems to connect the notion of one same subject to which representations belong: a mental unifying thing. There is no way of denying Kant's Cartesian heritage, which hinders his own exposition of his novel ideas.

\(^7^0\) Cf. Saugstad *op.cit.* p.391 and note 17. Kant says in *On a Discovery* (p.112 or p.Ak193-4n) that we use words even in judgements we do not yet express in sentences.
on the 'I'. The leading idea will be Kant's own statement that self-consciousness is *the ability to say I to ourselves*\(^{71}\).

The present investigation is informed by the externalist conviction that to understand Kant's claim about 'the highest point' it is necessary to understand previously Kant's account of how a subject that is corporeal develops techniques for the judgement of objects, account which is used to explain the necessary figurative synthesis of imagination. So, a great importance will again be given to the externalist conception of the transcendental subject. But, then, we also find Kant saying that 'the transcendental unity of apperception thus relates to the pure synthesis of imagination, as an *a priori* condition of all combination of the manifold in one knowledge' (A118). And in the same vein: 'The unity of apperception in relation to the synthesis of imagination is the understanding' (A119). If we bear in mind that the purpose of the Deduction is to clarify how the understanding stands in a necessary connection with appearances (cf.A119 and chapter 4), the characterization of the understanding just mentioned seems fundamental. In the type of passage just quoted Kant seems to be emphasizing the other direction of the reciprocal connection between the consciousness of self-identity and the consciousness of the necessity present in rules. In the previous section I concentrated on how the subject plays a fundamental role in the account of the necessity present in rules which we follow in judging how objects are and behave. In this section, I shall concentrate on that aspect of the subject which allows him to do what he does as a technique. We are concerned now not with the doing of something, but with the consciousness of self which allows the subject the awareness that he is doing something, doing this or that, doing the same. In the previous section the identity of an act by the subject was investigated. Now, it will be the subjection to a transcendental unity which renders interconnections according to rules possible, that is, the following of rules possible (cf.A108).

One should not forget that the 'highest point' is a *principle*. To account for this qualification I will make use of some metaphilosophical reflections of Kant. And having once entered this area of Kant's thinking, I will argue for the importance of Kant's *philosophical* clarification that self-consciousness is a fact. Thus far a feature of my account comes to the fore: in noting that not only the possession of categories is a fact (which means

---

that we judge objects according to the \textit{a priori} concepts categories were shown to be), but also self-consciousness is a fact, I intend to point out a characteristic of human beings according to Kant which is not philosophically reducible to any other simpler feature. We find in this observation Kant’s offering of an ultimate datum to our reflection, and we are not able to get behind it philosophically. We can only point out the features of the activity of judging by us human beings. Accordingly, I do reject the polemic of Kant’s interpreters about what is ‘the ultimate condition’ in Kant: self-consciousness or the conformity of appearances to the understanding. For the one camp self-consciousness would be the source of the categories, for the other self-consciousness would only be possible because we judge according to categories.

How can the externalist view of the transcendental subject as an embodied subject that uses external standards according to techniques of behaviour to judge the properties of objects help to clarify what self-consciousness is and why it is important for such a conception of that subject? Let me quote a text which is important for externalism in full.

The analytic unity of consciousness belongs to all general concepts, as such. If, for instance, I think red in general, I thereby represent to myself a property which (as a characteristic) can be found in something, or can be combined with other representations; that is, only by means of a presupposed possible synthetic unity can I represent to myself the analytic unity. A representation which is to be thought as common to \textit{different} representations is regarded as belonging to such as have, in addition to it, also something \textit{different}. Consequently it must previously be thought in synthetic unity with other (though, it may be, only possible) representations, before I can think in it the analytic unity of consciousness, which makes it a \textit{conceptus communis}. The synthetic unity of apperception is therefore that highest point, to which we must ascribe all employment of the understanding, even the whole of logic, and conformably therewith, transcendental philosophy. Indeed this faculty of apperception is the understanding itself (B133-4na).

Here we find Kant making some points which by now should not surprise us. For us to have the \textit{concept} of red this representation has to be thought in unity with other representations, that is, red as a concept, as a term with \textit{meaning}, has to be connected synthetically with other representations - such as of a \textit{thing} which apart from being red has other properties, the properties of a material sample for example. Only by this distinguishing of red as a colour from the other properties of a thing - as in the case of a sample its material or shape - can red become a concept applicable to objects in general.

Kant in this passage also makes clear that the object of his analysis is the faculty of understanding itself. Another important point is that his research leads him to ‘a point’
which conditions the whole of the application of the understanding, and this includes -
which is a fundamental clarification - the workings of logic and the task in hand itself, that
is, philosophy's endeavour to explain what conditions knowledge.

The problem is, therefore, to account for this 'highest point', the synthetic unity of
apperception. 'The principle of apperception is the highest principle in the whole sphere of
human knowledge' (B135).

Given that Kant asserts that the 'point' to be explained is a principle (which surely
distances his investigation from a psychological one of this faculty), it may be useful to
examine some of Kant's metaphilosophical reflections. Let us see what can be discovered
about self-consciousness from Kant's point about the investigation being of a principle.

Philosophical knowledge, Kant claims, is 'knowledge gained by reason from
concepts' (A713/B741), as opposed to mathematical knowledge which is gained from the
construction of concepts. 'Philosophy confines itself to universal concepts' (A715/B743), to
the 'discursive employment of reason in accordance with concepts' (A719/B747). These
claims point to Kant's view that 'there is a transcendental synthesis [framed] from concepts
alone' (ibidem), that is, the view that in philosophy we are also dealing with synthesis which
can be stated in 'transcendental synthetic propositions' (A721/B749). Thus, we find Kant
reflecting on the situation of philosophy in the following manner:

All our knowledge relates, finally, to possible intuitions, for it is through them alone
that an object is given [cf.B148-9]. Now an a priori concept, that is, a concept which
is not empirical, either already includes in itself a pure intuition (and if so, it can be
constructed), or it includes nothing but the synthesis of possible intuitions which are
not given a priori. In this latter case we can indeed make use of it in forming
synthetic a priori judgements, but only discursively in accordance with concepts,
ever intuitively through the construction of the concept (A719-20/B747-8).

Accordingly, the transcendental synthetic propositions 'contain nothing but the rule
according to which we are to seek empirically for a certain synthetic unity of that which is
incapable of intuitive representation a priori (that is, of perceptions). But these synthetic
principles cannot exhibit a priori any one of their concepts in a specific instance; they can
only do this a posteriori, by means of experience, which itself is possible only in conformity
with these principles' (A720-1/B748-9; cf. Kant's two examples in B162-3: the perceiving
of a house and the perceiving of the freezing of the water).

Philosophical proofs consequently deal only with discursive principles which always
require deductions (cf.A733/B761). They are, therefore, acroamatic, that is, discursive
proofs themselves, 'since they may be conducted by the agency of words alone (the object in
thought), rather than demonstrations which, as the term itself indicates, proceed in and
through the intuition of the object' (A735/B763, my italics on 'words': AK).

From this contrast with the demonstrative proofs of mathematics a further
distinction also comes to the fore. Philosophical definitions are nothing other than mere
exposition of given concepts, whereas mathematical definitions are constructions of
concepts. So, the former are obtainable only by analysis and, Kant insists, the completeness
of this procedure is never apodeictically certain; the latter on the contrary operate with
absolute certainty synthetically (this is a reference to the progressive synthetic method). The
reason for this is that 'mathematical definitions make their concepts, [whereas] in
philosophical definitions concepts are only explained' (A730/B758).

Kant's methodological prescription follows suit:

That in philosophy we must not imitate mathematics by beginning with definitions,
unless it be by way simply of experiment. For since the definitions are analyses of
given concepts, they presuppose the prior presence of the concepts, although in a
confused state; and the incomplete exposition must precede the complete
(A730/B758).

Now, the whole interest of philosophy is 'to investigate the origin of the pure
concepts of understanding and in so doing to determine the extent of their validity', and this
is something that mathematicians do not do, 'they care only to make use of them'
(A725/B753). This is, however, no great fault on the mathematician's part, for 'in all this
they are entirely in the right, provided only they do not overstep the proper limits... But,
unconsciously, they pass from the field of sensibility to the precarious ground of pure and
even transcendental concepts, a ground (instabilis tellus, innabilis unda) that permits them
neither to stand nor to swim, and where their hasty tracks are soon obliterated' (A725-
6/B753-4). This is why, for Kant, doing philosophy helps: 'The possibility of mathematics
must itself be demonstrated in transcendental philosophy' (A733/B761). And what kind of
demonstration this is, Kant has already made clear.

The investigation of origin and validity is conceived by Kant as a critical enterprise,
which aims at the ultimate goal of reason's self-knowledge. Thus, its prominent feature is
self-reflexiveness. 'Indeed it is precisely in knowing its limits that philosophy consists; and
even the mathematician, unless his talent is of such a specialised character that it naturally
confines itself to its proper field, cannot afford to ignore the warnings of philosophy, or to
behave as if he were superior to them' (A727/B755). Philosophy, consequently, has as its true purpose 'to expose the illusions of a reason that forgets its limits, and by sufficiently clarifying our concepts to recall it from its presumptuous speculative pursuits to modest but thorough self-knowledge. Reason must not, therefore, in its transcendental endeavour, hasten forward with sanguine expectations, as though the path which it has traversed led directly to the goal, and as though the accepted premises could be so securely relied upon that there can be no need of constantly returning to them and of considering whether we may not perhaps, in the course of the inferences, discover defects which have been overlooked in the principles' (A735/B763).

This is why for Kant philosophy cannot be dogmatic, in the sense of achieving its proofs by appeal to a method which derives synthetic propositions directly from concepts. Philosophy, consequently, does not issue doctrines (cf.A736/B764). Philosophy is bound to look at possible experience, something which Kant asserts is - as the third thing connecting the terms in the synthetic judgements of philosophy - 'altogether contingent' (cf. A737/B765).

Thus no one can acquire insight into the proposition that everything which happens has its cause, merely from the concepts involved. It is not, therefore, a dogma, although from another point of view, namely, from that of the sole field of its possible employment, that is, experience, it can be proved with complete apodeictic certainty. But though it needs proof, it should be entitled a principle, not a theorem, because it has the peculiar character that it makes possible the very experience which is its own ground of proof, and that in this experience it must always itself be presupposed (A737/B765).

The purpose of my excursion to these metaphilosophical reflections is to get an understanding of the claim that the philosophical analysis Kant offers for the highest point to which all knowledge is submitted has the status of an account of a principle. This is the principle of apperception. And the kind of clarification we should hope for is one for which metaphysics is equipped. Now, because the clarification sought is of a principle that is transcendental and as such regulates the employment of the understanding itself, we are here

72. Cf. What Real Progress, p.Ak293 and ff., for the view that, in opposition to a conception of philosophy as a theoretically dogmatic enterprise, critical philosophy is practical dogmatic.

73. Cf. Humphrey (What Real Progress, 'Translator's Introduction', p.19): 'Metaphysics is a possible science because man's faculty for conceptualizing can function independently of his capacity for intuiting'. How far 'science' here should be from 'natural' and 'formal' sciences should be clear by now.
necessarily looking for the presentation of something - of the content of the principle - *in abstracto*. Philosophical knowledge does always consider the universal *in abstracto* by means of concepts only (cf. A734/B762). However, I believe that it is necessary for the comprehension of Kant’s point the previous comprehension of his view on how a subject that is corporeal develops techniques for the judgement of objects, as he presents it in the account of the figurative synthesis. Presently we have to deal, then, with the consideration of what is revealed *in abstracto* about human abilities (which are the doing of things, even though not always exercised).

The principle that exercises Kant is then the *transcendental* principle concerning what grounds the unity of representations. This states the identity of the self of which we are *a priori* conscious as the necessary condition of the possibility of all representations (cf. A116). This principle says ‘no more than that all my representations in any given intuition must be subject to that condition under which alone I can ascribe them to the identical self as my representations, and so can comprehend them as synthetically combined in one apperception through the general expression “I think”’ (B138). What this principle claims as being a fundamental condition of representing is the identity of the self and that we should be *a priori* conscious of this identity. Here we meet one of the deeply counterintuitive features of Kant’s claims. We need the identity of the self, but we must be sure of it *a priori*. This suggests the demand for the identity of a thing, of which we should nonetheless be certain *a priori*. No thing has its identity guaranteed this way, goes the retort. However, what should govern our approach to Kant are his own claims, and about his principle he says that it is an identical, that is, analytical, proposition, which ‘reveals the necessity of a synthesis [*sine qua non*] the thoroughgoing identity of self-consciousness cannot be thought’ (B135). The proposition in spite of making ‘synthetic unity a condition of all thought,... [is] itself analytic’ (B138). Now, Kant’s intent here is that although he is making a claim that concerns synthetcity it is itself an analytical claim (cf. B407 and ff.). This means that the identity attributed to the self as a condition of possibility is *formal*: ‘The identity of the consciousness of myself at different times is therefore only a formal condition of my thoughts and their coherence’ (A363). This is why Peter Hacker says that the apparent uniqueness of the self (and this includes its assured identity) is *formal*; and he compares it with what Kant says about ‘consciousness in general’. He remarks as well that this privileged position belongs to the ‘I’ of the geometrical eye, not to the physical or even
meta-physical eye\textsuperscript{74}. But this \textit{formal} feature has to be clarified further. According to externalism, the \textit{a priori} identity of the self, or of self-consciousness, makes reference to the ability to use 'I' possessed by a human being, and it is this ability that is part and parcel of the ability to follow rules in making judgements that are objectively valid (possibly true or false). Let us now see how this is so in Kant.

The point about the identity of the self is no small claim by Kant. Kant is trying to give an account of the necessity that belongs to the very conception of the categories (cf.B168). This is an account of how there is a \textit{necessary} agreement between the understanding and the 'appearances' (cf.A119), that is, of experience with the concepts of its objects (cf.B168). Kant's most important result is that this agreement depends on the determination of 'appearances' in space and time in general. And the ultimate basis for this determination is the original synthetic unity of apperception (cf. B168-9), which is supposed to operate as the form of the understanding in its relation to space and time, \textit{i.e.} the forms of the sensibility. Kant's point is that this unity is the ground for the determination of 'appearances' because it is the 'unity of consciousness that alone constitutes the relation of representations to an object, and therefore their objective validity and the fact that they are modes of knowledge; and upon it therefore rests the very possibility of the understanding' (B137, cf. also A127). This is why we can expect as a result of this faculty's operation the production of judgements: 'I find that a judgement is nothing but the manner in which given modes of knowledge are brought to the objective unity of apperception. This is what is intended by the copula "is"' (B141-2). The judgement's use of this copula is just a sign of its intrinsic objectivity\textsuperscript{75}, that is, a claim to be true or false. The fundamental feature of the identity of the self in relation to the production of judgements is what we have to contend with, and this is its \textit{formal} character. As we saw, this seems to point to a kind of logical self, to the conception of the self as a logical subject, which is outside the empirical world. Externalism hopes to explain this feature of the self or of self-consciousness Kant stresses without appeal to such meta-physical notions.


\textsuperscript{75} As Allison points out: 'It [a judgement] is an "objective unity", and as such, it is correlated with the objective unity of apperception. ...Objective validity is, therefore, a definitional feature for Kant, not merely a value that can be assigned to some judgements' (1983, p.72).
In the *Anthropology* Kant asserts that man is an earthly being endowed with reason. In the same work Kant makes explicit what he is implying by this: 'I as a thinking being am one and the same subject with myself as a being in the world of sense'\(^{76}\). If the identity asserted here holds, then we can easily understand why Kant affirms in the ‘Loose Leaf Leningrad 1’ that ‘a cosmological Existence [which pertains to me as a Cosmological Apperception] is only the existence of a thing in appearance’\(^{77}\). So the ‘corporeal’ side of Kant’s story cannot be denied. The *Anthropology* is also very keen on affirming the ‘intelligible aspect’ though. The performance side, the action side, gets firmly stressed. Kant says that his interest in man, the result of which is the *Anthropology*, is pragmatic. He wants to investigate ‘what *man* as a free agent makes, or can and should make of himself’. He wants to see how we know our ‘way about in the world’, insofar as ‘we have participated in it’\(^{78}\).

Kant links this aspect of man, which he wants to study, with a property we have already encountered when considering meaning. We possess according to Kant an autonomy of *theoretical* reason. He asserts that there is ‘a freedom of the power of judgement and the autonomy of the mind, *by which the state of its ideas is under its control*’\(^{79}\).

Given the ‘corporeal’ side of Kant’s views, it may still seem unclear how exactly the ‘performance’ or ‘action’ side is supposed to have the fundamental role Kant envisages for it. This side is given its rationale by the pure self-consciousness of the subject. For externalism, however, this faculty should be understood as the ‘ability [of the corporeal transcendental agent] to express what he [the subject as agent] does in words’\(^{80}\). Hard evidence for this reading of Kant can be found in the *Anthropology*: ‘Pure apperception [is the] consciousness of what we are doing; for this belongs to the power of thinking’\(^{81}\). Now, crucial for this consciousness, for the expression of what one does, is the mastery of the use of ‘I’.


\(^{77}\) ‘On Inner Sense’, p.I lines 31-2, my italics: AK.

\(^{78}\) Cf. *op.cit.* pp.Ak119 and 120, respectively.

\(^{79}\) *Anthropology*, p.Ak131.

\(^{80}\) Saugstad (1992), p.382.

I would like to explain now why we should conceive the consciousness which indicates the power of thinking as the ability to use the word *I*. Firstly, I would like to point out Kant's awareness of the importance of words. In accounting for the unity of apperception in its relation to the possibility of knowledge we are in truth, according to Kant, dealing with a principle of which the philosophical explanation must be 'conducted by the agency of words alone (the object in thought)' (A735/B763, quoted supra). From the fact that we are dealing with a principle which is also said to condition the very investigation in hand, we should expect that principle to condition the use of words as well. Now, in speaking of the capacity of consciousness in general we are certainly speaking of what can be characterized as a capacity to deal with things *in thinking*. But for this we need to have *the object in thought*, which is done through 'the agency of words alone'. So, as consciousness of self is also a thought, it does also require an *object in thought*, which would be provided by 'I'.

Secondly, in *What Real Progress* Kant equates quite clearly the consciousness of self with the ability to say I to ourselves. This is again asserted to be the basis for the possibility of understanding, and what separates us from animals\(^82\). In the *Anthropology* Kant is also explicit. When speaking of self-consciousness, Kant affirms that the idea I raises man above all other beings living on earth. But, then, he says that when a child 'starts to speak in terms of "I"', and this is also when 'a light seems to dawn on him', then the child *thinks himself*, before that it merely *felt himself*. So, even if a language (a natural one) does not have a special word to express self-consciousness, it certainly *thinks* I, because it can express this idea with other means, as when, for example, in Latin the declination of the verb indicates its first person use, as with *cogito*\(^83\).

Lastly, in the first *Critique* there are many passages connecting self-consciousness, transcendental apperception, transcendental consciousness to 'I' as a representation (cf. A381-2, B278); sometimes saying that they are the same (cf. A117n, and B132, where the representation 'I think' is called *pure apperception*).

Now, in *What Real Progress* (loc.cit.) Kant conveys the idea that the consciousness of self is a fact which as a capacity allows me to distinguish 'me from myself', but of how

\(^82\) Cf. *op.cit.*, p.Ak270.

\(^83\) Cf. *op.cit.*, p.Ak127.
this is possible Kant says that any explanation is 'absolutely impossible'. Kant, therefore, highlights the presence of this ability in several observations, always with an eye at the right conception of the transcendental subject though. In the Anthropology he writes: 'Man’s ‘I’ is indeed twofold in terms of form (manner or representation), but not in terms of matter (content)'. This is a reference to the distinction between being conscious of myself and knowing myself, the latter being attainable through the recognition of psychological traits and dispositions. This form of dealing with myself also involves knowing things like my height, weight, my type of blood, etc. Kant's claim is that I in fact use 'I' in two kinds of situations: such as when I express my thoughts; and as when I say in which state or condition I am in, or what features I have. But this still does not make a distinction in the I's content: it is still a purely intellectual representation (cf.B423na) and its representation still has no content conceived as an object (cf.A381-2). In What Real Progress Kant says that the twofold 'I' does not mean a twofold personality: 'But a twofold personality is not meant by this double I. Only the I that I think and intuit is a person; the I that belongs to the object that is intuited by me is, similarly to other objects outside me, a thing'. Consequently, Kant understands that 'to many the double I in the consciousness of myself - namely, the I of inner sensuous intuition and the I of the thinking subject - seems to presuppose two subjects in a single person'. But, what we have is really two forms of I, not two subjects as two personalities.

Kant, therefore, insists that consciousness in general can be divided into a discursive consciousness and an intuitive consciousness. The point of this distinction is to stress that when talking is of the discursive consciousness it should be represented as not involving

---


85. Cf. also Reflexion 5453, AkXVIII186: 'All inner experience is (has) a judgement in which a predicate is empirical and the subject is I. Independently of experience, therefore, there remains merely the I for Rational Psychology, for the I is substratum of all empirical judgements' (Allison's translation, 1983, p.262). In A346/B404 Kant writes of the transcendental self-consciousness as follows: 'Consciousness in itself is not a representation distinguishing a particular object, but a form of representation in general'. Cf. Allison (1983), pp.280-3.


knowledge in a strict sense, since what we are dealing with is merely a pure apperception of our 'mental activity', as Kant puts it. However, this 'mental activity' is only reflexio:

Knowledge [of empirical things in general] (since it is based on judgements) requires reflection \textit{(reflexio)}, and, accordingly, consciousness of our activity in combining the manifold of ideas according to a rule of the unity of the manifold - that is, it requires \textit{concepts and thoughts} in general\textsuperscript{88}.

Kant can then argue that the \textit{discursive} consciousness is simple, that is, that the '‘I’\textit{ of reflection} contains no manifold and is always the same in every judgement, because it is merely the formal element of consciousness\textsuperscript{89}. In the first \textit{Critique} this ‘I’ is said to be a ‘simple representation’ through which ‘nothing manifold is given’ (cf.B135). For this reason there is no question of claiming \textit{knowledge} of this thinking (or conscious) subject (or self).

After asserting in \textit{What Real Progress} that the conjunction of the I that I think and intuit is a \textit{person}, and that a person \textit{has} (and not is) a body (‘the I that belongs to the object that is intuited by me is, similarly to other objects outside me, a thing’, but this I is not a person), Kant states that ‘I’ does only represent the subject, or as I put it, the subject represents itself as subject only when it uses ‘I’.

Absolutely no further knowledge regarding the I in the first sense (the subject of apperception) - the logical I, considered as an \textit{a priori} representation - is possible, neither with respect to what sort of being it is nor [with regard to] its natural constitution. It is as if the substance were what remains when all of its inhering accidents have been stripped away, but cannot in any way be further known [when stripped of its accidents], because its accidents are precisely that by means of which I can know its nature\textsuperscript{90}.

This last quote presents us with a reason to why Kant maintains in other places that ‘in the synthetic original unity of apperception, I am conscious of myself not as I appear to myself, nor as I am in myself, but only that I am’ (B157). Also that ‘[‘I’ in the universal proposition ‘I think’] is only the formal condition, namely, the logical unity of every thought [cf.A363], in which I abstract from all objects; but nevertheless it is represented as an object which I think, namely \textit{I myself and its unconditioned unity}’ (A398, my italics: AK). Kant also maintains that ‘apperception is something real, and its simplicity is already given in the mere fact of its possibility’(B419). In the quote Kant uses the notion, perhaps

\textsuperscript{88}. \textit{Op.cit.} p.Ak142, my italics: AK; cf. the same place for the other point as well.

\textsuperscript{89}. \textit{Loc.cit.}

\textsuperscript{90}. \textit{What Real Progress}, p.Ak270.
misleadingly, of a substance as a substratum of accidents which inhere in it to make a point about the I of the subject of apperception, *i.e.* that the use of 'I' is a real use, that it is not totally free-floating (that is, a use which is not in any contrast with 'you' or 'he')\(^91\). But about the substratum to which 'I' points we can not know anything, not even 'what sort of being it is'. This is a radical restriction indeed: 'I' as presenting a subject only\(^92\). But, then, this is what Kant means to point out by qualifying transcendental self-consciousness with 'formal'.

Externalism does not want to deny the role of this kind of restriction, but it does want to put it into its proper context. Kant's intention is to deny that the apperceptive use of 'I' is dependent on any criteria that appeals to what is given in experience as subject. In the Deduction B Kant says that 'I think', which we can take as the expression of transcendental apperception or self-consciousness, 'is an act of spontaneity, that is, it cannot be regarded as belonging to sensibility' (B132). Kant intends to stress the role of the subject as agent, performer and follower of rules. In truth, no awareness of a subject as having a natural constitution or being of a certain kind could possibly be dependent on what is for Kant the most important feature of this transcendental subject: its \textit{a priori} identity. Externalism promises to respect this feature without turning the transcendental subject into a \textit{transcendent} object which operates as a \textit{deus ex machina} in our philosophical reflections.

As we have already seen in various contexts, Kant's goal is to argue that the \textit{a priori} ground of all concepts is the \textit{numerical} identity of the transcendental apperception (cf.A107, 114). As we also saw, Henrich thinks that 'numerical identity belongs to self-consciousness as such', as 'a mere thought'\(^93\). In our brief discussion of Henrich we argued for the inadequateness of the view that Kant could be concerned with the 'numerical identity' of a \textit{type} of expression on the basis of the intrinsic implausibility of that suggestion. But, then, we should perhaps not place much weight on the notion of 'numerical identity' in this context at all, given that this concept is used by Kant himself in another context (in relation to persons) in a wholly different but clear manner:

The identity of the consciousness of myself at different times is therefore only a formal condition of my thought and their coherence, and in no way proves the numerical identity of my subject (A363).

Kant has in fact something special in mind when he turns the identity of the subject into the condition or ground for the use of all concepts. It is not, however, the conception of a platonic form, nor of an empirical thing as such. Externalism argues that it is the concept of an embodied subject with the ability of being conscious of itself as following a rule, and this consciousness is what is expressed with ‘I’. So, Kant’s position here, according to externalism, is that experience is only possible through the judging which is the following of rules, but then this judging has to be a conscious activity on the part of the subject, or an activity regulated by the use of ‘I’ (it should be clear that ‘I’ does not need to appear in acts of judgement, as if it were required that an assertion should be expressed through ‘I think...’). This is the point Kant makes in the *Groundwork* when he says that we have to conceive reason as being consciously directed in its judgements only from inside, otherwise we could not attribute the determination of our power of judgement to our reason. It would, then, have to be credited to impulse, without being in this case a power of judgement we possess. Kant concludes with a conceptual articulation: ‘Reason must look upon itself as the author of its own principles independently of alien influences’.

In the loose leaf “On Inner Sense” Kant distinguishes the identity of self-consciousness we are interested in as that of the subject who says only ‘I am’, that is, a subject insofar as she is a pure apperception (who is an apperception percipientis). This is contrasted with the subject’s knowledge that she is a thing of a past, present and future time, and Kant asserts that insofar as the subjects say I was, I am and I will be, we are concerned with the knowledge of an empirical history. Thus far the subject has an empirical apperception, an apperception percept. It is with this kind of distinction in mind that Kant talks of the need for self-knowledge of self-intuition, meaning that that knowledge can only be obtainable of something that can be considered as conditioned by an a priori form, which in the case of self-knowledge is time, ‘which is sensible and belongs to the receptivity of the determinable [in me]’ (B158na). But this is clearly distinguished from the discursive consciousness. So, we do not have a self-intuition of the determining, of the conditioning, in

---


95. Cf. *op.cit.*, I, lines 22-5.
ourselves. What we have is only the consciousness of the spontaneity of this act. This is because 'I cannot determine my existence as that of a self-active being' (B158na). So, according to Kant, we are limited, when it comes to this kind of claim, to representing to ourselves only the spontaneity of our thought, that is, of the determination. We can therefore, precisely in virtue of this spontaneity, entitle ourselves intelligences (cf.B158na)96.

What we find here is Kant stressing the connection between discursivity and the consciousness of the determining in ourselves, that is, of ourselves as spontaneous beings. Now, this consciousness is just what we have when we are ‘conscious of the identity of function’ whereby we synthetically combine the manifold in one knowledge (cf.A108). Furthermore, the reason for this equation is that ‘being an act of the self-activity of the subject, it [the synthesis] cannot be executed save by the subject itself’ (B130). Here we find the explanation that the consciousness of our spontaneity can only be the consciousness of self-activity, which is the consciousness that the subject executes a synthesis. Self-activity presupposes consciousness of the conjoining of one representation with another (cf.B133). This kind of capacity, of being conscious of what we do or are doing insofar as it involves meaning, as it is normative, is then linked by Kant with an essential feature of our understanding, a feature which is sometimes thought of, misleadingly, as a limitation: our understanding is not a faculty of intuition, it is a discursive and active faculty.

Now the understanding in us men is not itself a faculty of intuitions, and cannot, even if intuitions be given in sensibility, take them up into itself in such a manner as to combine them as the manifold of its own intuition. Its synthesis, therefore, if the synthesis is viewed by itself alone, is nothing but the unity of the act, of which, as an act, it is conscious to itself, even without [the aid of] sensibility, but through which it is yet able to determine sensibility (B153).

Kant does not attempt to go behind this in the clarification he offers in the central parts of the first Critique. He nonetheless makes claims about the synthesis as the unity of the act, an act of the subject, and about the subject’s necessary consciousness of the unity of the act. These claims allow one to appreciate that Kant’s purpose is to connect the notion of a subject as an agent with the requirement that thus far the subject must be aware of itself in this capacity. The subject then appears as the subject that does not question its identity

96. Allison uses this type of point to present Kant’s refutation of materialism, cf. (1989), specially pp.202-3.
insofar as it is able to follow a rule (cf.B145). 'The original and necessary consciousness of the identity of the self is thus at the same time a consciousness of an equally necessary unity of the synthesis of all appearances according to concepts, that is, according to rules, which not only make them necessarily reproducible but also in so doing determine an object for their intuition, that is, the concept of something wherein they are necessarily interconnected' (A108).

So, the identity of the subject which is Kant's concern is the identity of the subject in its conscious following of rules, which is what makes a concept possible in the first place, because there is no following of rules without the possibility of this consciousness. 'Synthesis according to concepts [is that] in which alone apperception can demonstrate a priori its complete and necessary identity' (A112). As Kant remarks in the Paralogisms: 'Self-consciousness in general is therefore the representation of that which is the condition of all unity, and itself is unconditioned. We can thus say of the thinking "I"... that it does not know itself through the categories, but knows the categories, and through them all objects, in the absolute unity of apperception, and so through itself. Now it is, indeed, very evident that I cannot know as an object that which I must presuppose in order to know any object, and that the determining self (the thought) is distinguished from the self that is to be determined (the thinking subject) in the same way as knowledge is distinguished from the object' (A401-2; cf. also B422 on why the self-consciousness of the subject that follows rules must be presupposed, for otherwise no thinking of object would be possible; and still B278).

We can conclude by noticing that Kant is presenting the following picture of what is fundamental to the possibility of objective knowledge. To be able to follow a rule the subject must recognise its act as its own, that is, must be able to take itself as following the rule, must be able to consciously follow the rule. It is the ability of transcendental self-consciousness, with its a priori identity as the subject that is following the rule, which is fundamental in accounting philosophically for rule-following. This is why Kant claims that nature, insofar as it can be an object of experience 'in respect of its conformity to law' (through the rules which are the categories), is dependent upon such a feature called by Kant 'our subjective ground of apperception'. Nature, in its formal aspect, as the existence of things insofar as they are determined by universal laws\(^7\), insofar as the understanding

\(^7\) Cf. Prolegomena, #14. Cf. also A124-9, B163-5 and A216/B263.
through the categories is the ground of nature's necessary conformity to law (*natura formaliter spectata*) (cf.B165), is dependent on the transcendental apperception as the 'radical faculty of all our knowledge', as the capacity which accounts for 'the first sources of our thought' (cf.A114). This is why Kant refers to the transcendental apperception as the highest point to which his philosophical analyses take us (cf.A116, A127, B134na).

What transcendental apperception presupposes to constitute such a fundamental faculty for Kant, is the use of the bare representation 'I', which makes all other representations possible (cf.A117na). 'I', as one and the same, as denoting universal self-consciousness (cf.B132-3), is the abiding and unchanging correlate of all representations if we are to be conscious of them (cf.A123). What Kant wants to stress is that this possibility of consciousness as a faculty has to be admitted for the sake of the possibility of representations. That is, what we have to recognize, according to Kant, is that the pure understanding, as a discursive faculty, is a formal and synthetic principle of all experiences by requiring the use of categorial rules (the use of the categories), and thus makes 'appearances' have a necessary relation to the understanding (cf.A119). But, then, the understanding, in the guise of the transcendental apperception, is what makes action based on rules possible, thus it can with all justice be designated the faculty of rules (cf.A127), that is, as the basis for the possibility of the following of rules by any subject that possesses understanding.
1. Introduction

It is relatively easy to praise Kant for his criticism of Cartesianism. Kant comes across as a powerful thinker, and many feel in their bones that there is something deeply wrong with Cartesian philosophical claims anyway. Furthermore, Kant's diagnosis that the position confuses the unity of consciousness with the intuition of the subject as object (cf.B421) seems to strike just the right chord. However, it is not so easy to say exactly what Kant saw fit to criticize in Cartesianism. For Kant uses in his own positive claims terms that have a very Cartesian ring to them. As we have already seen, he asserts that in transcendental self-consciousness I am conscious of myself only insofar as I am (cf.B157). He also says that apperception is something real (cf.B419), and he claims that even though not much is represented in the concept of 'the thing which thinks', at least it yields the notion of a transcendental subject of the thoughts = X (cf. A346/B404). Kant is also prone to speak in this context of a 'substratum of accidents'. Supposedly, it is also in this connection that Kant speaks of 'I think' as an actus (cf.B423n). The difficulty is created by the fact that these terms and claims in Kant fulfil a dual role: they are used to present Kant's own position and to formulate, using sometimes very similar phrases, the targets of his criticism (Cartesianism and its heirs).

However, no one should doubt Kant's willingness to present an alternative to the views presented to him by the tradition. Kant should not be said to be engaged in presenting a true 'alternative' though, given that, once again, his interest is 'transcendental': the conditions of knowledge are his subject-matter, not the establishment of another doctrine. Therefore, in keeping with the methodological reflections we considered in the last chapter, although Kant's assertions in this area can also be called 'Rational Psychology', they do not exist 'as doctrine, furnishing an addition to our knowledge of the self, but only as discipline' (B421). Accordingly, Kant criticizes both Cartesian and Empiricist conceptions of self-consciousness for their common assumption: that self-consciousness is a form of self-knowledge. The former equates it with knowledge of an immaterial subject, the latter with knowledge of particular experiences conceived of as constituting such a subject. Kant's
views about the 'I think' imply neither that we know the subject thereby, nor that we are aware of particular experiences within us. This is why, particularly, against the latter kind of position, Henrich insists that Kant's interest in 'I think'-instances is not an interest in the 'states of a thinking and sensing individual', that is, in mental states. According to Henrich, undue focus on what are deemed to be these mental states hinders 'the attempts in the Anglo-Saxon Kant literature to secure a basis for the Deduction in the postulate of the possibility of self-knowledge'. Therefore, Kant does not side either with 'spiritualism' or with what leads to 'materialism', since his intention is rather to undermine both positions (cf.B416-21; A383 and ff.).

For example, we have seen that the subject or self Kant is interested in is simple, identical with itself, etc, in the formal or transcendental sense of these qualifications. Knowledge, on the contrary, requires intuitions in general (cf.B406-7). Knowledge of the 'inner', in particular, is conditioned by what is the 'outer', that is, by what is 'outwardly' intuited. Accordingly, in the Observation to the Second Antinomy Kant remarks: 'Self-consciousness is of such a nature that since the subject which thinks is at the same time its own object, it cannot divide itself, though it can divide the determinations which inhere in it... Nevertheless, when this subject is viewed outwardly, as an object of intuition, it must exhibit [some sort of] compositeness in its appearances; and it must always be viewed in this way if we wish to know whether or not there be in it a manifold [of elements] external to one another' (A443/B471; cf. also Kant's remarks on why there is no a priori knowledge of a thinking being, viz. because here is nothing abiding in 'inner intuition', A381 and ff.. These remarks lead Kant to affirm that 'nothing is left for us but to study our soul under the guidance of experience', A382).

One should not, however, entertain the possibility that the 'conditioning' of the 'inner' by the 'outer' in Kant is a causal affair. As Allison notes, inner experiences are always correlated with outer experiences, and what Kant wants here is to present them as 'bound up in the way of identity' (BxII). Thus, in the Fourth Paralogism A, the relation between outer objects (extended beings) and inner objects (my thoughts) is denied to be an

---


2. Cf. Westphal (1995), for the view that Kant has excellent theoretical reasons for denying psychological determinism, one of them being that Kant proscribes causal judgements not only in Rational Psychology but also in Empirical Psychology. Cf. also A 386 and ff.
inference (cf.A372). In the B version, the distinction between myself as a thinking being with presumed ‘internal objects’ and the things ‘outside’ me - including my body - is made through an analytical proposition (cf.B409; also B418 and A386-7, where inner and outer senses are distinguished through transcendental considerations).

This relation between the ‘inner’ and the ‘outer’ in Kant is part of his general point that objects are known immediately, and that there is no epistemic priority possessed by inner experience or self-knowledge over outer experience. ‘I am just as certainly conscious that there are things outside me, which are in relation to my sense, as I am conscious that I exist as determined in time’ (Bxlin, cf. the whole note). We are here in the territory of the Refutation of Idealism, whose exploration would take us away from our main goal.

In this final chapter, I will present Kant’s criticism of the Cartesian and Empiricist positions in view of what has been established in previous chapters. This will close the argument of this dissertation by revealing what is really novel in Kant’s philosophy of self-consciousness.

2. The Critique of Cartesianism

Descartes certainly seems to have a very specific view of the subject which is of interest in philosophy. This subject is for Descartes a substratum of attributes and modes, i.e. a truly metaphysically conceived substance. In the “Preface to the Reader” of the Meditations, he calls the mind a ‘pure substance’: ‘the human mind is not made up of any accidents in this way [as the body is], but is a pure substance. For even if all the accidents of the mind change, so that it has different objects of the understanding and different desires and sensations, it does not on that account become a different mind; whereas a human body loses its identity merely as a result of a change in the shape of some of its parts’. As we have seen, the above mentioned ‘accidents’ of the mind, its modes in fact, such as particular thoughts, are known with absolute certainty. There is, according to Descartes, indubitable knowledge concerning my thinking and its contents. This view follows from the Cartesian definition of thought and the consequent account that these modes belong to a substance. So,

3. Cf. Allison (1983) p.304, where the point mentioned earlier can also be found.

'I' in Cartesianism makes reference to a substratum, and its modes are immediately known as its states.

Another important feature of Cartesianism is the way it attempts to guarantee the identity of the subject as a particular immaterial substance through thinking alone. In the continuation of the "Preface" Descartes affirms that exactly insofar as the mind is a pure substance - that is, a particular thing that does not lose its identity in changes of its accidents, as bodies do - it is by its very nature immortal. This is why Kant says that Cartesianism tries to prove that the identity of the subject as an intellectual substance is the ground for ascribing personality to it (cf.A345/B403). In A367 Kant explains that 'intellectual' has the import of being beyond what can be given directly in perception. But it is exactly in these terms that Cartesianism conceives of the identity of our particular immaterial souls.

Kant criticizes both features of Cartesianism. He criticizes the view of the substantiality, the purity or simplicity, and the identity of the self as presented by this position. For him, these are not 'material' features of a thing but rather 'formal' characteristics of what we humans do. The identity of the transcendental subject, for example, is not the identity of a subject in the world, conceived either as a material thing or as a spiritual thing. It is rather, as we have seen, the identity of the subject insofar as she is a rule-follower.

Kant's general charge against Cartesianism is, then, that it commits the subreption of hypostatising consciousness, it transforms apperception in an apperceptionis substantiatae (cf.A402). Let us see how Cartesianism is criticized by Kant in its more specific claims.

Kant wrote two versions of his analyses of the specific errors of Rational Psychology, the doctrine which enshrines Cartesianism. He called the inferences of this doctrine in trying to establish its claims 'paralogisms', i.e. fallacious syllogisms based on the incomprehension of a transcendental point being used in them (cf.A341/B399). Kant accuses these syllogistic inferences of being sophisma figurai dictionis, i.e. errors of reasoning by the utilization of an ambiguous middle term (cf.A402, B411). According to Kant, the notions that are not being properly understood are chiefly those of substance (and consequently of subject or self in this philosophical debate), simplicity and identity.

Before we look at what the rational-psychologistic and the critical construals of these notions are, we should bring to bear on the debate a constraint that comes from the very
nature of our intellectual enterprise, that is, of philosophy. Kant insists that, and Cartesianism seems to have no quarrel with this, the ‘knowledge’ we are trying to achieve in doing Rational Psychology has to be independent of all experience, of everything that is concreto, and must concentrate only on concepts (cf.A342/B400). Kant stresses that we are here restricted in our analysis to having as our ‘sole text’ the ‘I think’ (A343/B401). This accounts for the purity which is said to accompany this science. Kant shares this view of what is possible to achieve in philosophy, but we already get a clue to his unique line of thought when in the beginning of the discussion he says that the predicates dealt with in Rational Psychology are ‘transcendental predicates’ (cf.ibidem).

Whereas for Cartesianism substance can be that immutable pure substratum of modes or accidents, for Kant ‘substance’ has to be used with a different acceptation. Cartesianism tries to establish that we are such a substance by considering that we are a subject of thoughts and that thoughts must inhere in a substratum which is not made up of its accidents. However, for Kant the categorial claim that something is a substance requires for its serviceability in the production of genuine knowledge, i.e. empirical knowledge, the permanence of an object given in experience. Only by incorporating the notion of such an object to our concept of a substance can we claim to have a piece of knowledge in saying that something is a substance (cf.A349-50). Kant, therefore, dismisses the Cartesian claim to knowledge of the real subject by saying that it is ‘a mere pretence of a new insight’, and ‘does not carry us a single step further’ (A350).

The simplicity or purity Descartes sees in the substratum is also challenged by Kant. It is not challenged through the means of a principle of significance, as if we could test the purity of the substratum Descartes asserts. We have seen in chapter 3 that this Strawsonian approach is not Kant’s. Kant rather concentrates again on how we use the notion of simplicity. He reminds the Cartesian that the concept of something absolutely simple, which does not have its purity affected by the change of its accidents has no use in a genuine attempt to acquire knowledge about something existing in the world. As substance is used in conjunction with the permanent, so the simplicity of a substance is used relative to what is empirically composite (cf.A400, where Kant explains what can be simple in the field of appearances). Absolute simplicity or unity carries with it, on the contrary, a connotation of necessity which Kant insists can never be derived from or given application in experience (cf.A353). Kant makes the same point in his demonstration of the inconsistencies of reason
in the Second Antinomy, where the absolutely simple is asserted to exist and not to exist. Furthermore, we should remember that for Kant there is no *infima species* (cf. A655-6/B683-4), of which thesis the ontological reading could be - had Kant a sort of nomenclaturist conception of concepts and language - that there are no simples in the world. However, Kant's strategy is rather to clarify what is meant by the *transcendental* claim that the subject is simple.

Kant asserts that the simplicity connected in this context with the subject is only the simplicity of the subjective ‘*I*’. In explaining what this means, Kant considers the difficulty of representing to ourselves a ‘thinking being’. He claims that this can only be done by assuming that ‘*I*’ is presupposed in all thinking. Thus, what we have to do is to place ourselves in the position of such a putative thinking being by substituting our own subject for the object we are trying to represent. That is, we cannot represent such a being to ourselves ‘save by putting ourselves, *with the formula of our consciousness*, in the place of every other intelligent being’ (A354, my italics: AK). What Kant means is that we have to ascribe to this being the use of ‘*I*’. The simplicity here, according to Kant, is then only that of a representation, *which governs only the verb in its relation to a person*. It is obvious that in attaching ‘*I*’ to our thoughts we designate the subject... only transcendentally, without noting in it any quality whatsoever - in fact, without knowing anything of it either by direct acquaintance or otherwise. ...The representation of [the transcendental subject] must, no doubt, be simple, if only for the reason that there is nothing determinate in it’ (A355, my italics: AK).

It is important to notice that Kant is stressing again in these passages that we have no knowledge at all of the transcendental subject, but at the same time he is connecting the use of ‘*I*’ to persons and intelligent beings. Such qualifications call for an externalist account. Accordingly, the correct explanation is that the transcendental subject is the corporeal agent insofar as she has the ability to use ‘*I*’. Kant makes the negative point as follows: ‘The simplicity of the representation of a subject is not *eo ipso* knowledge of the simplicity of the object itself, for we abstract altogether from its properties when we designate it solely by the entirely empty expression ‘*I*’, an expression which I can apply to every thinking subject’ (A355). To fully appreciate Kant’s meaning here it is important to realize that the ‘designating’ of the subject is not the making of a reference to it, but is rather attributing the mastery of a representation (*‘I’*) to a subject insofar as it ‘govern’ certain verbs in
connection with this subject as a person. 'I' is an empty representation (in itself), not a concept; but a bare consciousness which accompanies all concepts. Thus, 'consciousness in itself is not a representation distinguishing a particular object' (cf. A346/B404; also A382). Furthermore, through 'I think' I do not know an object, I am only aware of a determining self, not of a determinable one, determining of judgements (cf. B407). Kant, as a consequence, states that 'I' is not an empirical representation, it is rather purely intellectual insofar as it belongs to thought in general (cf. B423n).

The dialectical illusion in rational psychology arises from the confusion of an idea of reason - the idea of a pure intelligence - with the completely undetermined concept of a thinking being in general. I think myself on behalf of a possible experience, at the same time abstracting from all actual experience; and conclude therefrom that I can be conscious of my existence even apart from experience and its empirical conditions. In so doing I am confusing the possible abstraction from my empirically determined existence with the supposed consciousness of a possible separate existence of my thinking self, and I thus come to believe that I have knowledge that what is substantial in me is the transcendental subject. But all that I really have in thought is simply the unity of consciousness, on which, as the mere form of knowledge, all determination is based (B426-7).

What Kant is interested in stressing is the conception of something simple in the concept only: 'If I know something simple in the concept only and not in the [field of] appearance, I have really no knowledge whatsoever of the object, but only of the concept which I make for myself of a something in general that does not allow of being intuited' (A400). This is why knowledge is out of place in this context and the concept of a transcendental subject, although not lacking objective validity, has no objective reality, that is, among other things it does not serve as such to make claims about how matter in general is dissimilar from the realm of thought (cf. A357).

In the Third Paralogism Kant criticizes the Cartesian attempt at securing the numerical identity of the substratum that underlies all the accidents perceivable when we think, will, etc. Kant cannot agree with this picture's conception of persons as immaterial substances with numerical identity in virtue only of their character as intellectual subjects of thoughts. For Kant the conception of the identity of a person has other implications. Personhood requires intuitions of the subject whereby it is given as an object, and identity of person requires identity of one's own substance throughout the changes of its states. This again refers us to the permanent in intuition, or in the field of appearances, which is the basis of any claim about a substance, including about that substance we attribute to a person.
Kant allows the talk of a transcendental conception of personhood, but taken in this way 'we can never parade it as an extension of our self-knowledge through pure reason, and as exhibiting to us from the mere concept of the identical self an unbroken continuance of the subject' (A366). Yet again Kant takes the stance that for a genuine claim to knowledge about the numerical identity of a substance to be admissible it must be conceived as a synthetic judgement about the substance in question. Numerical identity claims for Kant are linked with space and time and can only be established through the justification of synthetically \textit{a posteriori} judgements. However, the statement of consciousness of my identity with myself is for Kant nothing more than an analytical proposition, so no genuine claim to knowledge of a numerical identity of a substance is thereby made (cf.B408-9). I shall return to the Kantian conception of persons below when dealing with the critique of empiricism. I will close the discussion of Kant's differences with Cartesianism with a point about how difficult it is to truly appreciate the nature of transcendental requirements, and this will give me the opportunity to criticize one mistaken view held by Allison.

It is quite correct to understand Kant's point here as implying that the identity of the transcendental subject is not 'inferred' nor the result of a 'judgement of identity', but rather that it is a point about a necessary identity. I take it that this is what Kant intends to show also with his example of the elastic balls in A363-4n. To be one and the same person requires more than the use of 'I', as we know for example from the competent use of 'I' by amnesiacs - that is, it does not provide knowledge of identity or self-knowledge.

However, Kant also asserts that 'in my own consciousness, therefore, identity of person is unfailingly met with' (A362). This has to be taken with care. Allison believes that Kant is here making a point about what he calls 'self-consciousness in time', as opposed to the 'transcendental consciousness' or the 'logical I', that is, the necessarily identical transcendental self-consciousness. Allison tries to make this distinction through an illustration. He says that I can be uncertain about whether I had a headache I now remember having had yesterday, but not about whether I was the one who had it if at all someone had it. According to Allison this is a point about self-consciousness in time and not about the transcendental self-consciousness. I cannot understand this distinction. I think this is a distinction without a difference, basically because time does not come into the relevant considerations in the illustration. The only thing that matters is the mastery of the use of
'I'. Time does not come into the consideration of what conditions the identity of the transcendental subject. As Henrich puts it, 'time does not belong to the conditions to which thought and self-consciousness as such are necessarily subject'. This is because 'Kant does not attribute change of states to self-consciousness when its identity is at issue in the context of the Deduction', and the reason for this is that 'what cannot be determined in time has no state'. Although time is mentioned in Allison's illustration, the guarantee of identity there depends exclusively on the logical features of the use of 'I'. There is no actual passing of time involved in the claim that I was the one who had headache yesterday if I remember having had it. This is why Kant claims that 'the subject, in which the representation of time has its original ground, cannot thereby [by thinking the categories] determine its own existence in time' (B422). A transcendental condition is being confused with a special determination of the subject in what is deemed to be its mental constitution.

3. The Critique of Empiricism

It is sometimes claimed that it is far more difficult to assess what Kant's view of the bundle theory was, than to understand what Kant's response to Descartes was. In our discussion of Kant on the paralogisms of Cartesianism, we have seen in several places - especially in connection with Strawson - that the Kantian critique is not what it is sometimes assumed to be. However, in relation to the Empiricists, we should start with the fact that, as Mendus points out, Kant's concern was not to provide a solution to the problem of personal identity as the Empiricists conceived of it. Neither the consideration of what unites a bundle, nor identity over time are Kant's concern when he states the identity of the subject in self-consciousness; because this identity is already built into the possibility of self-consciousness.

5. Cf. Allison (1983) pp.304-9 for his discussion. Cf. also Hurley (1994, pp.139-41) who thought that this Allisonian conception of self-consciousness in time really did grasp Kant's points about the transcendental unity of apperception, with the result that Kant would be really tackling issues related to what gives unity to a succession of mental states. Unfortunately this is not what Kant addresses in his transcendental investigation.


itself. So for the Empiricists the interest of Kant’s message concerns the appreciation of the logical situation involving the self. Kant is not concerned with one particular empiricist development, he wants to challenge the assumptions that lead to the empiricist type of account.

We have seen how Locke and Hume reacted against the use of ‘substance’ by Descartes as a means to guarantee personal identity. We have also seen how Locke, for example, maintained that a person is a thinking being with reason and reflection, which has its identity in different times and places secured only by the consciousness Locke conceived to be thinking. In this approach not much attention was given to the sameness of man. In the end everything depended on Locke’s account of the thinking he understood as self-recollection or memory. As for Hume, we have seen that memory was not enough, that it only ‘discovered’ personal identity, and that causation had also to be brought in. In the end, however, Hume claimed that thought alone found personal identity, and that it was secured by a consciousness or reflected thought or perception, which was a view that could not convince Hume himself. The first aspect of Hume’s position is, as we have seen, very Lockean. But, then, the concerns Hume expressed at the end of his investigation pointed to flaws in the logical appreciation of the matters involved.

Kant indicates what his interests are in a passage of the Anthropology, making clear where he stands in relation to the Empiricists. He asks a typical empiricist question about personal identity, then he gives his verdict and goes some distance towards clarifying what are the fundamental logical features of the issue at stake.

Given the various changes within a man’s mind (of memory or the principles he accepts), when he is conscious of these changes can he still say that he remains the very same (as far as his soul is concerned)? The question is absurd. For it is only because he thinks of himself in these various states as one and the same subject that he can be conscious of these changes; and man’s ‘I’ is indeed twofold in terms of form (manner of representation), but not in terms of matter (content).9

In this passage we can see what Kant believes makes sense to ask about the identity of the subject. Kant counters the empiricist question about personal identity by reminding

---

9 Op.cit. p.Ak134n. Cf. Henrich (1989, pp.265-6) for the point that Kant knew of Locke’s problem and the solution he offered through Christian Wolff. But Kant had other issues and solutions in mind, simply because he maintained that ‘the fact that self-consciousness genuinely exhibits “logical” identity does not permit one to infer that a substance with the capacity for remembering underlies it’ (loc.cit.).
them of the role of the transcendental subject, that is, the proper use of 'I'. But at the same time a space is provided for genuine questions of identity. For Kant the identity of a particular person as instantiating inner and outer predicates cannot be proved logically, it has rather to be determined empirically (cf. B408-9 and A382 where Kant says that we can only study the soul under the guidance of experience). However, in dealing with questions about whether someone who changed completely his personality in the ordinary sense, after an accident for example, is still the same person, or whether someone with a great number of transplanted organs remains the same, one thing according to a Kantian perspective has to be taken for granted: our ordinary conception of persons. It is here that we find in Kant a very Strawsonian conception. In the last sentence of the quoted passage we find a reference to a twofold form of 'I' that does not amount to a twofold content of the word. We also saw in the last chapter that Kant never envisaged the subject to have two personalities because of different uses of 'I'. What I take Kant to be saying here is that the notion of a person is used in connection with 'the I that I think and intuit', that is, in connection with an I that belongs to an object that is intuited by me. This means that the notion of a person is that of something which is 'similarly to other objects outside me, a thing'. Perhaps the crucial difference between Kant and Strawson is that, despite the fact that persons have M- and P-predicates applied to them, not all P-predicates for Kant are of the sort normally taken to designate experiences. Henrich correctly insists that self-consciousness in Kant should not be construed in terms of mental states of a thinking and sensing individual. In Kant, then, some P-predicates are applied according to a constraint to the effect that we follow rules independently of actual experiences: we have understanding, intentions, etc. However, there is, nonetheless, a fundamental tenet on which Kant and Strawson agree, which can be exposed by their common reaction to materialism.

I have been arguing that Kant's conception of a transcendental subject is externalist, and one feature of externalism is that the subject in question is conceived of as a corporeal agent. Now, it is tempting to understand this subject, and consequently the notion of a person, as simply a material body that is so complex as to be able to use language, etc. This

10. Cf. Mendus (1984, p.137) on Kant's lack of interest in 'post-fission' persons. Whether they continue or not as the same is an open question. Cf. also Kant's 'fancy', intellectual 'adventure', of the creation of the soul of children through the dynamical division of the parent's souls (B415-8n).

11. These are passages from What Real Progress (p.Ak270), already quoted in the last chapter.
is, for example, Williams' position about what he calls the macro-level of the mind-body problem. For this author, persons are nothing other than material bodies that think, are conscious, etc. This is a line of thought that cannot conceive of the elements in the externalist position without incurring materialism. But Kant is not only not inclined to entertain spiritualism, he also rejects materialism.

Persons for Kant do not get their use of language or their P-predicates explained reductionally by appealing to the complexity of the functioning of their material components (cf. B415-8n where Kant argues against the intelligibility of materialism and discards the issue of emergentism with its question about whether the unity of apperception in thought could arise out of the composite). Persons in the Kantian perspective have an intelligible character that elevates them from nature as such, giving them the status of a unique category\(^\text{12}\). For Kant, being a person involves having features that are not reducible to each other\(^\text{13}\).

At A342/B400 Kant says that it comes from the nature of our faculty of representation that we distinguish two kinds of objects, that is, an object of the inner sense or a soul and an object of the outer senses or a body. Kant also makes it clear that it comes from the way we represent things that M-predicates should be distinguished from P-predicates. Thoughts do not have predicates involving place, motion, shape, and other spacial determinations (cf.A387). This prompts Kant to hold a dualism 'only in the empirical sense' (A379). This is a view that recognises the differences in the way we represent to ourselves M- and P-predicates: outer and inner senses 'are extremely unlike each other' (A386). It maintains that we have a soul whose permanence during life is

\(^{12}\) Kant never formulated the claim that 'person' forms a logically primitive concept, that it is a category on its own. But the points he makes concerning the use of this term do imply, in my view, just such a claim. Glock & Hyman (1994) defend the Kant-Strawsonian view that 'person' is a logically primitive concept against Williams' attack by clarifying the nature of categorial claims and stressing that, for example, Carter is not identical with his body, because 'Carter' and 'Carter's body' are not interchangeable either salva veritate or salva significatone. In fact, it does not make sense to say that Mary by loving Carter is only loving Carter's body. She may love Carter and dislike his body. It also does not make sense to suggest that Carter by being proud of himself is either only proud of his body or, even worse, that if Carter is in fact proud of his body then Carter's body is proud of Carter's body.

\(^{13}\) I suspect Cassam agrees with this. However, it is for basically Davidsonian reasons (cf.1989). I will not argue with this here. However, my account of self-consciousness in Kant should speak for itself against such an attempt, which is incidentally an attempt to make Strawson's views even more ontologically respectable.
evident *per se* since the thinking being *as man* is also an object of the outer senses (cf. B415). It does not amount to a 'transcendent' dualism. It rather holds that 'appearances in both fields [the inner and the outer] must be connected with each other according to rules which [substance] introduces into that connection of our outer as well as of our inner perceptions whereby they constitute one experience' (A379). According to Kant, we should not confuse modes of representing objects with determinations of how they are in themselves, otherwise we will end up holding a 'crude dualism' that claims that there is a communion between two kinds of substances, a thinking and an extended type (cf. A392, also B427-8).

Concomitant with the point that what we are dealing with here are modes of representing, is Kant's view of what we should realise as having priority in the present discussion. Kant sympathizes with the common expression that 'men think' (A360), and implies that men and women are the thinking beings who are permanent during life as long as they have bodies (cf. B415). This is a way of stating that 'the possibility of thought... [has] an example only in the empirical intuitions of our human life' (B416n). Consequently, Kant is happy with our use of 'soul', as long as we do not bestow on it a Cartesian import (cf. A350-1).

So, from Kant's perspective, we can agree with Strawson about the categorially primitive status of persons. Furthermore, it is also the case that, not only for Strawson, but also for Kant, persons are not just their bodies, and consequently that materialism has also, together with the crude Cartesian dualism, to be refused. Kant maintains that 'souls' are quite different from 'their bodies' (cf. A358), precisely because many instantiations of P-predicates cannot be 'outwardly intuited', but instead belong to the inner sense; and we should not take extension, impenetrability, cohesion and motion, in sum, 'everything which outer sense can give us' (A358) as characteristics of thoughts, feelings, desires and resolutions. But then Kant also wants to make the point that the 'soul' is not distinct from matter (cf. A358-61). In stating this claim (cf. also A383-6), Kant turns to a mode of speaking that appears to be 'transcendent', as if what, beyond the conditions of knowledge, is responsible for the transcendentally ideal conception of matter could also be responsible for the transcendentally ideal conception of the soul. But, given that the 'soul' can only be known in experience, and that the empirical knowledge of ourselves - the only knowledge we can in fact achieve - is also under the strictures of transcendental conditions, the point
Kant is trying to make with this contention should not be inflated out of proportion. Kant wants to preclude the temptation of holding that, because there is a distinction between P- and M-predicates, there is a metaphysical distinction between immaterial and material substances (compare how Kant criticizes the ‘solutions’ to, but more interestingly the presuppositions of, the problems of communion, of the soul before birth and after death, A384-96). In this line, Kant seems to commit himself to the view that, even though the inner and the outer are distinguished analytically form one another, my own existence as a thinking being can only take place as an existence in human form (cf. B409; cf. also B420-1, for talk of our existence in this present life, and B426-7, quoted above, for the notion of a thinking being as an abstraction from what exists empirically determined).

Kant’s strategy here is to undermine the very assumption that makes both Cartesian dualism and its counterpart, materialism, positions one would want to consider. Materialism itself can only be a tempting position if one implicitly assumes that it is intelligible to hold the distinction between materialism and Cartesian dualism. Kant’s target is this claim to intelligibility. He maintains that the dichotomy itself is flawed. According to Kant both positions are misconstruing an important feature of some P-predicates. For Kant the solution to our troublesome reflection about the ‘soul’ and matter in this context depends on the recognition that men have an *inward respect* that has a special character (cf. A360). In this passage Kant is attentive to the fact that it is this feature that people are normally trying to explain by their metaphysical distinction between matter and soul. But he refuses the classical approaches to this feature of our dealings with ourselves.

We have seen how much Kant makes of the fact that we use ‘I’. This use is equated with self-consciousness, which then, by being an instance of men’s *intelligible character*, is distinguished from any other of his features, especially from men’s empirical qualities. It is this which represents the basis for our ‘souls’ to have an ‘inward respect’. We have seen already what the place is of such a view in Kant’s philosophy. Here it is important only to stress the centrality of such a feature for Kant. Consequently, against Empiricism the Kantian criticism is that the position transforms what pertains to this ‘inward respect’ of

14. Allison offers a Kantian attack on materialism by appealing, not to the primitive categorial status of persons, but rather to the role of the ubiquitous ‘I’. According to Allison, it is the account of conceptual thinking that cannot be for Kant the object of a causal explanation, rather we have to appeal to *intelligible* considerations (cf. 1989, *passim*).
humans, as it is revealed in self-consciousness, into an experiential possession of a particular of some sort, be it idea, perception, or whatever. Beyond the issue of the particularity of thoughts, ideas and perceptions, what is wrong with this transformation is that it leads to a further misconstrual: self-consciousness has then to achieve the attribution of these particulars to an object, which is supposedly done through a referential use of 'I'.

Consequently, Kant’s attention to this ‘inward respect’ distances him from Strawson to a certain extent as well, for the latter is still very much of a follower of the empiricist tradition. Strawson obviously avoids, say, the Humean metaphysically dubious *dictum* ‘perceptions are distinct existences’. However, he picks up the spirit of the position and tries to give it an ontologically respectable face: experiences are to be dependent particulars. Strawson makes much more of man’s body and shape, acknowledging even Kant’s influence. But then Strawson wants to account for the cognitive objectivity of claims concerning experiences. He believes that by giving experiences a sort of public basis in an empirical body he can give the use of P-predicates in general a suitable structure that can justify those claims. I have already dealt with the nature and shortcomings of this way of facing these problems, and showed how Kant differs from Strawson in this respect\(^{15}\).

**4. Kant’s Distinctive Metaphysics of Mind**

Against the background of the Cartesian and the Empiricist positions on self-consciousness it is possible to appreciate what is new and distinctive in Kant’s metaphysics of mind. I will conclude this chapter and this dissertation by calling attention to three points that have emerged from my discussions as crucial for a correct understanding of Kant’s philosophical position.

The first point I want to single out concerns the account of thinking. Against the Cartesian concentration on the accidents of the soul or mind, Kant is keen to point out that thinking is linked rather with the discursive capacity of the human understanding. Against

---

\(^{15}\) It is interesting to note that Henrich thinks that Strawson’s appeal to this structure of a common language is a version of Wittgenstein’s Private Language Argument. Consequently, Henrich, by refusing Strawson’s account of Kant, distances Kant’s project from Wittgenstein’s (cf. 1989, pp.278-9). Also Hacker resists the approximation of Kant and Wittgenstein, now because of Strawson’s misconstrual of Wittgenstein’s argument (cf.1986, pp.211-2). My explanation for this is that Strawson’s commitments to the empiricist spirit prevent him from giving a correct account of either Kant or Wittgenstein.
the model of thinking as states of a substance Kant insists on our faculty of judgement. This insight develops into his view that thinking is also not experiencing anything. This is now a claim that can be targeted especially against the empiricist views on thinking. It is in fact overtly opposed in spirit to much of what fosters Empiricism. Consequently, the claim also exposes the rationale behind what is new about Kant's position, which simultaneously allows him to present a positive view of transcendental apperception and to criticize the mistakes of the alternative positions. Kant articulated this claim explicitly very late in his philosophical career (1788-90), in a Reflexion entitled “Is it an Experience that we Think?” He answers this question at the end in the negative. His reasons are the following.

He distinguishes first the thought a priori [sic] of a square, which he maintains is not an experience, from the recognition of a drawn figure that was drawn according to the concept of square, which Kant claims involves experience. This is wholly in accord with an externalist position. Thus Allison reminds us well: the thought of the square consists of nothing else than the rule or procedure for constructing the figure, whereas the actual construction in space is an object from which one can learn about the properties of squares. Consequently, Kant maintains that experience involves an empirical consciousness of the seeing of an object in time, whereas thought, although it happens in time, ‘makes no reference to time in the thinking of the properties of the figure’. Kant's point is clear. Bearing in mind that it is men and women that think, thinking can be linked with time. This is without problems. But the thought, the content of it, this is not something that is in time or in space. Allison, following Ewing, accepts the point about the content of a

---

16. 'Wohl aber kann dieses gesagt werden [es sei eine Erfahrung], wenn ich eine schon gezeichnete Figur in der Wahrnehmung auffasse, und die Zusammenfassung des Mannigfaltigen derselben vermittelst der Einbildungskraft unter dem Begriff eines Quadrats denke' (AkXVIII319).

17. Cf. (1983), p.276 for the first notion distinguished. Allison, however, maintains a sort of mentalism concerning the second notion. He suggests that all that Kant has in mind in this discussion is a determination of an 'inner object' that then should be able to be experienced. This is his 'internal' representation or mental state (cf. pp.276-8). But notice what Kant is concerned with: ‘In der Erfahrung und durch dieselbe [of the physical construction according to externalism] werde ich vermittelst der Sinne belehrt [this cannot be the imagination]; allein wenn ich ein Object der Sinne mir bloß willkürlich denke [this is an imagining], so werde ich von demselben nicht belehrt und hänge bei meiner Vorstellung in nichts vom Objecte ab, sondern bin gänzlich Urheber derselben [in the sense of being responsible for the imagining]' (AkXVIII319; my additions: AK).

18. Cf. loc.cit.. Parts of this Reflexion were rendered into English by Ewing (cf. 1924, pp.136-7). Allison uses Ewing's rendition in his discussion, as I have.
thought, but maintains that this does not apply to what he takes, in a special sense, to be an ‘act of thinking’. The latter, he claims, happens in time and can be transformed into an object experienceable by the subject herself. According to Allison, this can be a true ‘object of inner experience’. Consequently, he reads a distinction (between a consciousness of what happens when one ‘institutes’ an experience and a consciousness of the act of instituting) into a Kantian text where none is meant. Kant’s text actually collapses the distinction which Allison thinks creates a space for him to introduce his ‘subjective objects’. Kant’s whole text is as follows (Allison leaves out the first sentence of the paragraph):

The consciousness, when I institute an experience, is a representation of my existence, insofar as it is empirically determined, that is, in time [existence in the only possible sense, as an agent in the world]. Were this consciousness in turn itself empirical, then the same time determination would again have to be represented as contained under the conditions of the time determination of my state [as an object in the world].

Externalism accounts for this in the following manner. To think is not to experience, and to follow a rule in, say, constructing a figure according to a concept does not require the experience of anything, even though thought as such cannot be separated from ourselves as subjects, that is, from transcendental subjects existing as corporeal agents in the world, the only way in fact one can exist. Now, the reference to time is related to Kant’s view of self-knowledge and cannot be dealt with here. What is important, however, is Kant’s intention to attribute spontaneity to thinking and to the transcendental consciousness of our following of rules. In addition, time determination is dependent on the transcendental conditions, and is consequently dependent on what conditions the whole of the functioning of the understanding, viz. the principle of the transcendental apperception with its reference to our consciousness of following rules. So, if this consciousness were itself dependent on a prior determination of time, we could not conceive of ourselves as spontaneous in the determination of the transcendental conditions of whatever we determine a priori, that is, in achieving the necessary relation between the understanding and appearances. This would be contrary to the whole spirit of Kant’s philosophy of autonomy and immanence in the field of theoretical knowledge.

19. AkXVIII319, lines 24-8, my translation and additions: AK. Cf. Allison (1983, p.277) for his use of this text.
These reflections lead one to the recognition of a second point that is fundamental in Kant’s philosophy, viz. the importance of the first-person perspective. As we have already seen in various connections, Kant gives great importance to our conception of ourselves as subjects. This involves the claim that ‘I’ - this simple and completely empty representation - is crucial in ascribing to ourselves the self-consciousness that makes us subjects. This is why for Kant this consciousness and, consequently, ‘I’ do not distinguish any particular object (cf. A346/B404), but are the form to be used to assert that I ‘represent’ something. This is to say that it is only through it that one can claim to be thinking of something (cf. loc. cit.). We have also seen what role this faculty plays in Kant: it is only through its mediation that we can claim to be following rules, insofar as it allows us to have the consciousness of so doing. Put in other terms, this is ‘the subjective I’ that has to be presupposed in all thinking (cf. A354). Kant calls this the ‘I’ that thinks and has to be regarded always as subject (cf. B407).

Nothing could be more distant from Kant’s position than to suggest that his insistence on the first-person perspective leads to solipsism and scepticism about other minds. Kant never bothered to argue explicitly against this form of scepticism, but his views offer enough material to at least show that he did not think that his philosophy would lead to it. The intelligible features of the transcendental subject were always conceived of as features of every intelligent being, and the ones Kant claimed to know were we humans. In A354 Kant talks about how I can represent to myself a thinking being, and he says that ‘I must put myself in his place, and thus substitute, as it were, my own subject for the object I am seeking to consider’. That is, I must consider this object as using ‘I’ for himself. This is not an appeal to analogy. This is only a reference to the ascription of an intelligible character to another possible thinking being, which has always to be made problematically because the intelligible is not what appears or is manifest to the senses as such, but rather what is dependent on the performance-aspect of what we perceive. So, if such a being uses ‘I’ in the expression of self-consciousness, then this is what is important for Kant, and there is no speculation about the literally inner of such a being. This is why Kant also says that we can intuit the signs of thoughts in the field of appearances only (cf. A359), because it is men that think (cf. A360).

With this point Kant refers to the constitutive transcendental conditions of all representing or meaning. Objectivity about them does not depend on an a posteriori
agreement between subjects, as if each of them had to consult and express to the benefit of others how she sees things. According to Allison, and we have seen that he is correct in this, objectivity is in Kant a definitional feature of judgments in what concerns their meaning. This is because they will be judgements only if they comply with the condition of the unity of apperception as it is set out in paragraph 19 of the Deduction B. To this requirement I have given an externalist construal, showing how, for example, the transcendental syntheses of imagination, by being techniques for doing things according to rules, require the transcendental consciousness of the following of the rules. This transcendental condition is in fact the basis for objectivity and consequently intersubjectivity, and not vice-versa. The following remarks by Kant, where a reference to the object is given a normative character, illustrate the point by stating what ‘truth’ means, and not how truth can be established.

Persuasion is a mere illusion, because the ground of the judgement, which lies solely in the subject, is regarded as objective. Such a judgement has only private validity, and the holding of it to be true does not allow of being communicated [Kant means shared]. But truth depends upon agreement with the object, and in respect of it the judgement of each and every understanding must therefore be in agreement with each other (consentientia uni tertio, consentiunt inter se). The touchstone whereby we decide whether our holding a thing to be true is conviction [when it rests on objective grounds] or mere persuasion is therefore external, namely, the possibility of communicating it and of finding it to be valid for all human reason. For there is then at least a presumption that the ground of the agreement of all judgements with each other, notwithstanding the differing characters of individuals, rests upon the common ground, namely, upon the object, and that it is for this reason that they are in agreement with the object - the truth of the judgement being thereby proved (A820-1/B848-9; my italics and additions: AK).

This quotation makes it clear that ‘agreement with the object’ comes first, being then eventually followed by intersubjectivity. Now, the ‘third thing’ mentioned in the Latin phrase, which makes the agreement of the understandings possible, is something external. This can only be a reference to the requirement that the object be spatial - distinct from the understandings of the subjects - and consequently the ‘common ground’ which can be ‘communicated’ and ‘found valid for all human reason’. Such claims cannot be accounted for except by externalism20.

Furthermore, Kant never saw the errors we commit in philosophizing as due to a psychological subjectivism; instead he always insisted that they are not errors of men but of

reason itself insofar as transcendental conditions are misconstrued, which means that these
errors spring from the very nature of reason itself (cf.A339/B397). This is also in perfect
agreement with Kant’s metaphilosophical reflections about philosophy as a discipline.

The importance of Kant’s conception of ourselves as corporeal agents in the world,
who nonetheless give an intelligible aspect to their actions, and his acute awareness of the
possibilities and limits of philosophy lead him to a further feature of this intellectual
enterprise we are engaged in. This refers to a certain limitation in transcendental
investigations, and is the third and last element I want to present as distinctive of Kant’s
critical position.

Kant considers the possession of the categories to be a fact. However, this factual
claim is not so straightforward as it appears. Kant envisages an argument aimed to show that
there are these concepts, that is, an argument to show that they exist. So, Kant offers to
show how we can ‘discover’ the categories. This argument is called by him ‘metaphysical
deduction’ (cf. B159). He says that it proves ‘the a priori origins of the categories... through
their complete agreement with the general logical functions of thought’ (ibidem). I would
like to present Kant’s views concerning this issue with the intention of stressing his
reflections on the factual claim from a transcendental perspective.

Where the argument is to be found (A70/B95-A83, B116), it is not so much a proof
of a priori origin that we get; what is presented is rather the identification of the categories
Kant said existed. Kant, as a matter of fact, does take as data for this identification the forms
of judgement which are the subject-matter of General Logic. With this ‘clue’, however,
what Kant really intends is to present the ‘categories’ on a systematic basis, that is, in a
complete catalogue (cf. A69/B93-4; A80-1/B106-7)\(^{21}\). Kant’s claim here are admitted to be,
even by the most charitable commentators of his views, unwarranted\(^{22}\). I will not deal with
these claims by Kant here, nor with the actual details of the derivation of the categories from
the forms of judgement. Let us see what Kant says about the fact we are concerned with.

\(^{21}\) For the purpose of the Metaphysical Deduction cf. Allison (1983), p.115 and ff., and Johnson

\(^{22}\) Cf. Allison (1983), pp.128-9, especially note 40 (pp.350-1).
When we concentrate on the factual claim, we find that Kant regards our possession of the functions or forms of judgement described by General Logic as a fact of which any further explanation is impossible:

This peculiarity of our understanding, that it can only produce \textit{a priori} unity of apperception solely by means of the categories, and only by such and so many, is as little capable of further explanation as why we have just these and no other functions of judgement, or why space and time are the only forms or our possible intuition (B145-6).

In a letter to Marcus Herz, Kant again states that the existence of the functions of the understanding is a fact. He remarks to his addressee that someone (a Mr. Maimon) wants to know from him 'how can I even prove the necessity of these functions of the understanding whose existence is again [Kant had asserted that the agreement between \textit{a priori} intuitions and \textit{a priori} concepts is ‘given as a fact’] merely a fact'. Later on, he precludes any insight into the origin of the faculties of sense and understanding when we have in mind the necessary agreement between the two:

If we wanted to make judgements about their origin [sense and understanding] - an investigation that of course lies wholly beyond the limits of human reason - we could name nothing beyond our divine creator; once they are given, however, we are fully able to explain their power of making \textit{a priori} judgements (that is, the \textit{quid juris})\textsuperscript{23}.

I think that there can be no doubt about this being Kant’s considered opinion on the matter. As a consequence, any project of providing a metaphysical deduction of, as it were, the logical functions themselves, perhaps from a single principle or a superior faculty, must be ill-motivated for Kant. Whether and to which extent this constitutes a stumbling-block on the way of Kant’s intentions of presenting a complete list of categories is a worry that I am not able deal with here. However, it seems fair to say that Kant’s factual claim goes against, in spirit, his wishes for completion and systematization.

The points made above about the categories and the forms of judgements are connected with a further factual claim by Kant, \textit{viz.} that our understanding is discursive. And in a similar fashion, the basis for this is the observation of our activity of judging. The point I want to stress concerns the status of these claims. They are considered to be ultimate in our transcendental reflections, and Kant never speculates about ‘the reason’ for them. The

\textsuperscript{23} Both quotes are from Kant’s letter to Marcus Herz (May 26, 1789), \textit{in:} \textit{Kant: Philosophical Correspondence 1759-99}, pp.152, 154 respectively. Cf. also \textit{Prolegomena}, #36, pp.Ak318-20, for similar points.
implicit suggestion is that it is not possible to go *philosophically* behind these data. In parallel with these points about our understanding, Kant asserts an ultimate datum about our intuitive nature: 'how in a thinking subject outer intuition... is possible... this is a question which no man can possible answer. This gap in our knowledge can never be filled' (A393). This is a type of question on which Kant observes: 'Now on speculative principles no one can give the least ground for any such assertion. Even the possibility of what is asserted cannot be established; it can only be assumed' (A394)\(^2^4\). This is also why according to their nature philosophical claims have a limit, they reflect in a sense the fact that in philosophy trying to say something, to characterize things further, can lead us astray. A *transcendental* claim concerning the simple *in the concept* reveals in Kant the following limitation: 'But if I know something as simple in the concept only and not in the [field of] appearance, I have really no knowledge whatsoever of the object, but only of the concept which I make for myself of a something in general that does not allow of being intuited. I say that I think something as completely simple, only because I have really nothing more to say of it than merely that it is a something' (A400). I think this has to do with our *finitude* and the *ultimate contingency* which qualifies the accord between our understanding and nature. This should obviously not imperil the Kantian account of the necessary relation between understanding and appearances. It is rather a point about the amount of realism we should subscribe to. An intuitive understanding would not be so limited, but 'our understanding is a faculty of conceptions. This means that it is a discursive understanding for which the character and variety to be found in the particular given to it in nature and capable of being brought under its conceptions must certainly be contingent'\(^2^5\).

The three points I have stressed show clearly that Kant’s perspective on matters of philosophical reflection, his ‘Copernican Revolution’, was ‘anthropocentric’; unlike the ‘theocentric’ one, held by his predecessors as well as by many of his successors\(^2^6\), who tried to view things *sub specie aeternitatis*. As we have seen, however, there is nothing to be

\(^{24}\) For a diametrically opposite view, cf. Locke (E, p.440), where the possibility of a *real* definition of man (that would explain the source of all our ultimate faculties) is asserted and at the same time put beyond our reach.

\(^{25}\) *Critique of Judgement*, #77, p.62.

sorry about, because our finitude is also the source of our autonomy, which in the end makes nature and freedom possible (cf.A543/B571).
BIBLIOGRAPHY

WORKS BY KANT


Kant: Pre-Critical Writings and Correspondence with Beck (1968), Manchester University Press, Manchester; G.B. Kerferd and D.E. Walford (tr.).

Kant: Philosophical Correspondence, 1759-99 (1970), The University of Chicago Press, Chicago; Arnulf Zweig (ed. and tr.).

Critique of Pure Reason (1933), Macmillan, London; Norman Kemp Smith (tr.).

Kritik der Reinen Vernunft (1993), Felix Meiner Verlag, Hamburg; Raymund Schmidt (ed.).


The Critique of Judgement (1992), Clarendon Press, Oxford; J.C. Meredith (tr.).


Logic [Jäsche] (1974), Bobbs-Merrill, Indianapolis; R.S. Hartman and W. Schwarz (tr.).

What Real Progress Has Metaphysics Made in Germany since the Time of Leibniz and Wolff? (1983), Abaris Books, New York; Ted Humphrey (tr.).

Anthropology from a Pragmatic Point of View (1974), Martinus Nijhoff, The Hague; Mary Gregor (tr.).


WORKS BY OTHERS


