

KANT'S DEDUCTION OF THE CATEGORIES

Robert Watt

Balliol College, Oxford

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Philosophy

Total word count: 74,405

Michaelmas Term 2013

Table of Contents

Introduction	p.4
§1 The Objective Validity of the Categories	p.25
§1.1 Objective validity and Kant's concept of representation	p.27
§1.2 Empiricist and rationalist theories of the Categories	p.40
§1.3 Kant's theory of the Categories	p.50
§1.4 Dickerson on Kant's concept of representation	p.61
§2 Kant's Concept of Empirical Cognition	p.72
§2.1 The principle of Kant's Deduction of the Categories	p.74
§2.2 Empirical cognition and the act of judgment	p.90
§2.3 Kitcher on Kant's concept of empirical cognition	p.100
§2.4 Empirical cognition and empirical knowledge	p.106
§3 Kant's Principle of the Original Synthetic Unity of Apperception	p.113
§3.1 The unity of apperception and the act of judgment	p.116
§3.2 Kant's concept of original apperception	p.135
§3.3 Kant's concept of an act of synthesis	p.150
§3.4 The unity of apperception and the Categories	p.168

§4	The Objective Validity of Judgment and the Original Synthetic Unity of Apperception	p.182
§4.1	Objective validity and necessary, universal validity	p.184
§4.2	Necessary, universal validity and the unity of apperception	p.198
§4.3	Other theories of judgment	p.214
§4.4	Kitcher on empirical cognition and the unity of apperception	p.225
	Conclusion	p.239
	Bibliography	p.245

Introduction

At the heart of Immanuel Kant's *Critique of Pure Reason* is the famous argument that he calls his Deduction of the Pure Concepts of the Understanding (A84/B116, Axvi), or his Transcendental Deduction of the Categories (cf. A92/B124, A128). This argument has intrigued and infuriated students of Kant's work more than any other part of the *Critique*. On the one hand, it is an argument that we clearly need to understand if we want to understand Kant's theory of knowledge. As one commentator has remarked, it is in his Deduction of the Categories that Kant

has formulated his most profound thoughts and presented the decisive foundation for his theory of knowledge. Whoever understands these pages possesses a key to the understanding and evaluation of the entire work.
(Henrich 1969, p.640)

On the other hand, the argument is famously obscure. It is dense and abstract. It is peppered with obscure technical terms, few of which are clearly defined and most of which are used inconsistently. It is organized in such a way as to make it hard for the reader to distinguish the introduction of a new idea from the reformulation of an old idea or to see how all of the different ideas and lines of argument are related to one another. Several commentators have compared it to a jungle (Bennett 1966, p.100, Strawson 1966, p.93, Van Cleve 1999, p.79), and one has gone as far as to write:

The crossing of the Great Arabian Desert can scarcely be a more exhausting task than is the attempt to master the windings and twistings of the Transcendental Deduction. (Paton 1936, p.547)

The aim of this thesis is to explain how Kant's Deduction of the Categories should be understood. I believe that many of the claims that have been made about what is good and what is bad in this argument have been based on false assumptions about its aims and its structure. I also believe that there are important ideas in Kant's Deduction of the Categories that have escaped the notice of commentators and that deserve serious critical attention. Finally, I believe that there is reason to care about improving our understanding of the value of Kant's argument even apart from mere historical curiosity. Kant's Deduction of the Categories is an original and ambitious attempt to make progress on philosophical questions that are as pressing today as they were in the late 18th Century. By clarifying and evaluating Kant's answers to these questions we may reasonably hope to find ourselves in a better position to answer them for ourselves.

I shall start by saying something about the basic themes of Kant's Deduction of the Categories. Kant's aim in his *Critique of Pure Reason* as a whole is to answer a question about the possibility of a particular science - namely, metaphysics. Kant thinks that this science is similar in at least two important respects to the mathematical sciences. First, he thinks that the science of metaphysics and the mathematical sciences are species of he calls '*a priori* cognition', as opposed to 'empirical cognition' (e.g. B2). That is to say, he thinks that they are based on 'pure reason' as opposed to 'experience'

(Ak IV:265-266, B2). The distinction that Kant has in mind here has to do with the different ways in which our conclusions about the world can be justified.

It is often the case that we reach conclusions about the world that are justified on the basis of what we have seen or touched or in some other way experienced. I can reach the conclusion that there is a cup of tea on the desk in front of me. And I can reach the conclusion that the sun will rise tomorrow. The first of these conclusions is justified on the basis that I can see a cup of tea on the desk in front of me. The second is justified on the basis that I have seen that the sun has risen on every day of my life so far. And there are many areas of scientific enquiry where our conclusions are justified in exactly the same way. For instance, if a medical scientist reaches the conclusion that a particular treatment counteracts the effects of a particular disease, her conclusion will be justified on the basis of her observation of what has happened to patients who have been prescribed this treatment. However, there are other areas of scientific enquiry where our conclusions are not justified in this way. They are not justified empirically, which is to say, on the basis of what we have seen or touched or in any way experienced. This is what Kant thinks is true of the mathematical sciences. I can reach the conclusion that the square on the hypotenuse of a right angled triangle is equal to the sum of the squares on the other two sides. And I can reach the conclusion that $7 + 5 = 12$. But these conclusions are not justified on the basis of what I have seen or touched or in any way experienced. It is not because I have seen numerous right angled triangles in the course of my life and found that all of them are such that the square on their hypotenuse is equal to the sum of the squares on the other two sides that I am entitled to conclude that Pythagoras' theorem is true. However, this conclusion is still

justified, which means that it must be justified in some other way. And Kant thinks that this is true of our conclusions in metaphysics as well (Ak IV:265).

The second respect in which Kant thinks that metaphysics is similar to the mathematical sciences is insofar as its conclusions are 'synthetic', or 'ampliative', as opposed to merely 'analytic', or 'explicative' (Ak IV:266, A6-7/B10-11). It is a controversial question how exactly this claim should be understood. Loosely, what Kant has in mind is the idea that when we draw conclusions in the mathematical sciences, what happens is not just that we become conscious of a fact about the world that we already knew, but rather that we come to know a fact about the world that we did not know before. Conclusions in the mathematical sciences are discoveries or additions to what we know (Ak IV:267). They are not mere clarifications of what we knew already, which is to say, additions to what we know clearly. Kant's favourite example of an analytic or explicative conclusion is the claim that all bodies are extended. His favourite example of a synthetic or ampliative conclusion is the claim that all bodies are heavy. Suppose that someone possesses the concept of a body. Suppose that she knows what a body is. And suppose that she subsequently reaches the conclusion that all bodies are extended. This is not a discovery on her part. She already knew that all bodies are extended. For she already had the concept of a body. And part of what it is to have the concept of a body is to know that all bodies are extended. What has happened is that she has become conscious of a fact that she already knew. Suppose instead that this person has reached the conclusion that all bodies are heavy. This could well be a discovery. For this person could possess the concept of a body and yet fail to realize that all bodies are heavy. To know what a body is it is not necessary to know that all bodies are heavy. Kant thinks that the conclusions

of the mathematical sciences are synthetic in this sense. He doesn't think that what happens when we reach the conclusion that the square on the hypotenuse of a right angled triangle is equal to the sum of the squares on the other two sides is that we become conscious of a fact that we already knew. For it is no part of our concept of a right angled triangle that it has this property. I can know what a right angled triangle is and yet fail to know that all right angled triangles have this property. And once again Kant thinks that the same is true of the conclusions reached in metaphysics.

In summary, when Kant refers to the science of metaphysics, what he has in mind is a sort of science where we reach conclusions that are synthetic or ampliative and in which these conclusions are not justified on the basis of what we have seen or touched or in any way experienced. But there is one more point that needs to be made. Kant thinks that the mathematical sciences are synthetic and *a priori*, yet he thinks that they are also distinct from metaphysics. So we are left with the question: How does the science of metaphysics differ from the mathematical sciences? Kant's answer to this question is that whereas metaphysical conclusions are reached 'from concepts', the conclusions of the mathematical sciences are reached by means of the 'construction of concepts' (Ak IV:272, A713/B741). What he means by this is that there is a special operation of the mind that is required for us to reach conclusions in mathematics. And he defines this special operation of the mind as the *a priori* exhibition of the intuition corresponding to a concept (A713/B741). Now it is not at all easy to understand what Kant means by this 'construction of concepts', or *a priori* exhibition of the intuition corresponding to this concept. And the question how this special operation of the mind should be understood lies far beyond the scope of this thesis. For my purposes it will be sufficient to note that what Kant means by the science of

metaphysics is that science other than the mathematical sciences where we reach conclusions that are not justified on the basis of what we have seen or touched or in any way experienced, and that are also synthetic in the sense that I just explained.

It will be helpful at this point to have some examples of claims made in the course of metaphysics. One example that Kant likes to use is the claim that the world had a beginning in time – that is to say, the proposition that there was a time at which the world started to exist (e.g. B18). Another example is the Law of Causality, which is to say, the claim that everything that starts to exist or ceases to exist in time has a cause. A third example is the claim that the world is composed only of simple entities – that is to say, indivisible entities or atoms – and the composite objects that are made up of these entities. Other examples are claims about the properties of the soul and its relation to the body – say, the proposition that the soul is simple or that it does not interact with the body, or the proposition that the soul is material, or that it is immaterial.

Kant thinks that there is such a science as metaphysics. He thinks that it is possible for us to reach conclusions about the world that are justified in spite of the fact that they are not justified on the basis of what we have seen or touched or in any way experienced, and that amount to additions to what we know, not just additions to what we know clearly, and that, finally, are not conclusions of mathematics. However, Kant also thinks that there is a problem that must be confronted by anyone who thinks there is such a science as metaphysics. Someone who is committed to the possibility of this science needs to be able to address the concerns of someone who cannot see how a science of this sort is possible and who concludes on this basis that it is not in fact possible. The trouble is that if we confine our attention to those conclusions that we

reach that add to what we know and not just to what we know clearly, it is hard to see how conclusions of this sort could be justified except on the basis of what we have seen or touched or in some way experienced. Take the metaphysical conclusion that the world started to exist at some point in time. Suppose that this conclusion is justified, and that it is justified in spite of the fact that it is not justified on the basis of what we have seen or touched or in any way experienced. It is hard to see how this could be the case unless this conclusion was reached on the basis of mere analysis, in which case it would not constitute an addition to what we know, only to what we know clearly. However, it is implausible to claim that it is any part of our concept of the world that it has started to exist at some point in time. Suppose then that this conclusion constitutes an addition to what we know. In that case it is hard to see how we could reach such a conclusion except on the basis of what we have seen or touched or in some way experienced. But if this is how we have reached the conclusion then it is not a metaphysical conclusion at all.

The problem here rests on the assumption that a conclusion can be justified *a priori* – that is to say, it can be justified independent of experience – only if it adds only to what we know clearly, or, to put this the other way around, that if it is not the case that a conclusion adds only to what we know clearly, it is not the case that this conclusion can be justified *a priori*. If we accept this assumption then we will be driven to the conclusion that the science of metaphysics is impossible. And we may think that there is corroboration for this conclusion. The impossibility of metaphysics as a science may appear to provide the best explanation of the lack of progress in the science of metaphysics over the last two and a half thousand years.

The question of the possibility of the science of metaphysics is the basic philosophical question that Kant has in mind in his Deduction of the Categories. He makes this clear in the Preface to the first edition of the *Critique*, where he writes:

I am acquainted with no investigations more important for getting to the bottom of that faculty that we call the understanding... than those that I have undertaken.... under the title Deduction of the Pure Concepts of the Understanding... This inquiry... has two sides. One side refers to the objects of the pure understanding, and is supposed to demonstrate and make comprehensible the objective validity of its *a priori* concepts; thus it belongs essentially to my ends. The other side deals with the pure understanding itself... and although this exposition is of great importance in respect of my chief end, it does not belong essentially to it, because the chief question always remains: What and how much can understanding and reason cognize free of all experience? and not: How is the faculty of thinking itself possible? (Axvi-xvii)

There is a lot that calls for clarification in this passage. For now, the crucial point is Kant's claim that what is essential to his Deduction of the Categories is the part that helps him to attempt to answer the question: What is it possible for us to know independent of experience – that is to say, by means of pure reason? While Kant doesn't take care at this point to emphasize that what really matters to him is the possibility of a particular sort of *a priori* science, viz. the science of metaphysics, as opposed to the mathematical sciences, it is reasonable to assume that this is what he has

in mind. For earlier on in the Preface of the first edition of the *Critique* he has made this quite clear. He has told us that what he means by a critique of pure reason is ‘the decision about the possibility or impossibility of a metaphysics in general’ (Axvii).

The lines that I just quoted also make it clear that there is a second question that Kant has in mind in his Deduction of the Categories. This is a question about our faculty of thinking, which is another name for what Kant calls our faculty of understanding. Specifically, it is the question how we are able to engage in acts of thinking, which is to say, the question what conditions must be satisfied in order for us to make use of our faculty of understanding. In the course of this thesis I shall defend an interpretation of Kant’s Deduction of the Categories on which it can be understood as an attempt to answer precisely this question.

It should be noted that there is another more familiar philosophical question that is also often associated with Kant’s Deduction of the Categories. This question has to do with the existence of an external world, which is to say, the existence of a world of objects distinct from the states of our minds. Suppose that I come to the conclusion that there is a cup of tea on the desk in front of me or that the apple in my hand is round. These conclusions commit me to the existence of external objects. They commit me to the belief that there are cups of tea and apples in the world, as well as mere ideas of cups of tea and apples or desires for cups of tea and apples. Some philosophers have maintained that all of my conclusions about the existence of objects in the external world are false. They have insisted that there are in fact no cups of tea or apples in the world. Other philosophers have taken a more moderate stance. They have claimed that all of our conclusions about the existence of objects in the external world are unjustified. They have claimed that we can never really know whether or not

there is an external world. Now one of the most famous questions in philosophy – or, at any rate, one of the most famous questions in modern philosophy – is the question how we can justify our conclusions about the existence of objects in the external world. And many commentators on the Deduction of the Categories have assumed that this famous question about the existence of the external world is at least one of the basic questions that Kant is trying to answer. As Ameriks puts it, these commentators have treated the Deduction of the Categories

as basically aiming to establish objectivity, i.e. to prove that there is an external and at least partially lawful world, a set of items distinct from one's awareness, and to do this from the minimal premise that one is self-conscious. (1978, p.276-277)

One recent commentator who reads Kant's argument in this way is Van Cleve (1999). He writes that in the course of the Deduction of the Categories Kant defends the claim that 'the validity of the Categories enables our representations to be parlayed into knowledge of objects distinct from ourselves' (1999, p.73). He adds that 'this additional thesis may furnish us with (or at least make an important contribution toward) a refutation of skepticism about the external world' (1999, p.73, cf. p.76).

One of the aims of this thesis is to show that this way of understanding Kant's Deduction of the Categories is without foundation in the text of the *Critique*. At no point does Kant tell us that it is any part of his aim in the Deduction of the Categories to prove that there really is an external world, that is to say, a world of apples and cups, chairs and tables etc. (cf. Ameriks 1978, p.280). And if we look carefully at what Kant

actually does in the Deduction of the Categories then we find no evidence that this is even part of what he wants to achieve.

It often seems as if this way of understanding Kant's Deduction of the Categories is based on false assumptions about the connections between Kant's claims and the claims made by other philosophers. Take Van Cleve as an example. He remarks in the course of his discussion of the Deduction of the Categories that we may hope that a particular interpretation is the correct one '[insofar] as we look to Kant for a satisfactory reply to Hume's skepticism about the external world' (1999, p.74). What is striking about this remark is the association of Hume with the idea that our conclusions about the existence of objects in the external world are unjustified. The reason that this is striking is that Hume is never associated with this idea by Kant himself. There is in the second edition of Kant's *Critique of Pure Reason* a clearly marked refutation of the doctrine that the external world, which is to say, the world of apples and books, chairs and tables, does not exist – or, at any rate, cannot be proved to exist. Kant calls this doctrine 'empirical idealism' (A491/B519), and he calls the part of the *Critique* where he sets out to refute it the Refutation of Idealism (B274). However, Kant does not mention Hume's name in the Refutation of Idealism or in any of his other discussions of empirical idealism. The philosophers that he associates with empirical idealism are Berkeley and the Eleatics (B274, Ak IV:374). When Kant refers to Hume it is almost always in connection with Hume's claims about the concept of cause and effect (B5, Ak IV:257), or with his scepticism about the possibility of the science of metaphysics (B9, A760/B788, Ak IV:258).

This provides us with one plausible explanation of the fact that Van Cleve associates Kant's Deduction of the Categories with the question as to the existence of

the external world. Van Cleve notices that the Deduction of the Categories is designed to address a sort of scepticism championed by Hume. And since Van Cleve associates Hume with scepticism about the existence of the external world, he assumes that this must be the sort of scepticism that Kant has in mind in his Deduction of the Categories. But this is just a mistake. The sort of scepticism that the Deduction of the Categories is designed to address – and that Kant associates with Hume – is scepticism about a particular sort of *a priori* science, viz. the science of metaphysics.

Another explanation of the fact that so many commentators have treated Kant's Deduction of the Categories as if it is designed to prove the existence of an external world has to do with the central claim that Kant makes in the course his argument, viz. the claim that he calls the Principle of the Original Synthetic Unity of Apperception. This principle appears to be a claim about our consciousness of our self. Many commentators have interpreted it as the claim that we are able to achieve a special sort of consciousness of our self – or, alternatively, that we must be able to achieve a special sort of consciousness of our self if we are to become conscious of anything at all. And it is plausible to suppose that what has led so many commentators to read Kant's Deduction of the Categories as a rejoinder to the sceptic about the external world is the assumption that if the argument starts from a claim to the effect that we are or must be conscious of our self, its intended conclusion must be a claim to the effect that we are conscious of objects other than our self – that is to say, a claim to the effect that we are conscious of an external world (cf. Ameriks 1978, p.283). There would be an obvious historical precedent for this sort of argument. Descartes' line of argument in his *Meditations* starts from a claim about our consciousness of our self – or to be precise, from a claim to the effect that we have a special sort of knowledge of the existence of

our self – and builds on this basis a proof of, among other things, the existence of an external world of apples, books, chairs and tables etc.

In the course of this thesis I defend a very different interpretation of Kant's Principle of the Original Synthetic Unity of Apperception. I argue that this principle is not simply a claim to the effect that we achieve or are able to achieve a special sort of conscious of our self, nor is it the claim that we must be able to achieve a special sort of consciousness of our self if we are to achieve any sort of consciousness at all. Rather, it is a claim to the effect that we must be able to achieve a special sort of consciousness if we are to be able to make use of our faculty of thinking or understanding. The Principle of the Original Synthetic Unity of Apperception is an answer to the second of the two questions that Kant mentions at the end of the passage that I quoted from the Preface to the first edition of the *Critique*. I also argue in this thesis that what Kant deduces from the Principle of the Original Synthetic Unity of Apperception is not the claim that there must be an external world – that is to say, a world of apples and books, chairs and tables etc. – but rather a claim about the role in our thinking of special concepts, viz. the concepts that Kant calls the Categories or Pure Concepts of the Understanding.

So far in this introduction I have tried to clarify the underlying themes of Kant's Deduction of the Categories. In the rest of this introduction I shall describe the structure of the thesis and the contents of the individual sections. The thesis is divided into four chapters. The first chapter is concerned with the aim of Kant's Deduction of the Categories. I have already quoted some lines from the Preface to the first edition of the *Critique* where Kant writes that this argument – or, to be precise, the part of the argument that 'belongs essentially to [his] ends' (Axvii) – is intended 'to demonstrate

and to make understandable the objective validity of [the] *a priori* concepts [of the pure understanding]' (Axvi). In the first section of the first chapter of this thesis, I address the question what exactly it means to demonstrate and make understandable the objective validity of the Categories or Pure Concepts of the Understanding. I argue that it is not sufficient to demonstrate the objective validity of these concepts to show that there are in fact objects that fall under these concepts. I argue instead that the question as to the objective validity of the Categories is the question whether, and in what way, these concepts in our minds are able to serve as representations of objects. Naturally, this raises the question of what exactly it means to say that a concept serves as a representation of an object. I argue that Kant thinks that a state of our mind serves as a representation of an object insofar it stands in a very special relation to an object, viz. a relation of necessary agreement or conformity. I show that Kant's aim in his Deduction of the Categories is to show that the Categories or Pure Concepts of the Understanding stand in this relation of necessary agreement or conformity to objects.

In the second section of the first chapter, I consider some of the possible ways of explaining the necessary agreement or conformity of the Categories with their objects. My aim here is to shed further light on the idea of a relation of necessary conformity and to substantiate my claim that it is a relation of necessary conformity between the Categories and their objects that is the focus of Kant's Deduction of the Categories. I examine a number of theories of the objective validity of the Categories to which Kant refers in his famous letter of 21st February 1772 to Marcus Herz. These include the theory of hyperphysical influence that he ascribes to Plato and Malebranche and the theory of pre-established intellectual harmony that he ascribes to Crusius. In the third section of the first chapter, I address the crucial question how Kant himself

intends to explain the relation of necessary agreement or conformity between the Categories and objects. I argue that Kant's explanation turns on the claim that it is necessary for an object to conform to the Categories or Pure Concepts of the Understanding if it is to count as an object of experience. This amounts to the claim that there is a relation of necessary conformity between the Categories and objects of experience – and, therefore, to the claim that the Categories serve as representations of objects of experience. In the fourth and final section of the first chapter of this thesis, I discuss a rival interpretation of Kant's concept of representation recently defended by Dickerson (2004).

In the second chapter of this thesis I focus on the general strategy of Kant's Deduction of the Categories. Kant thinks that in order for an object to count as an object of experience – that is to say, an object of empirical cognition – it must be both an object of intuition – specifically, an object of empirical intuition – and also an object of thought. In the first section of the second chapter, I show that what Kant thinks we need to do in order to show that the Categories serve as representations of objects is to show that an object must conform to the Categories if it is to be an object of thought – that is to say, if it is to be an object of our faculty of understanding. Kant thinks that the Categories serve as conditions of the possibility of empirical cognition precisely because he thinks that they serve as conditions of the thought of an object. And so the strategy of his Deduction of the Categories is to investigate the conditions that must be satisfied if we are to think of an object. I defend this claim against the objection that what Kant actually does in his Deduction of the Categories is to show that the objective validity of the Categories is required for even the mere intuition or mere consciousness of an object. In the second section of the second chapter, I turn to the question what

exactly Kant thinks it means to think of an object – that is to say, what he thinks it means to make use of our faculty of understanding. I argue that according to Kant a person has made use of his faculty of understanding, or has thought of an object, insofar as he has made a judgment about an object – for instance, the judgment that the cup on the desk in front of me is round, or the judgment that there are clouds over Oxford. I defend this interpretation of Kant’s concept of the thought of an object against an important objection based on an observation about Kant’s concept of a state of cognition. In the third section of the second chapter, I consider a very different interpretation of Kant’s concept of empirical cognition recently defended by Kitcher (2011). Kitcher argues that someone has achieved a state of empirical cognition in the sense that Kant has in mind in his Deduction of the Categories insofar as she has made a judgment about an object that is rational in the sense that she is aware of her reasons for making this judgment. And Kitcher argues that Kant’s strategy in his Deduction of the Categories is to investigate the conditions that must be satisfied if someone is to make a judgment about an object that is rational in this sense. I show that Kitcher’s argument is based on a questionable interpretation of Kant’s claims about the role of concepts as ‘marks’, or ‘grounds of cognition’. Finally, in the fourth section of the second chapter of this thesis, I consider the claim that what Kant means by a state of empirical cognition is a state of empirical knowledge. I examine two objections against this claim. The first turns on Kant’s distinction between the different ways of holding something to be true in the Transcendental Doctrine of Method. The second, due to Van Cleve (1999), turns on the implications of this claim for Kant’s definition of truth at the start of the Transcendental Logic.

In the third chapter of this thesis I focus on the pivotal claim in Kant's Deduction of the Categories. This is the claim that Kant calls the Principle of the Original Synthetic Unity of Apperception. In the first section of the chapter, I argue that this principle is a claim about the conditions that must be satisfied if we are to make a judgment about an object – or, in other words, a claim about the conditions that must be satisfied if we are to make use of our faculty of understanding. I argue that this principle is the claim that if we are to make use of our faculty of understanding then those conditions must be satisfied that must be satisfied if we are to achieve what Kant calls the original synthetic unity of apperception. If this is correct then the Principle of the Original Synthetic Unity of Apperception is not simply the claim that we have achieved or are able to achieve a special sort of consciousness of our self. Nor is it the claim that we must be able to achieve a special sort of consciousness of our self if we are to achieve any sort of consciousness at all. I defend this interpretation of Kant's Principle of the Original Synthetic Unity of Apperception on the basis of what he writes in §17 and §19 of the second edition Deduction of the Categories. I also explain how this interpretation is consistent with Kant's more famous formulation of his principle in §16 of the second edition Deduction of the Categories.

In the rest of the third chapter of this thesis, I set out to explain what exactly Kant means by the original synthetic unity of apperception. I start in the second section of the chapter by examining Kant's concept of apperception in general, and his concept of original apperception in particular. I argue that what Kant means by apperception is the consciousness of a state of mind as one that serves as a representation of an object. In this I follow Dickerson (2004). However, I go further

than Dickerson in arguing that what Kant means by original apperception is the consciousness of a state of mind as one that serves as a representation of an object in a special way – namely, in virtue of the fact that the object conforms to the state of mind and not the other way around. In the third section of the third chapter, I turn to Kant’s concept of an act of synthesis. I argue that what Kant means by an act of synthesis is an act of the mind by means of which several of our concepts or judgments become connected to one another precisely in virtue of the fact that the mind has formed a representation of a connection between these concepts or judgments. I should mention that my interpretation of both Kant’s concept of original apperception and his concept of synthesis turn on my interpretation of Kant’s concept of representation in the first chapter of this thesis. In the fourth and final section of the third chapter, I use my interpretation of Kant’s concept of original apperception and his concept of an act of synthesis to complete my interpretation of his concept of the original synthetic unity of apperception. I argue that what Kant means by the Principle of the Original Synthetic Unity of Apperception is the claim that in order for someone to make use of his faculty of understanding, he must not only connect several of his representations in an act of synthesis, but also become conscious – by means of his faculty of original apperception – that part of what he has done is to perform an act of synthesis that is necessary in the sense that it is performed whenever any other act of synthesis is performed. To put this point another way, Kant thinks that someone who makes a judgment about an object must be conscious that he has connected his representations in accordance with a necessary rule of synthesis. At the end of the fourth section of the third chapter, I explain the final stages in Kant’s Deduction of the Categories. I argue that there is a short path between the Principle of the Original

Synthetic Unity of Apperception and the claim that the use of our faculty of understanding requires Categories or Pure Concepts of the Understanding. The key claim is that what Kant thinks a Pure Concept of the Understanding consists in is the consciousness of an act of synthesis that is necessary in the sense that I just described.

The fourth chapter of this thesis is also the most original. It goes to the heart of the question whether there is anything of value in Kant's attempt to trace the connection between the use of our faculty of understanding and the Categories or Pure Concepts of the Understanding. My aim in this fourth chapter is to examine the considerations on the basis of which Kant accepts the Principle of the Original Synthetic Unity of Apperception. In the course of my discussion I make extensive use of the passage of Kant's *Prolegomena* where he introduces his famous distinction between objectively valid and merely subjectively valid empirical judgments. I argue in the first section of this chapter that one of the pillars upon which Kant's acceptance of the Principle of the Original Synthetic Unity of Apperception rests is his belief that a person makes use of his faculty of understanding only if he represents a connection between his representations as necessarily, universally valid, which is to say, only if he thinks that these representations ought to be connected in this way in every consciousness whatsoever. In the second section of the chapter, I argue that the other pillar upon which Kant's acceptance of the Principle of the Original Synthetic Unity of Apperception rests is his belief that someone can represent a connection between his representations as necessarily, universally valid in the sense I just described only if he is conscious of the necessary unity of his consciousness – or, equivalently, only if he is conscious of the necessary unity of the synthesis of his representations. The idea that I ascribe to Kant in the first two sections of this chapter can be sketched as follows: Kant

thinks that when a person makes use of his faculty of understanding, part of what he thinks is that there is a requirement for his consciousness of the various connections between his representations to exhibit a sort of coherence. However, Kant also thinks that someone can think of this requirement of coherence only if he is conscious of the fact that his consciousness has a sort of unity, which is to say, loosely, that this consciousness constitutes a sort of whole. And Kant thinks that someone can become conscious that his consciousness has this sort of unity only if he is conscious that all of his consciousness of the various connections between his representations is structured in the same way.

In the third section of the fourth chapter of this thesis, I compare Kant's theory of judgment to the theories defended by some other 18th and 19th Century philosophers. I discuss Hume's theory that the difference between entertaining the thought there is a cup on the desk in front of me and making the judgment that there is a cup on the desk in front of me has to do with the force or vivacity of our representations. I also consider the theory that this difference has to do with the different effects of these states of our mind on our actions and feelings, and the theory that it has to do with the question whether or not the state of mind is the result of the tendency of our mind to form associations between representations. I argue that Kant's theory of judgment has a major advantage over these other theories in that it recognizes the distinctively normative component of an act of judgment. It involves the recognition that in judging that, say, there is a cup of tea on the desk in front of me, I require that others should not form a contrary judgment. Finally, in the fourth section of the fourth chapter, I examine Kitcher's (2011) recent attempt to explain Kant's

theory of the connection between a state of empirical cognition and the original synthetic unity of apperception.

§1 The Objective Validity of the Categories

It is clear from the text of the *Critique* that Kant's aim in his Deduction of the Categories is to explain how it is that the Categories or Pure Concepts of the Understanding are able to count as objectively valid. More precisely, Kant's aim is to address the concerns of someone who cannot see how the Categories could possibly count as objectively valid, and who concludes on this basis that the Categories are not objectively valid. My aim in this chapter is to explain as clearly as I can what Kant means by the objective validity of the Categories. In the course of my discussion I consider a number of different theories about the objective validity of the Categories, including the one that Kant defends in the Deduction of the Categories itself.

I defend two key claims. The first is that a concept counts as objectively valid in the sense that Kant has in mind in his Deduction of the Categories insofar as it serves as a representation of an object. This is a controversial claim. Recent commentators have tended to interpret Kant's concept of objective validity in terms of instantiation. Some have assumed that a concept counts as objectively valid in the sense that Kant has in mind insofar as there is at least one object of experience that falls under this concept (e.g. Van Cleve 1999), while others have assumed that a concept counts as objectively valid insofar as all objects of experience fall under this concept (e.g. Guyer 1992). I show that these interpretations are mistaken. If we look carefully at what Kant thinks is required for us to explain how the Categories are able to count as objectively valid, we will see that there must be more to the objective validity of a concept than mere instantiation.

If my first claim is correct then we can draw a further conclusion: If we want to understand Kant's concept of objective validity, we need to understand his concept of representation. We need to know what Kant thinks is required for a concept to serve as a representation of an object. The second claim that I defend in this chapter is that a concept serves as a representation of an object in the sense that Kant has in mind in his Deduction of the Categories insofar as two conditions are satisfied. The first condition is that there is a relation of agreement or conformity between the concept and the object, and the second condition is that it is no accident that there is this relation of agreement or conformity between the concept and the object. Since this is what Kant thinks is required for a concept to serve as a representation of an object, his aim in his Deduction of the Categories is to show not only that there are objects that stand in a relation of conformity to the Categories – that is to say, objects that fall under the Pure Concepts of the Understanding – but also that this relation of conformity is no accident. Kant wants to explain why it is necessary for certain objects to stand in a relation of conformity to the Categories.

At the start of §1.1, I focus on the Categories or Pure Concepts of the Understanding themselves. These are the concepts of unity, plurality and totality, reality, negation and limitation, inherence, causality and community, possibility, actuality and necessity. I explain briefly how each of these concepts should be understood, and I explain how Kant thinks that these concepts differ from other less abstract concepts such as the concept of roundness or the concept of a cup of tea. In the rest of §1.1, I focus on Kant's concept of objective validity. I use the famous letter that Kant wrote to his friend Marcus Herz on 21st February 1772 to show, first, that a concept counts as objectively valid insofar as it serves as a representation of an object,

and second, that a concept serves as a representation of an object insofar as it stands in a relation of necessary conformity to an object. My aim in §1.2 is to explore some of the ways in which Kant thinks that a concept can stand in a relation of necessary conformity to an object. To this end, I consider three explanations of the objective validity of the Categories that Kant dismisses in his letter to Herz. One of these explanations is a familiar empiricist theory of the Categories. The other two theories are what Kant calls the theory of hyper-physical influence and the theory of pre-established intellectual harmony. In §1.3, I introduce Kant's theory of the objective validity of the Categories that Kant defends in the *Critique* itself. And I show that this theory can also be understood as an attempt to explain how the Categories are able to serve as representations of objects in the sense I have discussed in §1.1 and §1.2. Finally, in §1.4, I consider an alternative interpretation of Kant's concept of representation recently defended by Dickerson (2004).

§1.1 Objective validity and Kant's concept of representation

When Kant refers to the Categories or Pure Concepts of the Understanding, what he has in mind are the concepts that we cannot avoid using if we want to think about objects in the world. To think that there are clouds over Oxford, I must make use of at least two concepts. I must make use of my concept of a cloud and I must make use of my concept of Oxford. But there is no need for me to make use of these concepts if I want to think that the apple in my hand is round or that the sun will rise tomorrow in the East. I can very easily think about objects in the world without making use of my concept of Oxford or my concept of a cloud. However, Kant thinks that there are some

other concepts of which I must make use if I want to have any thoughts at all about the objects that I find in the world (e.g. A80/B106). These concepts are the '*subjective conditions of thinking*' (A89/B122), the 'concepts of an object in general' (B128). And they are the concepts that Kant calls the Categories or Pure Concepts of the Understanding.

Kant thinks that there are exactly twelve of these Categories (A80/B106, Ak IV:303), and he divides them into Categories of Quality, Categories of Quantity, Categories of Relation and Categories of Modality. The Categories of Quality are the concepts of Reality, Negation and Limitation and the Categories of Quantity are the concepts of Unity, Plurality and Totality. The Categories of Relation are the concepts of Inherence, Causality and Community, and the Categories of Modality are the concepts of Possibility, Actuality and Necessity. When Kant claims that I cannot avoid using these concepts if I want to think about objects in the world, he doesn't mean that every thought about objects in the world requires the use of all twelve concepts. His claim is rather that every thought about objects in the world requires the use of at least one Category of Quality, at least one Category of Quantity, at least one Category of Relation and at least one Category of Modality.

It will be helpful at this point briefly to consider how each of the twelve Categories should be understood. The Category of Reality and the Category of Negation are respectively the concept of a presence and the concept of a lack or an absence (A291/B347). Darkness and cold count as negations since the concept of darkness is the concept of a mere absence of light and the concept of cold is the concept of a mere absence of heat (A291/B347, A575/B603). Heat and light count as realities since the concept of heat is not the concept of a lack of cold, nor is the concept

of light the concept of a lack of darkness. The Category of Limitation is the concept of a presence that is limited in some way by an absence or an absence that is limited in some way by a presence (Ak XXIX:998). One of Kant's examples of a limitation is shade. The idea here is that the concept of shade is the concept of an area of darkness that is limited in extent by an area of light.

The Category of Unity is the concept of an individual. Socrates counts as a unity since it is part of the concept of Socrates that he is one object and not many objects. The Category of Plurality is the concept of some but not all of the objects of a particular sort. If I think of some of the objects that count as apples or of three of the objects that count as cups of tea then I make use of the Category of Plurality. As for the Category of Totality, this is the concept of all of the objects of a particular sort. One of the objects that Kant thinks counts as a totality is nature. He defines nature in the *Prolegomena* and in his *Metaphysical Foundations of Natural Science* as the sum total of all objects of experience (Ak IV:467, Ak IV:295).

The Category of Inherence is the concept of the relation between a substance and one of its accidents (A80/B106), where a substance is an object that has properties but that is not itself a property of any object (B129, Ak IV:335) and an accident is a property of a substance (A186/B229). Heat inheres in a stone that has been left out in the sunshine, since the heat is a property of the stone but the stone is not itself a property of any other object. The Category of Causality is the concept of the relation between a cause and its effect (A80/B106), where one object counts as the cause of another object insofar as the existence of the second object is a consequence of the existence of the first object. The sunshine is the cause of the heat in the stone since the fact that there is heat in the stone is a consequence of the fact that there is sunshine.

As for the Category of Community, this is the concept of a set of substances where the state of each substance – that is to say, the sum total of the properties of each substance – is a consequence of the states of all of the other substances (A221/B269, B257-258).

The Category of Possibility is the concept of what could exist. My concept of the victory of the England cricket team in the forthcoming Ashes series is the concept of a possibility. The Category of Actuality is the concept of what does in fact exist (A225/B272). My concept of the roundness of the apple in my hand is the concept of an actuality. Finally, the Category of Necessity is the concept of what must exist (A226-227/B279). If I think that the stone has been left out in the sunshine and that it is a consequence of this that the stone will shortly become hot, my concept of the hotness of the stone is the concept of a necessity (cf. A227/B279-280).

Having clarified each of the twelve Categories, I turn to the Deduction of the Categories itself. It sometimes looks as if the aim of Kant's Deduction of the Categories is just to prove that the Categories or Pure Concepts of the Understanding have a particular property – namely, objective validity. Kant tells us that the Deduction of the Categories is a 'justification of [the] objective validity [of the Categories]' (A96). Its aim is to 'confirm' (A91/B123), or to 'demonstrate' (A90/B122), that the Categories are objectively valid. But there is also evidence that the aim of Kant's Deduction of the Categories is to explain the fact that the Categories are objectively valid. Kant tells us that the aim of his Deduction of the Categories – or, at any rate, the aim of the most important aspect of the Deduction of the Categories – is 'to make comprehensible the objective validity of [the] pure *a priori* concepts [of the pure understanding]' (A128). Kant defines a transcendental deduction in general as 'the explanation of the way in which concepts can relate to objects *a priori*' (A85/B117). And he claims in the

Deduction of the Categories that he has deduced ‘the concepts of space and time’, which ‘relate to objects completely *a priori*’ (A85/B118), because he has ‘explained and determined their *a priori* objective validity’ (A87/B119-120).

The reason why Kant defines his Deduction of the Categories in terms of both demonstration and explanation is that his aim to address the concerns of someone who doubts that the Categories are objectively valid precisely because he cannot see how concepts of this sort could possibly count as objectively valid. This sceptic about the objective validity of the Categories makes two assumptions. First, he assumes that a concept is able to count as objectively valid only if it is a concept of a particular sort, viz. one that is empirical, as opposed to pure *a priori*. Second, he assumes that the Categories are pure *a priori* concepts. And he concludes on this basis that the Categories are not objectively valid. Kant’s aim in his Deduction of the Categories is to undermine the first of the two assumptions on the basis of which the sceptic reaches his conclusion. Kant wants to explain how the Categories are able to count as objectively valid in spite of the fact that they are pure *a priori* concepts – and, in this way, to undermine the sceptical conclusion that the Categories are not in fact objectively valid. Put simply, in order to defend the claim that the Categories or Pure Concepts of the Understanding are objectively valid, Kant must explain how it is that they are able to count as objectively valid.

I come now to the crucial question: What exactly does it mean to say that a concept is objectively valid in the sense that Kant has in mind? As I mentioned in the introduction to this chapter, Kant’s concept of objective validity is often interpreted in terms of instantiation. Van Cleve thinks that a concept counts as objectively valid in the sense that Kant has in mind insofar as it applies to objects of experience – that is to

say, insofar as there is at least one object of experience that falls under this concept (1999, p.89). He writes that a concept counts as objectively valid insofar as it is 'instantiated', or 'exemplified in the world we experience' (1999, p.73). If this interpretation is correct then the concept of a cup of tea is objectively valid if and only if the objects of experience include at least one cup of tea, and the Categories are objectively valid if and only if at least one of the objects of experience is a reality, at least one of the objects of experience stands in a relation of causality, etc.

There is some textual evidence to support this interpretation. Kant writes for example:

I take e.g. the concept of cause... It is not clear *a priori* why appearances should contain anything of this sort (one cannot adduce experiences for the proof, for the objective validity of this concept must be able to be demonstrated *a priori*), and it is therefore *a priori* doubtful whether such a concept is not perhaps entirely empty and finds no object anywhere among the appearances. (A90/B122)

Here it seems that what Kant means by a demonstration of the objective validity of the concept of a cause is just a demonstration that the objects of experience include an object of this sort – that is to say, at least one cause. But if this is correct then Kant must think that a concept counts as objectively valid insofar as there is at least one object that falls under this concept.

Another commentator who interprets Kant's concept of objective validity in terms of instantiation is Guyer (1992, p.125). Whereas Van Cleve thinks that a

concept counts as objectively valid insofar as it applies to at least one object of experience, Guyer thinks that a concept counts as objectively valid insofar as it applies to all objects of experience – that is to say, insofar as every object of experience falls under this concept. He writes that

a concept has objective reality if it has at least some instantiation in experience but objective validity only if it applies to all possible objects of experience. (1992, p.125)

If Guyer is correct then the Categories count as objectively valid insofar as every object of experience is a unity, a plurality and a totality, stands in a relation of inherence, causality and community, etc. By contrast, the concept of a cup of tea does not count as objectively valid since not every object of experience is a cup of tea. The trouble with Guyer's interpretation of Kant's concept of objective validity is that it is simply implausible to claim that every object of experience falls under each of the twelve Categories. Take a particular instance of heat – say, the heat in a stone that has been left out in the sun. This is surely an object of experience. And Kant clearly thinks that it falls under the Category of Reality. It is a presence as opposed to a mere lack or absence. But in that case it cannot also fall under the Category of Negation. Similarly, Socrates is an object of experience. And he falls under the Category of Unity. However, he does not fall under the Category of Plurality. The concept of Socrates is not the concept of some but not all of the objects of a particular sort.

There is also a serious objection to Van Cleve's interpretation of Kant's concept of objective validity. Remember that Kant's aim in his Deduction of the Categories is

to explain how it is that the Categories or Pure Concepts of the Understanding are able to count as objectively valid. If Van Cleve's interpretation is correct then what this means is that Kant's aim is to explain how there can be objects of experience that fall under the Categories. He wants to explain how there can be unities and pluralities, relations of inherence and causality, etc. in the world of experience. Kant's aim is to address the concerns of someone who denies that there are unities and pluralities, relations of inherence and causality etc., in the world of experience – and, crucially, who denies this because he cannot see how there could possibly be unities and pluralities, relations of inherence and causality etc., in the world of experience. The trouble with Van Cleve's interpretation of Kant's concept of objective validity is that it is implausible to claim that what Kant is doing in his Deduction of the Categories is addressing the concerns of someone who cannot see how there could possibly be unities and pluralities, relations of inherence and causality etc., in the world of experience.

It is helpful at this point to compare Kant's Deduction of the Categories to his refutation of the theory that he calls dogmatic idealism. A dogmatic idealist is someone who denies the existence of objects in space – someone who denies the existence of apples and books, chairs and tables etc. – and who denies this precisely because he thinks that objects of this sort are impossible (A377). Kant thinks that the way to refute this sort of idealism is to undermine the idealist's argument for the claim that objects in space are impossible. So if Van Cleve is right about Kant's concept of objective validity, Kant's aim in his refutation of dogmatic idealism is very similar to his aim in his Deduction of the Categories. In the first case, Kant wants to explain how there can be objects in space – and, in this way, to undermine the claim that there are

in fact no objects in space. In the second case, Kant wants to explain how there can be unities and pluralities, relations of causality and inherence etc. – and, in this way, to undermine the claim that there are in fact no unities and pluralities, relations of causality and inherence etc. Now, Kant leaves us in no doubt about the reason why the dogmatic idealist thinks that objects in space are impossible. He writes that someone can be a dogmatic idealist ‘only because [she] believes [she] can find contradictions in the possibility of matter in general’ (A377). The dogmatic idealist thinks that we can prove a contradiction from the assumption that objects in space exist. And he concludes on this basis that objects in space cannot exist, from which it also follows that they do not exist. In Kant’s *Deduction of the Categories*, by contrast, there is no reference whatsoever to any argument for the claim unities and pluralities, relations of inherence and causality etc. are impossible. Kant doesn’t even mention the idea that the *Categories* or *Pure Concepts of the Understanding* might be contradictory. It is reasonable to conclude that his aim in this part of the *Critique* is not to explain how objects that fall under the *Categories* – that is to say, unities and pluralities, relations of inherence and causality etc. – can possibly exist.

So how should we understand Kant’s concept of the objective validity of the *Categories*? The answer to this question becomes clear if we consider the famous letter that he wrote to his friend Marcus Herz on 21st February 1772. This is the letter where Kant announces his discovery of ‘the key to the whole secret of metaphysics’ (Ak X:130). He writes:

I asked myself this question: What is the ground of the relation of that in us which we call ‘representation’ to the object? If a representation

comprises only the manner in which the subject is affected by the object then it is easy to see how it is in conformity with this object, namely, as an effect accords with its cause, and it is easy to see how this modification of our mind can represent something, that is, have an object. Thus the passive or sensuous representations have an understandable relation to objects. (Ak X:130)

Kant starts from the observation that there are states of our mind that count as representations of objects. The question that he raises is this: How is it that these states of our mind are able to count as representations of objects? Kant thinks that this question is easy to answer in the case of representations that are the effects of objects upon the mind, but difficult to answer in the case of representations that are not related to objects in this way. He continues:

In my dissertation I was content to explain the nature of intellectual representations in a merely negative way, namely, to state that they were not modifications of the soul brought about by the object. However, I silently passed over the further question of how a representation that refers to an object without being in any way affected by it can be possible... [If] intellectual representations depend on our inner activity, whence comes the agreement that they are supposed to have with objects...? (Ak X:130-131)

This last question is what Kant has in mind when he refers to 'the key to the whole secret of metaphysics' (Ak X:130). He also expresses this question as follows:

[As] to how my understanding may, completely *a priori*, form for itself concepts of things with which concepts the facts should necessarily agree... this question, of how the faculty of understanding achieves this conformity with the things themselves, is still left in a state of obscurity. (Ak X:131)

This is clearly a question about what Kant calls in the *Critique* the Pure Concepts of the Understanding or the Categories. It is the question how the Categories or Pure Concepts of the Understanding are able to serve as representations of objects. And I propose that this is also precisely the question that Kant sets out to answer in his Deduction of the Categories. If this is correct then we are entitled to draw an important conclusion: What Kant means by the question how it is that the Categories are able to count as objectively valid is the question how it is that the Categories are able to serve as representations of objects. Kant thinks that a concept counts as objectively valid insofar as it serves as a representation of an object. And his aim in the Deduction of the Categories is to address the concerns of someone who assumes that an *a priori* concept cannot serve as a representation of an object, and who concludes on this basis that the Categories or Pure Concepts of the Understanding are not in fact representations of objects.

There is another important conclusion that we can draw from Kant's letter to Herz. Notice the assumptions that Kant makes in the lines that I have quoted from this letter about the conditions that must be satisfied by a state of mind if it is to count as a representation of an object. First, Kant clearly thinks that there must be some sort of relation of agreement or conformity between the state of mind and the object. He tells

us that it is easy to see how the ‘sensuous representations’ serve as representations of objects because it is easy to see how they are ‘in conformity with’ their objects (Ak X:130). By contrast, it is hard to see how the Pure Concepts of the Understanding serve as representations of objects because it is hard to see where ‘the agreement that they are supposed to have with objects’ comes from (Ak X:130-131). Second, Kant clearly thinks that this agreement or conformity between the state of mind and the object must be necessary. He thinks that the relation between the state of mind and the object must be grounded on something (Ak X:130). It can be no accident that the state of mind and the object stand in a relation of conformity to one another. There must be some connection between the state of mind and the object that ensures that they agree with one another.

These two assumptions about what is required for a state of mind to serve as a representation of an object are also clearly present when Kant discusses the objective validity of the Categories in the *Critique* itself. For example, Kant expresses the claim that there are two ways in which concepts can be related to objects of experience as follows:

there are only two ways in which a necessary agreement of experience with the concepts of its objects can be thought: either the experience makes these concepts possible or these concepts make the experience possible.
(B166-167 cf. Ak IV:319)

This makes it reasonable to conclude that a state of mind counts as a representation of an object in the sense that Kant has in mind in his Deduction of the Categories only if

two conditions are satisfied. First, the state of mind and the object must stand in a relation of agreement or conformity to one another. Second, it must be no accident that the state of mind and the object stand in this relation of agreement or conformity to one another.

It isn't hard to think of objects that serve as representations of other objects in virtue of the fact that they stand in a relation of necessary conformity to those objects. Consider the Tube Map. Part of the reason why the Tube Map serves as a representation of the network of lines and stations on the London Underground is that the map and its object stand in a relation of conformity to one another. Just as there is no direct line between Piccadilly Circus and Westminster stations, so there is no direct line on the map between the node marked Piccadilly Circus and the node marked Westminster. But it also matters that this relation of conformity between the Tube Map and the network of lines and stations on the London Underground is no accident. It matters that, for example, it is only because there is no direct line between Piccadilly Circus and Westminster stations that there is no direct line on the map between the nodes marked Piccadilly Circus and Westminster. This is part of the reason why the Tube Map serves as a representation of the network of lines and stations London Underground. Suppose that someone on a distant planet draws an abstract picture that just happens perfectly to resemble the Tube Map. This picture stands in the same relation of conformity to the London Underground network as the Tube Map itself. However, unlike the Tube Map, it does not serve as a representation of the London Underground network. For the conformity of the picture to the London Underground network is purely accidental (cf. Putnam 1981, p.1). The Tube Map counts as a representation of its object in virtue of the fact that there is a relation of necessary

conformity between the lines and nodes on this map and the network of lines and stations on the London Underground.

If what I have claimed so far in this chapter is correct then we can say a little more about the mistake that Van Cleve makes in his interpretation of Kant's concept of objective validity. Van Cleve assumes that it is sufficient for one of the Pure Concepts of the Understanding to count as objectively valid that there is some object in the world of experience that falls under this Category. If some object of experience is a unity, it follows that the Category of Unity is objectively valid. But this is not what Kant understands by objective validity. He thinks that the Category of Unity is objectively valid only if it serves as a representation of an object in the world. And it is not sufficient for the Category of Unity to serve as a representation of an object in the world that some object in the world falls under the concept of unity, any more than it is sufficient for, say, a picture of a woman to represent the Queen of England that the woman in the picture closely resembles the Queen of England. If it is entirely accidental that there is this relation of resemblance between the woman in the picture and the Queen of England, the picture does not count a representation of, or depiction of, the Queen of England. Similarly, if it is entirely accidental that there are objects in the world that fall under the Category of Unity, this concept in my mind does not count as a representation of these objects.

§1.2 Empiricist and rationalist theories of the Categories

So far in this chapter, I have defended an interpretation of Kant's concept of objective validity, and I have defended an interpretation of Kant's concept of representation. I

have claimed that a concept counts as objectively valid in the sense that Kant has in mind in his Deduction of the Categories insofar as it serves as a representation of an object. And I have claimed that a concept counts as a representation of an object in the sense that Kant has in mind insofar as, first, it stands in a relation of agreement or conformity to an object, and second, it is no accident that it stands in this relation of agreement or conformity to an object. I have concluded on this basis that Kant's aim in his Deduction of the Categories is to explain how it is that the Categories or Pure Concepts of the Understanding – that is to say, the Categories of Unity, Plurality and Totality, Inherence, Causality and Community, etc. – are able to stand in a relation of necessary conformity to objects.

In the course of his letter of 21st February 1772 to Marcus Herz, Kant not only indicates the question that he thinks constitutes 'the key to the whole secret of metaphysics' (Ak X:130), viz. the question how it is that the Pure Concepts of the Understanding are able to count as objectively valid, but also criticizes three answers to this question. My aim in this section is to consider each of these answers in turn. The point of this is to clarify and lend further support to the interpretation of Kant's concept of objective validity that I defended in §1.1. If we look carefully at the explanations of the objective validity of the Pure Concepts of the Understanding that Kant rejects in his letter to Herz, we find that they all consist in a claim about what makes it necessary for these concepts to stand in a relation of conformity to objects.

One explanation of the objective validity of the Categories that Kant clearly rejects in his letter to Herz is that the Categories represent objects in the same way that the Tube Map represents the network of lines and stations on the London Underground, or in the same way that a portrait of Queen Elizabeth II represents the

Queen Herself. The Tube Map serves as a representation of its object because it is a consequence of the fact that the lines and stations on the London Underground network are related in the way that they are that the lines and nodes on the Tube Map are related in the same way. It is because there is no direct line between Piccadilly Circus and Westminster stations on the London Underground network that there is no direct line on the map between the nodes marked Piccadilly Circus and Westminster. But Kant doesn't think that it is because there is unity and plurality, inherence and causality etc. in the world that we have the Categories of Unity and Plurality, Inherence and Causality, etc. He doesn't think that the fact that we have the Categories of Unity and Plurality, Inherence and Causality etc. is a consequence of the fact that there is unity, plurality, inherence, causality, etc. in the world. If the Categories served as representations of objects in this way then they would be what Kant calls empirical concepts. But Kant is adamant that the Categories are not empirical concepts. That is why he calls them Pure Concepts of the Understanding.

Why is Kant so sure that this is not the correct explanation of the objective validity of the Categories? The reason has to do with Hume's famous discussion of causality. Kant takes Hume to have established that what accounts for the fact that we have the concept of causality is not the fact that there are relations of cause and effect in the world. The argument for this conclusion turns on the claim that the concept of causality includes the representation of a necessary connection between two objects - or, to be clear, the claim that part of what it is to think of a relation of cause and effect is to think of a necessary connection between two objects. Hume thinks that one object counts as the cause of another object only if the existence of the first object makes it necessary for the second object to exist. As Kant puts it,

this concept [i.e. the concept of a cause] always requires that something A be of such a kind that something else B follows from it necessarily and in accordance with an absolutely universal rule. (A91/B124, see also Ak IV:257)

So if the concept of causality is to count as an empirical concept, there must be necessary connections between objects in the world, and it must be a consequence of the fact that there are necessary connections between objects in the world that we have a concept that includes the representation of a necessary connection. Now our senses are the capacities by means of which objects in the world cause us to have representations. So, loosely, if it is a consequence of the fact that there are necessary connections between objects in the world that we have a concept that includes the representation of necessary connections, necessary connections between objects in the world must impress themselves upon our senses. We must see or feel or in some other way perceive such connections. The trouble is that we never see or feel or in any way perceive necessary connections between objects in the world. Both Hume and Kant conclude on this basis that the concept of causality is not an empirical concept after all (cf. A91/B124).

Kant thinks that a similar point can be made in the case of each of the twelve Categories. He thinks each of these concepts includes the representation of a necessary connection. As he writes in the *Prolegomena*,

the concept of the connection of cause and effect is far from being the only

concept through which the understanding thinks connections of things *a priori* [i.e. thinks of a necessary connection]... (Ak IV:260)

Since Kant accepts Hume's claim that we never see or feel or in any way perceive any necessary connections between objects in the world, he concludes that it cannot be the fact there are unities and pluralities, relations of inherence and causality, etc. in the world that accounts for the fact that we have the Categories of Unity and Plurality, Inherence and Causality, etc. And this is why he thinks that the Categories of Unity and Plurality, Inherence and Causality etc. are Pure Concepts of the Understanding. This is why he rejects the empiricist theory that these concepts serve as representations of objects in the same way that the Tube Map serves as a representation of the network of lines and stations on the London Underground, or in the same way that a portrait of the Queen of England serves as a representation of the Queen herself.

In the course of his letter to Herz, Kant considers and rejects two other theories of the objective validity of the Categories. On one of these other theories, the connection between the Categories and their objects that accounts for the conformity between the two is simply the reverse of the connection that is supposed to exist on the empiricist theory that I have already discussed. Kant writes:

if that in us which we call 'representation' were active with regard to the object, that is, if the object itself were created by the representation (as when divine cognitions are conceived as the archetypes of things), the conformity of these representations to their objects could be understood. Thus the possibility of... an intellectus archetypus... is at least

comprehensible. (Ak X:130)

Suppose that I draw up the blueprint for a new house. And suppose that I pass on this blueprint to an engineer who builds a house in accordance with my plan. There is a sense in which the blueprint that I have drawn serves as a representation of the house that the engineer has constructed. And the fact that the blueprint counts as a representation of the house is not due simply to the fact that the house that the engineer has constructed has the same properties as the house depicted on the blueprint. After all, if there is a qualitatively identical house built by some other engineer, this other house is not represented by the blueprint that I have drawn. What makes it the case that my blueprint represents the house that my engineer has constructed is the fact that it is precisely because I indicated on the plan that the house should take a particular form that the house that the engineer has constructed takes this form. Now suppose that a state of mind serves as a blueprint for an object in the same way that my plan of the house serves as the blueprint for the house itself. Suppose that it is precisely because my state of mind takes a particular form that the object itself takes this particular form. In that case, my state of mind serves as a representation of the object in the same way that the blueprint serves as a representation of the house. I take it that this is what Kant means when he writes that ‘if that in us which we call ‘representation’ were active with regard to the object... the conformity of these representations could be understood’ (Ak X:130).

Kant refers in the lines that I have just quoted from Kant’s letter to Herz to the theory that God’s concepts serve as representations of objects because they serve as the blueprints for – or, to put this point another way, the archetypes of – those objects.

This is the theory defended by Aquinas, who claims that God's concepts serve as representations of objects in the world not in virtue of the fact that they are caused by those objects but rather in virtue of the fact they cause those objects – or, to be precise, in virtue of the fact that they cause those objects in conjunction with God's will. Now in the course of his letter to Herz, Kant considers the possibility that this theory about the relation of God's concepts to objects also allows us to understand how it is that the Categories – that is to say, the pure concepts of our faculty of understanding – are able to serve as representations of objects in spite of the fact that they are not empirical concepts. He writes:

Plato assumed a previous intuition of divinity as the primary source of the pure concepts of the understanding and of first principles. Malebranche believed in a still-continuing, perennial intuition of this primary being. (Ak X:131)

On each version of this theory, the Categories serve as representations of objects because they are in fact concepts in God's mind, and God's concepts serve as representations of objects in virtue of the fact that they are the archetypes of those objects. The two versions of the theory differ with respect to the question how it is that we are able to become conscious of God's concepts. On the version of the theory that Kant ascribes to Plato, we become conscious of these concepts by means of recollection. The idea here is that there was a time in the distant past when our minds were united with God's. At this time, God created the world in conformity with the concepts in His mind. Since then our minds have been divided from God's, but we can still call

back to consciousness the concepts in conformity with which the world was created. And so our concepts serve as representations of objects in virtue of the fact they served as the archetypes of those objects. On the version of the theory that Kant ascribes to Malebranche, by contrast, we become conscious of God's concepts simply because God allows us to see into His mind.

Kant's objections to this theory are the same as his objections to a third theory of the Categories, which I should also briefly mention. This theory turns on the concept of pre-established harmony. Kant writes that

Crusius believed in certain implanted rules for the purpose of forming judgments and ready-made concepts that God implanted in the human soul just as they had to be in order to harmonize with things. (Ak X:131)

The idea here is that the Categories serve as representations of objects in virtue of the fact that God creates both the Categories – that is to say, special concepts in our mind – and also the objects in the world, and ensures that the concepts and the objects conform to one another. In his letter to Herz, Kant calls this third theory the theory of pre-established intellectual harmony. He calls the second theory, viz. the one that he ascribes to Plato and Malebranche, the theory of hyper-physical influence (Ak X:131).

Kant raises two objections to these rationalist theories of the Categories in his letter to Herz:

the deus ex machina is the greatest absurdity one could hit upon in the

determination of the origin and validity of our cognitions. It has, besides its vicious circularity in drawing conclusions concerning our cognitions, also this additional disadvantage: it encourages all sorts of wild notions and every pious and speculative brainstorm. (Ak X:131)

The first objection that Kant mentions here can be restated as follows: We cannot appeal to the existence of God in the course of a proof that the Categories are objectively valid. For we can appeal to the existence of God in the course of such a proof only if our concept of God is objectively valid. Yet our concept of God is objectively valid only if the Categories are objectively valid. Perhaps this is because the concept of God is the concept of a particular cause – say, the concept of the cause of order in the world or the concept of the first or ultimate cause. In any case, if we appeal to the existence of God in the course of a proof that the Categories are objectively valid then we reason in exactly the same sort of circle as Descartes when he appeals to the existence and benevolence of God in the course of a proof that whatever we clearly and distinctly perceive is true.

Kant's second objection to the theories of hyper-physical influence and pre-established intellectual harmony is expressed a little more clearly in a footnote to the *Prolegomena*. Here Kant ascribes to Crusius the claim that 'a spirit who can neither err nor deceive originally implanted... natural laws in us' (Ak IV:319), and then writes:

But, since false principles are often mixed in as well – of which [Crusius'] system itself provides not a few examples – then with the lack of sure criteria for distinguishing an authentic origin from a spurious one, the use

of such a principle looks very precarious... (Ak IV:319)

Suppose that someone claims to have discovered a concept – say, the concept of Doomsday – that serves as a representation of an object in spite of the fact that it is not an empirical concept. If we use either the theory of hyper-physical influence or the theory of pre-established intellectual harmony to explain how it is that a pure concept can serve as a representation of an object then we have no reliable way to assess the claim that this person's concept really does represent an object. Had this person claimed that his concept was empirical, we could test this claim to objective validity by reference to the source of the concept. We could ask to be shown what the person had seen or touched or in some way perceived that had led him to form this concept. But if the theory of hyper-physical influence or the theory of pre-established harmony is correct then we have no such test available to us. We have no way to tell for sure whether or not the concept of Doomsday is one that God has implanted in this person's mind, or whether it is in fact a concept in the mind of God.

This concludes my survey of the three theories of the Categories that Kant considers and rejects in his letter to Herz. Each of these theories is an attempt to explain how it is that the Categories are able to count as objectively valid. Crucially, on each of these theories there is a relation of agreement or conformity between the Categories and objects in the world. Moreover, on each of these theories there is something that ensures that there is this relation of agreement or conformity between the Categories and objects in the world. There is some connection between the way that our mind is determined and the way that the objects in the world are determined. The three theories differ with respect to the nature of this connection. On the first

theory, which is to say, the empiricist theory, the way that our mind is determined – that is to say, the fact that we have the Categories of Unity and Plurality, Inherence and Causality, etc. – is a consequence of the way that the objects in the world are determined. It is only because there are unities and pluralities, relations of inherence and causality, etc. that we have the Categories of Unity and Plurality, Inherence and Causality, etc. On the second theory, the theory of hyper-physical influence, the way that the objects in the world are determined is a consequence of the way that our mind is determined. What ensures that there is a relation of conformity between the objects and the Categories is the fact that the Categories are part of the cause of the objects. On the third theory, what ensures that there is a relation of conformity between the objects and the Categories is the fact that they have a cause in common. It is God who makes it the case that the way that our mind is determined corresponds to the way that the objects in the world are determined.

This survey of the theories of the objective validity of the Categories that Kant criticizes in his letter to Herz is helpful for two reasons. It helps us to obtain a clearer picture of the sort of relation between the Categories and objects that is of interest to Kant in his Deduction of the Categories. It sheds further light on the relation of necessary conformity that is the key to Kant's concept of representation. Moreover, it provides further support for the interpretation of Kant's concept of objective validity that I defended in §1.1. It is difficult to see how the theories that I have considered in this section can plausibly be understood as attempts to answer the question how the Categories are able to count as objectively valid if Kant's concept of objective validity is interpreted in terms of mere instantiation.

§1.3 Kant's theory of the Categories

At this point I turn to the question: What exactly is Kant's own theory of the objective validity of the Categories? How does Kant intend to explain the fact that the Categories count as objectively valid in spite of the fact that they are Pure Concepts of the Understanding? Consider these famous lines from the Preface to the second edition of the *Critique*:

Up to now it has been assumed that all our cognition must conform to the objects; but all attempts to find out something about them *a priori*... have, on this presupposition, come to nothing. Hence let us once try whether we do not get further with the problems of metaphysics by assuming that the objects must conform to our cognition... This would be just like the first thoughts of Copernicus, who, when he did not make good progress in the explanation of the celestial motions if he assumed that the entire celestial host revolves around observer, tried to see if he might not have greater success if he made the observer revolve and left the stars at rest. (Bxvi)

As I explained in the introduction to this thesis, Kant thinks that the science of metaphysics is an *a priori* science. He thinks, however, that there can be no science of this sort if all of the states of our mind that serve as representations of objects do so in virtue of the fact that they conform to the objects. Kant therefore suggests that some of the states of our mind that serve as representations of objects do so in virtue of the fact that the objects conform to the states of mind. In particular, he proposes that there are

both intuitions and concepts in our minds that represent objects in this way. He continues:

I can assume either that the concepts through which I [determine objects] also conform to the objects, and then I am once again in the same difficulty about how I could know anything about them *a priori*, or else I assume that the objects, or what is the same thing, the experience in which alone they can be cognized (as given objects) conforms to those concepts, in which case I immediately see an easier way out of the difficulty... (Bxvii)

I take it that the concepts that Kant has in mind here are the Categories or Pure Concepts of the Understanding. Kant thinks that it is precisely because objects conform to the Categories that the Categories serve as representations of objects. It is not because the Categories conform to the objects.

But what does it mean to say that the Categories serve as representations of objects in virtue of the fact that the objects conform to them? It looks at first as if what Kant has in mind here is precisely the sort of theory that he ascribes to Plato and Malebranche in his letter to Herz – that is to say, the theory that the Categories serve as representations of objects because they are the blueprints for those objects. When Kant refers to the idea that a state of mind serves as a representation of an object in virtue of the fact that the state of mind conforms to an object, what he has in mind is presumably the empiricist theory on which it is precisely because the object is determined in a particular way that the state of our mind is determined in this way. If this is correct then it is reasonable to assume that what Kant has in mind when he

refers to the idea that a state of mind serves as a representation of an object in virtue of the fact that the object conforms to the state of mind is the theory that simply reverses the connection between the objects and the concepts. And in Kant's letter to Herz that is the theory that the Categories serve as representations of objects in the way that God's concepts serve as representations of objects – that is to say, because they are the archetypes of those concepts. Objects can be said to conform to God's concepts in the sense that it is only because God conceives of objects as round or square or solid that objects are in fact round or square or solid. So it seems that Kant's own answer to the question how it is that the Categories are able to serve as representations of objects – and, therefore, as objectively valid – is that they serve as the blueprints for those objects.

This impression is reinforced when we consider those places in the *Critique* itself, and also in the *Prolegomena*, where Kant contrasts his own theory about the objective validity of the Categories with other theories (B166-167, Ak IV:319). He tends to mention two alternatives, viz. the empiricist theory and the theory of pre-established intellectual harmony. Now, as I pointed out in §1.2, in his letter to Herz Kant also mentions three theories about the objective validity of the Categories. These are the empiricist theory, the theory of pre-established harmony, and the theory of hyper-physical influence. And so it is natural to assume that the theory to which Kant subscribes in the *Critique* and the *Prolegomena* is the one that he contrasts with the empiricist theory and the theory of pre-established intellectual harmony. But in Kant's letter to Herz this is the theory of hyper-physical influence.

However, there are also at least two good reasons not to accept this interpretation of Kant's theory of the objective validity of the Categories. The first reason is that there is no evidence in the *Critique* that Kant has found a way to deal

with the objections to the theory of hyper-physical influence that he raised in his letter to Herz. If Kant has finally accepted one of the two versions of the theory of hyper-physical influence – if he has decided that the Categories are concepts in the mind of God of which we are conscious either by means of recollection or because God allows us to see directly into His mind – then we should expect him to explain why this theory does not create a vicious circularity in our demonstration of the objective validity of the Categories or make it impossible for us to tell for sure whether or not a given concept is objectively valid in spite of the fact that it is an empirical concept. However, at no point in the *Critique* does Kant even attempt to respond to these objections. There is no mention at all of God in the part of *Critique* where Kant discusses the objective validity of the Categories. Does Kant accept a third version of the hyper-physical influence theory on which Kant is himself the creator of objects in the world – that is to say, a version of the theory that ascribes to Kant the role traditionally reserved for God? This is extremely implausible.

A further, conclusive reason not to interpret Kant's theory of the Categories in the way that I suggested is that Kant explicitly rejects the theory on which the Categories serve as the archetypes for objects in the world in the section of the Deduction of the Categories that he calls the Transition to the Transcendental Deduction of the Categories. He writes:

There are only two possible cases in which synthetic representation and its objects can come together, necessarily relate to each other, and, as it were, meet each other. Either if the object alone makes the representation possible or if the representation alone makes the object possible. If it is the

first then this relation is only empirical... But if it is the second, then since representation in itself (for we are not here talking about its causality by means of the will) does not produce its object as far as its existence is concerned, the representation is still determinant of the object *a priori* if it is possible through it alone to cognize something as an object. (A92/B124-125)

Kant's claim in the first two sentences of this passage is that certain states of our mind are able to serve as representations of objects only if either one of two conditions is satisfied. Either these synthetic representations conform to the objects or the objects conform to the representations – that is to say, either it is because the objects take a particular form that the states of mind take this form or it is because the states of mind take a particular form that the objects take this form. If a state of mind serves as a representation of an object in virtue of the fact that the state of mind conforms to the object then it is an empirical representation. But this is not the way that Kant thinks that the Categories serve as representations of objects. He thinks that the Categories serve as representations of objects in virtue of the fact that the objects conform to the states of mind. However, Kant's claim in the fourth sentence of the passage that I just quoted is that the sort of representation that interests him in his Deduction of the Categories – that is to say, the Pure Concept of the Understanding – 'does not produce its object as far as its existence is concerned'. Kant adds that 'we are not here talking about its causality [i.e. the causality of the representation] by means of the will'. Now remember that on the theory of God's concepts that I discussed in §1.2, these concepts serve as representations of objects precisely because the concepts are the causes of the

objects in conjunction with God's will. So it seems to be this very theory of God's concepts that Kant has in mind in the fourth sentence. His point is that the Categories do not serve as representations of objects in the way that God's concepts are traditionally understood to serve as representations of objects.

How can Kant subscribe to the theory that the Categories serve as representations of objects in virtue of the fact that the objects conform to the Categories and yet reject the claim that they serve as representations of objects in the way that God's concepts are traditionally understood to serve as representations of objects? The answer to this question has to do with the way that the objects of representation are conceived. Consider once again the key lines from the Preface to the second edition of the *Critique*. It should be noted that what Kant does here is to contrast a case in which the concepts conform to their objects with the case in which the objects, 'or, what is the same thing, the experience in which alone they can be cognized' (Bxvii), conform to the concepts. He goes on to say that in this second case

I immediately see an easier way out of the difficulty, since experience itself is a kind of cognition requiring the understanding, whose rule I have to presuppose... *a priori*, which rule is expressed in concepts *a priori*, to which all objects of experience must therefore necessarily conform, and with which they must agree. (Bxvii-xviii)

This reference to experience is also found in the part of the second edition Deduction of the Categories where Kant discusses possible theories about the objective validity of the Categories:

Now there are only two ways in which a necessary agreement of experience with the concepts of its objects can be thought: either the experience makes these concepts possible or these concepts make the experience possible... If someone still wanted to propose a middle way between the only two already named ways, namely, that the categories.... were rather subjective predispositions... implanted in us... by our author in such a way that their use would agree exactly with the laws of nature along which experience runs... (B166-167)

Finally, here is the end of the passage from the Transition to the Transcendental Deduction of the Categories:

...if it is the second [i.e. if the representation alone makes the object possible], then since representation in itself (for we are not here talking about its causality by means of the will) does not produce its object as far as its existence is concerned, the representation is still determinant of the object *a priori* if it is possible through it alone to cognize something as an object. (A92/B124-125)

The crucial point about Kant's theory of the Categories is this: Like Plato and Malebranche, Kant thinks that the Categories count as objectively valid in virtue of the fact there are objects that conform to these Categories. However, unlike Plato and Malebranche, Kant thinks that the objects that conform to the Categories are objects

insofar as they stand in another relation to our mind, viz. objects insofar as they count as objects of experience for us. To be clear, Kant's claim is that if an object is to count as an object of experience for us then it is necessary for this object to conform to the Categories. It follows from this that if there are objects of experience then there are objects that stand in a relation of necessary conformity to the Categories. And this is sufficient to show that if there are objects of experience then the Categories serve as representations of these objects – and, therefore, as objectively valid. To put this point another way, Kant thinks that the Categories serve as representations of objects in virtue of the fact that they serve as blueprints for objects insofar as they count as objects of experience for us. If an object counts as an object of experience for us then it must conform to the Categories. By contrast, on the theory of the Categories that Kant ascribes to Plato and Malebranche, the objects that are supposed to conform to the Categories are not simply objects insofar as they are objects of experience for us. Plato and Malebranche think that the Categories serve as representations of objects as they are in themselves.

My argument so far in this chapter turns on the assumption that the question Kant raises in his letter to Herz is the same question that he sets out to answer in his Deduction of the Categories. It has been proposed, however, that there are important differences between these two questions. Longuenesse has argued that there is a 'radical change in the formulation of the problem' between Kant's letter to Herz and the Deduction of the Categories itself (1998, p.20). She notes first that there is a change in the terms that Kant uses when he is discussing possible answers to his question about the Categories. Whereas in his letter to Herz Kant refers to the idea that a representation might be caused by an object or that it might be the cause of the

object itself, in the Deduction of the Categories itself he refers to the idea that an object might make the representation possible or be made possible by the representation. Longuenesse writes that this change in terms 'is only a manifestation of a more fundamental shift' (1998, p.20). She continues:

Kant is no longer examining the relation of two heterogeneous elements (one "within" and the other "outside" representation), but the relation of two elements both internal to representation. We are thus no longer faced with an alternative between two causal relations opposite in direction, but with the cooperation of two complementary relations, which together constitute the "relation of a representation to its object". (1998, p.20-21)

What Longuenesse argues here is not easy to understand. But she evidently thinks that what has happened in the time between Kant's letter to Herz and the Deduction of the Categories is that he has changed his mind about the nature of the relation between a representation and its object. And she thinks that this change of the mind is precisely what allows Kant to find an answer the question of the objective validity of the Categories.

There is good reason not to accept Longuenesse's analysis of the difference between Kant's discussion of the objective validity of the Categories in the *Critique* and in his letter to Herz. First, we should not conclude from the change in the language that Kant uses to describe the different ways in which a state of mind can serve as a representation of an object that he has reformulated his question. If we look carefully at the various theories of the Categories that Kant mentions and criticizes in his letter

to Herz then we find that these theories are easily restated in terms of the language of one thing making another thing possible. Arguably, what Kant has done in his *Deduction of the Categories* is just to express the different ways in which a state of mind can serve as a representation of an object in clearer and more precise terms. Take the empiricist theory, viz. the theory that the *Categories* serve as representations of objects in the same way that, for example, the Tube Map serves as a representation of the network of lines and stations on the London Underground, or a portrait of the Queen of England serves as a representation of the Queen herself. When Kant refers in his letter to Herz to states of the mind that serve as representations of objects in this way, he describes them as representations that are ‘caused by the object’, and that conform to an object ‘as an effect accords with its cause’ (Ak X:130-131). However, it isn’t quite right to describe the Queen herself as the cause of the portrait or the Tube Map as the effect of the network of lines and stations on the London Underground. For a state of mind to serve as an empirical representation of an object, it isn’t necessary for that object to count as the cause of the representation. What matters is just that the mind would not be determined in the way that it is if it were not for the fact that the object is determined in the same way. What matters is that the object – or, rather, the form that the object takes – is what makes it possible for the mind to take this same form. In the same way, the Tube Map serves as a representation of the network of lines and stations on the London Underground in virtue of the fact that the lines and nodes on the map would not be related in the way that they are if it were not for the fact that the lines and stations on the London Underground are related in the same way.

As for Longuenesse's second point, she does not give us any good reason to think that what has occurred between Kant's letter to Herz and his Deduction of the Categories is a radical change in his formulation of the problem as opposed to a change in his view about the best way to solve the problem. Longuenesse is surely correct that in the Deduction of the Categories Kant thinks of the objects that are represented by the Categories in a very different way. In particular, Kant thinks that the Categories serve as representations of objects only insofar as they are objects of experience for us. There is no hint of this in Kant's letter to Herz. However, we should not conclude on this basis that Kant has changed his mind about what is required for a state of mind to serve as a representation of an object. On the contrary, there is evidence in the *Critique* that Kant still assumes that what is required for a state of mind to serve as a representation of an object is a relation of necessary conformity between the two. What matters is that there should be a relation of agreement or conformity between the state of mind and the object, and that it should be no accident that there is this relation of agreement or conformity between the state of mind and the object. Kant has only changed his mind about the nature of the objects that stand in this relation of necessary conformity or agreement to objects. He has decided that these objects can only be objects insofar as they count as objects of experience for us. I conclude that Longuenesse is wrong to speak of a 'radical change in the formulation of the problem' (1998, p.20). We are entitled to assume that the question about the Categories or Pure Concepts of the Understanding that Kant sets out to answer in his Deduction of the Categories is the very same question that he raised in his famous letter to Herz.

§1.4 Dickerson on Kant's concept of representation

In §1.1, I argued that the aim of Kant's Deduction of the Categories is to explain how it is that the Categories are able to count as objectively valid. And I argued that what Kant means by the objective validity of a concept is the fact that this concept serves as a representation of an object. Finally, I argued that a concept serves as a representation of an object in the sense that Kant has in mind insofar as, first, it stands in a relation of agreement or conformity to an object, and second, it is no accident that it stands in a relation of agreement or conformity to an object. In §1.2, I showed that each of the three explanations of the objective validity of the Categories that Kant considers and rejects in his famous letter to Herz is an attempt to specify the relation of necessary conformity that obtains between the Categories and objects. In §1.3, I defended a further claim about the aim of Kant's Deduction of the Categories. I argued that what Kant wants to show with this argument is that the Categories serve as representations of objects in virtue of the fact that it is necessary for an object to conform to the Categories if it is to count as an object of experience for us. To put this point another way, the Categories serve as representations of objects on Kant's theory in virtue of the fact that they serve as the blueprints for an object of experience.

If my claims in §1.1 and §1.2 are correct then one of the keys to making sense of Kant's Deduction of the Categories is to understand his concept of representation. Unless we realize that a state of mind serves as a representation of an object in the sense that Kant has in mind only if it stands in a relation of necessary conformity to an object, we will be unable to understand what it means to demand an explanation of the objective validity of the Categories or Pure Concepts of the Understanding, and we will

be unable to understand the theory of the Categories that Kant defends in the Deduction of the Categories itself. Because Van Cleve (1999), for example, fails to realize the importance of Kant's concept of representation – because he interprets Kant's concept of objective validity solely in terms of the instantiation of a concept by at least some objects of experience – he cannot explain why Kant thinks it is necessary for a Deduction of the Categories to show not only that there are objects of experience that conform to the Categories but also that the conformity of objects of experience to the Categories is necessary.

One recent commentator who has drawn attention to the importance of Kant's concept of representation is Dickerson (2004). He also believes that we cannot understand Kant's Deduction of the Categories unless we realize exactly what Kant thinks is required for a state of our mind to count as a representation of an object. However, Dickerson's interpretation of Kant's concept of representation is very different to the one that I defended in §1.1. And my aim in this section is to compare these two interpretations. Dickerson starts from the observation that Kant is a 'representationalist' (2004, p.5). By this he means that Kant is someone who thinks that 'the immediate objects of consciousness are internal representative states' (2004, p.5). And he claims that Kant is required to explain 'how the mind's awareness of its own internal states can yet amount to, or provide the basis for, an awareness of an independent reality' (2004, p.10). We should understand this as follows: Dickerson ascribes to Kant the claim that while we are certainly aware of an independent reality, our awareness of this reality is indirect in the sense that we are aware of this reality only by means of our awareness of states of our own mind that serve in some way as representations of this reality. The question that Kant is required to answer is how it is

that the awareness of these states of our own mind can allow us to achieve an awareness of an independent reality.

Dickerson writes that there are 'at least two familiar models for understanding [this indirect awareness of an independent reality] in the representationalist tradition' (2004, p.10). The first is what he calls the 'indirect realist' model, which he defines as follows:

By 'indirect realism' I mean a position that thinks of representing objects as involving (i) an act of awareness of an idea (or representation, impression, etc.) and (ii) an inference to the external cause of that idea. (2004, p.10)

On this model, which Dickerson notes is often associated with Locke and Descartes (2004, p.11), the states of our mind serve as representations of objects because they 'stand for external things via a relation of natural resemblance or symbolism' (2004, p.10). The idea here is that objects cause us to be in particular states of mind – specifically, they cause us to be in states of mind that are in some way similar to the objects themselves. We become aware of the properties of the objects in virtue of the fact that we draw inferences to the properties of these objects on the basis of the properties of the states of mind that they have caused us to be in.

It should be clear that on this first model, the states of our mind serve as representations of objects precisely because they conform to the objects – that is to say, they are able to serve as representations of objects in virtue of the fact that it is only because the objects have a particular form that the states of our mind have this same form. The idea is that the states of our mind count as representations of objects in the

same way that the Tube Map counts as a representation of the network of lines and stations on the London Underground. By the way, there is plenty of evidence to support the association of this model with Descartes. Descartes clearly commits himself to this model in his *Rules for the Direction of the Mind*. He writes first that ‘sense perception occurs in the same way in which wax takes on an impression from a seal’ (1985, p.40), adding that

[it] should not be thought that I have a mere analogy in mind here: we must think of the external shape of the sentient body as being really changed by the object in exactly the same way as the shape of the surface of the wax is altered by the seal. (1982, p.40)

He goes on to write that

when an external sense organ is stimulated by an object, the figure which it receives is conveyed at one and the same moment to another part of the body known as the ‘common’ sense (1982, p.41).

This common sense then ‘functions like a seal, fashioning in the phantasy or imagination, as if in wax, the same figures or ideas which come, pure and without body, from the external senses’ (1982, p.41). Descartes’ claim is that there are three steps in sense perception. First, the state of a part of our body, in particular, the state of one of our sense organs, is caused to have the same form as an object, just as a piece of wax is caused to have the same shape as a seal. Second, the state of another part of our body,

namely, a state of our 'common sense', is caused to have the same form as the sense organ. Third, a state of our mind is caused to have the same form as the state of our common sense. The result of this process is that a state of our mind conforms to the object that affected our sense organ in the first place. It is only because the object that affected our sense organ had, say, the particular shape that it did that we acquired an idea of this particular shape in our mind. I take it that this theory of Descartes' perfectly illustrates the indirect realist model that Dickerson has in mind.

The second model that Dickerson mentions in his discussion of Kant's concept of representation is what he calls the 'idealist' model, which he associates with Berkeley. By this he means

a position that thinks of representing objects as involving (i) an act of awareness of an idea and (ii) constructive acts in which that idea is linked with other ideas. (2004, p.11)

On this model, there is no independent reality apart from the ideas or representations in our mind:

Instead, the things of the external world are identified with or constructed out of ideas... and veridical perceptions are those complex arrays of ideas that obey certain rules of coherence and consistency. (2004, p.11)

As Dickerson also puts it,

On this model... being aware of reality is not a question of leaping beyond one's ideas through an inference to their causes... Rather, it is a question of enriching the content of one's ideas by connecting them with one another.

(2004, p.11)

It is a little harder to see how this second model counts as an explanation of the way that the states of our mind serve as representations of objects. The key point, I take it, is that the objects that are represented by the states of our mind are themselves mere ideas or representations – specifically, they are complex ideas constructed out of other ideas or representations. And, loosely, a state of mind is able to serve as a representation of an object insofar as it can be fitted into – that is to say, it can become a part of – one of these complex ideas. Since this model of representation involves the concept of construction (2004, p.11, p.13), it may be tempting to associate it with the second model of representation that I extracted from Kant's letter to Herz. This is the model on which a state of our mind serves as a representation of an object in the same way that the blueprint for a house serves as a representation of the house itself, which is to say, in virtue of the fact that the house has been constructed in such a way as to ensure that it conforms to the blueprint. But this would be a mistake. First, on the blueprint model, the object that is represented by a state of our mind is not itself a state of our mind. Or, at any rate, it is not necessary for it to be a state of our mind. Second, on the blueprint model, the state of mind and the object are distinct entities that stand in some sort of relation of agreement or conformity to one another. The state of mind is not a part of the object. The object is constructed in conformity with the state of mind, but the state of mind is not one of the entities out of which the object is constructed.

Dickerson argues that neither of these models – the indirect realist model or the idealist model – can be used to understand Kant’s concept of representation. He doesn’t think that Kant can be committed to the claim that a state of our mind represents an object only if it conforms to an object in the way that the impression in the wax represents the seal, or in the way that the Tube Map represents the network of lines and stations on the London Underground. Similarly, he doesn’t think that Kant can be committed to the claim that a state of mind represents an object only if it can be fitted in some way into a coherent and consistent complex of states of our mind. And he concludes on this basis that we need to find some other way to understand Kant’s concept of representation. Now, Dickerson is surely right to think that Kant is not committed to the first of these claims. Were Kant committed to this claim then he could not allow for the possibility of any representations beside empirical ones. What Dickerson fails to notice is that Kant does not simply deny that the states of our mind are able to represent objects in the way that the impression in the wax represents the seal. Kant’s claim is rather that not all of the states of our mind that serve as representations of objects represent their objects in this way. Because Dickerson fails to appreciate this point, he does not realize that standing in the sort of relation to an object specified by the indirect realist model is just one of several ways in which Kant thinks that a state of mind can fall under the concept of representation. And so Dickerson fails to ask himself what are the features of a state of mind that serves as a representation of an object in this way in virtue of which it falls under the concept of representation. He doesn’t see that such a state of mind serves as a representation of an object in virtue of the fact that, first, there is a relation of agreement or conformity between the state of mind and the object, and second, this relation of agreement or

conformity is necessary in the sense that it is no accident that the state of mind and the object conform to one another.

So how does Dickerson understand Kant's concept of representation? To illustrate his interpretation he draws attention to the phenomenon of 'seeing things in pictures' (2004, p.13). Suppose that I can see in front of me a piece of paper, upon which there is a set of black marks - specifically, lines and dots. And suppose that I interpret this as a picture of a smiley face. Dickerson thinks that there is a clear sense in which the set of black marks counts as a representation of a smiley face. The question is what the relation between the smiley face and the black marks consists in. Dickerson's first point is that 'the smiling face is not a separate object that lies 'behind' or 'outside of' the configuration of dots or lines' (2004, p.14). The configuration of dots and lines does not function 'like a signpost that points beyond itself to some further object (i.e., the smiling face)', nor is it 'as it were evidence for, or a natural sign of, the smiling face' (2004, p.14). However - and this is Dickerson's second point - the smiling face is also not identical to the configuration of lines and dots on the piece of paper. As Dickerson notes:

The configuration is composed of ink marks on paper, but the face is not composed of ink marks on paper - it is composed of eyes, a nose and a mouth. (2004, p.15)

The object of the picture - what Dickerson calls the 'depicted object' (2004, p.15) - is an object that is 'seen in', and in that sense represented by, the configuration of lines and dots on the piece of paper - which configuration of lines and dots Dickerson calls

the 'representational medium' (2004, p.13, p.15). Dickerson claims that '[it] is something analogous to this act of 'seeing in' that is the key to Kant's account of representation (2004, p.17). Dickerson thinks that a state of our mind serves to represent an object to us in the sense that Kant has in mind insofar as, loosely, we see an object in this state of our mind.

Here is a different example to illustrate the difference between Dickerson's interpretation of Kant's concept of representation and the one that I have defended in this chapter. Suppose that an artist has painted a portrait of the Queen of England. We can ask ourselves: What is the object that is represented to us by this portrait? One obvious answer is that it is the Queen of England herself. But there is another possible answer, viz. that it is the queen that we see in the portrait that the artist has painted. Suppose that the artist has made a mistake and has painted the Queen as having brown eyes when in fact her eyes are blue. There is a sense in which the object that is represented to us by the portrait is a person who has blue eyes. But there is another sense in which the object that is represented to us by the portrait is a person who has brown eyes. On the interpretation that I have defended in this chapter, the relation of representation that interests Kant is the relation that obtains between the portrait and the Queen of England herself. On the interpretation that Dickerson prefers, the relation of representation that interests Kant is the relation that obtains between the portrait and the queen that we see in the portrait - that is to say, the person who has brown eyes instead of blue.

Dickerson provides little in the way of direct textual support for his interpretation of Kant's concept of representation. He claims that '[there] are no key analyses or definitions in the *Critique* upon which an interpretation of Kant's notion of

representation can be grounded’, and concludes that ‘such an interpretation must instead be justified by its capacity to provide a coherent understanding of Kant’s text as a whole’ (2004, p.4). Dickerson is perfectly correct that there is no definition of the concept of representation in the *Critique*. However, as I have shown in §1.1 and §1.2, we can still work out what Kant thinks is required for a state of mind to count as a representation of an object if we look carefully at the different ways in which he thinks that a state of mind can count as a representation of an object. What we discover is that a state of mind counts as a representation of an object in the sense that Kant has in mind insofar as there is a relation of necessary conformity between the state of mind and the object. And if this is correct then it must be the sort of relation that obtains between the portrait of the Queen of England and the Queen herself – that is to say, between the portrait and the blue eyed Queen – that is of interest to Kant in his Deduction of the Categories. It is not, as Dickerson thinks, the relation between the portrait and the brown eyed Queen.

§2 Kant’s Concept of Empirical Cognition

In §1, I defended an interpretation of Kant's concept of the objective validity of the Categories. I argued that the Categories count as objectively valid in the sense that Kant has in mind insofar as they serve as representations of objects. And I argued that the Categories serve as representations of objects in the sense that Kant has in mind insofar as, first, there are objects that stand in a relation of conformity to the Categories, and second, it is no accident that these objects stand in a relation of conformity to the Categories. I also defended an initial claim about Kant's theory of the objective validity of the Categories. I ascribed to Kant the claim that the Categories count as objectively valid in virtue of the fact that it is necessary for an object to conform to these Categories if it is to count as an object of experience for us. On Kant's theory, the Categories count as objectively valid because they serve as the blueprints for an object of experience.

My aim in this chapter is to describe in more detail how Kant intends to explain the objective validity of the Categories. My point of departure is Kant's concept of experience - or, in other words, his concept of empirical cognition. Kant thinks that a state of empirical cognition has two components. He thinks that someone has achieved a state of experience only if he has formed both an empirical intuition of an object and also a thought or concept of an object. The first claim that I defend in this chapter is as follows: What Kant thinks we need to do in order to show that the Categories are objectively valid is to show that it is necessary for an object to conform to these Categories if it is to count as an object of thought. Kant thinks that it is precisely because the conformity of an object to the Categories is required for it to be an object of thought that it is required for it to be an object of experience or empirical cognition.

The other claim that I defend in this chapter is a claim about what Kant means by the thought of an object. I argue that someone has formed a thought of an object in the sense that Kant has in mind in his Deduction of the Categories – that is to say, he has made use of his faculty of understanding – insofar as he has made a judgment about an object – say, the judgment that there are clouds over Oxford, or that England will win the forthcoming Ashes series. If this claim is correct then Kant’s strategy in his Deduction of the Categories is to show that there is a connection between the act of judgment and the conformity of objects to the Categories. Kant wants to show that if we are to make a judgment about an object then this object must conform to the Categories. If he can show this then he can conclude that there is a relation of necessary conformity between the Categories and the objects of experience. But that is as much as to say that the Categories serve as representations of the objects of experience.

At the start of §2.1, I focus on Kant’s discussion of the principle of a transcendental deduction in general in the Transition to the Transcendental Deduction of the Categories. I argue on the basis of this passage of text that Kant’s strategy in his Deduction of the Categories is to show that objects must conform to the Categories if they are to count as objects of thought or understanding. In the rest of §2.1, I criticize some of the arguments that have been advanced on behalf of the claim that according to Kant the conformity of an object to the Categories is required for the mere intuition of an object or even the mere consciousness of an object. In §2.2 I introduce and defend my interpretation of Kant’s concept of the thought of an object. I argue that someone has formed a thought of an object in the sense that Kant has in mind insofar as he has made a judgment about an object. And I defend this claim against an

objection arising from the connection between Kant's concept of understanding and his concept of cognition. In §2.3 and §2.4 I consider two alternative interpretations of Kant's concept of empirical cognition. On the first interpretation, recently defended by Kitcher (2011), what Kant thinks is required for someone to achieve a state of empirical cognition is for her to make a judgment that is rational in the sense that she is conscious of her grounds for judging as she does. On the second interpretation, Kant thinks that someone has achieved a state of empirical cognition only if she has acquired empirical knowledge. I argue that neither of these interpretations is supported by the text of the *Critique*.

§2.1 The principle of Kant's Deduction of the Categories

Kant claims in a section of the Deduction of the Categories that he calls the Transition to the Transcendental Deduction of the Categories that

[the] transcendental deduction of all concepts *a priori* has a principle in accordance with which the whole investigation must be directed, namely this: that they [i.e. *a priori* concepts] must be recognized as conditions *a priori* of the possibility of experience (be that of the intuition that is found in [experience] or of the thinking). (A94/B126)

What Kant is telling us here is what we have to do if we want to deduce an *a priori* concept – if, that is, we want to explain how an *a priori* concept is able to serve as a representation of an object. It should be noted, however, that his claim is not only

about the Deduction of the Categories. When Kant refers to ‘all concepts *a priori*’, he has in mind not only the Categories – that is to say, the Pure Concepts of the Understanding – but also the concepts of space and time. This is an important point to bear in mind because Kant typically contrasts the Pure Concepts of the Understanding with what he calls the pure intuitions of space and time, which may lead us to assume that what he means by ‘all concepts *a priori*’ in the Transition to the Transcendental Deduction of the Categories is all Pure Concepts of the Understanding.

How do we know that this isn’t what Kant means by ‘all concepts *a priori*’ in this context? Kant claims that if we want to deduce an *a priori* concept then what we need to do is to show that this concept serves as an *a priori* condition of the possibility of experience. As he puts it a little later on in the Transition to the Transcendental Deduction of the Categories:

if one wants to know how Pure Concepts of the Understanding are possible, one must enquire what are the *a priori* conditions on which the possibility of experience depends... (A95-96)

And Kant claims that there are two ways in which we can show that a concept serves as an *a priori* condition of the possibility of experience. We can show that the concept serves as an *a priori* condition of the possibility of ‘the intuition that is found in [experience]’, or we can show that the concept serves as an *a priori* condition of the possibility of the ‘thinking’ that is found in experience. Kant makes this second claim

for the following reason. He defines experience as a particular sort of cognition, viz. empirical cognition:

the categories provide us with no cognition of things by means of intuition other than through their possible application to empirical intuition, i.e. they serve only for the possibility of empirical cognition. This, however, is called experience. (B147, also B165-166, Ak IV:472)

And he thinks that a state of cognition has two components. It includes the intuition of an object and also a corresponding thought of, or concept of, an object:

cognition has two parts: first, the concept, through which in general an object is thought... and second, the intuition, through which it is given... (B146, also A92-93)

A state of cognition counts as a state of empirical cognition – that is to say, a state of experience – insofar as the intuition that is involved in this state of cognition is itself empirical:

all experience contains, besides the intuition of the senses, through which something is given, also a concept of an object, which is given in the intuition or [that] appears. (A93/B126)

According to Kant, someone is in a state of experience only if she satisfies two conditions. First, she must have formed an empirical intuition of an object, and second, she must have formed a corresponding thought of an object. And so it seems that if an *a priori* concept is to serve as a condition of the possibility of experience this must be because it serves either as a condition of the possibility of the intuition that is found in experience, that is to say, as a condition of the possibility of empirical intuition, or as a condition of the possibility of the thinking that is found in experience.

However, even before Kant has reached the Transition to the Transcendental Deduction of the Categories, he has committed himself to the claim that the Pure Concepts of the Understanding do not serve as conditions of the possibility of the intuition that is found in experience. He writes a few pages before that

[the] categories of the understanding... do not represent to us the conditions under which objects are given in intuition at all, hence objects can indeed appear to us without necessarily having to be related to functions of the understanding... (A89/B122),

He also writes that 'intuition by no means requires the functions of thinking' (A91/B123). On the other hand, Kant has made it clear that the concepts of space and time can be deduced by showing that they serve as *a priori* conditions of the possibility of empirical intuition:

In the case of the concepts of space and time, we were able to make comprehensible with little effort how these, as *a priori* cognitions, must

nevertheless necessarily relate to objects... (A89/B121).

So it must be the concepts of space and time that Kant has in mind when he allows for the possibility that an *a priori* concept can be deduced by showing that it serves as an *a priori* condition of the possibility of empirical intuition. His claim about the principle of the transcendental deduction of all concepts *a priori* must be intended to cover the concepts of space and time as well as the Pure Concepts of the Understanding.

If my argument so far in this section is correct, we can draw the following key conclusion from Kant's claim about the principle of the transcendental deduction of all concepts *a priori*. Kant thinks that if we want to deduce the Categories or Pure Concepts of the Understanding, what we need to do is to show that these concepts serve as *a priori* conditions of the possibility of the thinking – that is to say, the thought or concept of an object – that is found in experience. And there is plenty of evidence to corroborate this conclusion. Kant remarks a little earlier on in the Transition to the Transcendental Deduction of the Categories:

Now the question arises whether concepts *a priori* do not also [lie in the mind *a priori*], as conditions under which alone something is, if not intuited, nevertheless thought as [an] object in general. (A93)

He also writes that

the objective validity of the categories, as *a priori* concepts, rests on the fact that through them alone is experience possible (as far as the form of

thinking is concerned). (A93/B126)

Later on in the first edition Deduction of the Categories, Kant states that

the categories... are nothing other than the conditions of the thinking in a possible experience, just as space and time contain the conditions of the intuition [in a possible experience]. (A111)

We should also consider Kant's description of his aim in the section of the *Critique* that contains his Deduction of the Categories, viz. the Analytic of Concepts:

I understand by an analytic of concepts not their analysis, or the usual procedure of philosophical investigations, that of analyzing the content of concepts that present themselves and bringing them to distinctness, but rather the much less frequently attempted analysis of the faculty of understanding itself, in order to research the possibility of *a priori* concepts by seeking them only in the understanding as their birthplace... (A65-66/B90)

Kant's aim in the Analytic of Concepts is to analyze a particular faculty of the mind, viz. the faculty of understanding, for the sake of showing that there can be *a priori* concepts of objects. Now, Kant defines the faculty of understanding as the faculty for thinking of objects or forming concepts of objects:

The faculty for thinking of objects of sensible intuition... is the understanding. (A51/B75)

these two faculties [sensibility and understanding] cannot exchange their functions. The understanding is not capable of intuiting anything, and the senses are not capable of thinking anything. (A51/B75).

So when Kant refers to an analysis of the faculty of understanding, what he has in mind is an analysis of the faculty of the mind that is responsible for thinking – and, therefore, for the thinking that is an essential ingredient in empirical cognition. His aim in the Analytic of Concepts – the section of the *Critique* that contains his Deduction of the Categories – is to look carefully at what is required for us to make use of our faculty of thinking of objects – and, in this way, to address the question how there can be concepts that count as objectively valid in spite of the fact that they are pure *a priori* concepts.

It will be helpful at this point to summarize the results of my analysis of Kant's claim about the principle of a transcendental deduction. I have argued that what Kant thinks we need to do if we want to explain how the Categories can be objectively valid is to look into the conditions that must be satisfied if we are to think of, or form a concept of, an object of intuition – if, that is, we are to make use of our faculty of understanding. Remember that Kant's aim in the Deduction of the Categories is to show that the Categories serve as representations of objects. And as I explained in §1.3, his theory is that the Categories serve as representations of objects in virtue of the fact that it is necessary for an object to conform to the Categories if it is to count as an

object of experience for us. To put this point another way, Kant wants to show that it is a condition of the possibility of an object of experience that it conforms to the Pure Concepts of the Understanding. Now, Kant thinks that an object counts as an object of experience or empirical cognition only if it satisfies two conditions: First, this object must be an object of empirical intuition. Second, this object must be an object of thought – that is to say, it must be an object of which we have formed a concept. But Kant doesn't think that it is a condition of the possibility of an object of empirical intuition that it conforms to the Pure Concepts of our Understanding. So he thinks that what we need to do if we want to show that it is a condition of the possibility of an object of experience that it conforms to the Pure Concepts of our Understanding – and, in this way, to show that the Pure Concepts of the Understanding serve as representations of objects – is to show that it is a condition of the possibility of an object of thought that it conforms to the Pure Concepts of the Understanding.

It is widely accepted that Kant's strategy in his Deduction of the Categories is to investigate the conditions that must be satisfied if we are to achieve a state of experience or empirical cognition (cf. Kitcher 2011, p.85-86, p.89). Van Cleve, for example, writes that

Kant's argumentative strategy is a paradigm instance of what is nowadays often called "transcendental proof": he argues that if the categories were not objectively valid, experience would be impossible. (1999, p.73)

Similarly, Ameriks thinks that Kant's Deduction of the Categories is structured around the claim that 'empirical knowledge ("experience") is possible only if... pure concepts

have validity...’ (1978, p.276).

It should be noted, by the way, that what Van Cleve and Ameriks ascribe to Kant is a weaker claim than the one that I have ascribed to him. They do not ascribe to Kant the claim that in order for an object to count as an object of experience this very object must conform to the Categories. Rather, they ascribe to Kant the claim that in order for us to have experience of an object – that is to say, in order for there to be an object of experience for us – there must be an object that conforms to the Categories. Notice also that what Van Cleve and Ameriks claim is a condition of the possibility of experience is the objective validity of the Categories. On the interpretation that I have defended, Kant thinks that it is the conformity of the object to the Categories that is a condition of the possibility of the experience of that object. Now, as I explained in §1, on Van Cleve’s view, all it means to say that the Categories are objectively valid is to say that some objects conform to them – or, to be clear, that some objects fall under them. So Van Cleve would accept the claim that I have defended. However, if the interpretation of objective validity that I defended in §1 is correct, the objective validity of the Categories does not consist simply in the fact that there are objects that fall under these concepts. Rather, it consists in the fact that the conformity of an object to the Categories is necessary – say, necessary in order for this object to count as an object of experience. So if my interpretation is correct then Kant’s aim in his Deduction of the Categories is not to show that the objective validity of the Categories is a condition of the possibility of experience. Rather, his aim is to show that the conformity of objects to the Categories is a condition of the possibility of our experience of those objects. To show this is to show that the Categories are objectively valid.

There is textual evidence in support of this claim. At one point in the second

edition Deduction of the Categories, Kant writes:

all synthesis... stands under the Categories, and since experience is cognition through connected perceptions, the categories are conditions of the possibility of experience, and are thus also valid *a priori* of all objects of experience. (B161)

Kant's claim here is that the Categories are valid of objects of experience in virtue of the fact that they serve as conditions of the possibility of our experience of objects. The claim is not that the objective validity of the Categories is a condition of the possibility of our experience of objects.

What is less widely accepted by commentators is the claim that Kant's strategy in his Deduction of the Categories is to investigate the conditions that must be satisfied if we are to engage in thinking that is involved in experience – if, that is, we are to make use of our faculty of understanding. Some commentators maintain that there is reason to deny that what Kant thinks we need to do in order to deduce the Categories is to show that an object must conform to the Categories if it is to count as an object of thought. One example is Van Cleve. Van Cleve thinks that there is conclusive evidence that Kant's strategy in his Deduction of the Categories – or, at any rate, his strategy in the sections that follow the Transition to the Transcendental Deduction of the Categories – is to investigate the conditions that must be satisfied if we are to have any consciousness whatsoever, which is to say, if we are to have experience in a very weak sense indeed. The conclusion of Kant's argument, according to Van Cleve, is that 'consciousness without the categories is impossible', or, that '[the] categories apply to all

representations of which I am conscious' (1999, p.79).

What does Van Cleve think is the evidence that Kant's strategy in his Deduction of the Categories is to show that the Categories serve as conditions of the possibility of consciousness in general? He quotes the final sentence of the following passage:

Actual experience, which is constituted by apprehension, association (reproduction), and finally recognition of appearances, contains in recognition, the last and highest of these merely empirical elements of experience, certain concepts which render possible the formal unity of experience, and with that all objective validity (truth) of empirical cognition. These grounds of the recognition of the manifold, so far as they concern solely the form of an experience in general, are the categories... [Only] by means of these fundamental concepts can appearances belong to cognition or even to our consciousness, and so to ourselves. (A124-125)

Van Cleve insists that the final sentence of this passage shows conclusively that Kant regards the Categories as conditions of the possibility of consciousness, not just as conditions of the possibility of thinking of, or forming a concept of, an object (1999, p.79). He allows that Kant did at one point consider trying to show that the Categories constitute conditions of the possibility of thinking, but he claims that in the lines just quoted Kant 'evidently repudiates' this strategy (1999, p.79).

It is hard to explain exactly why Van Cleve's reading of the passage quoted above is incorrect without touching on some claims that will only become clear in subsequent chapters. But the key points are as follows: First, it should be noted that Kant

connects the Categories to the ‘last and highest’ of the ‘empirical elements of experience’, which is to say, ‘recognition’. The Categories are, Kant tells us, ‘[the] grounds of the recognition of the manifold’, and, therefore, conditions of the possibility of experience. However, as will become clear in §3.4, Kant thinks that there is a close connection between the act of recognition and the act of forming a concept. He thinks that it is an act of recognition that constitutes a concept (cf. A103- 104). So even in the passage that Van Cleve quotes as conclusive evidence that Kant has abandoned his claim that the Categories serve as conditions of the possibility of thinking of, or forming a concept of, an object, Kant restates his commitment to this claim. Second, we can allow that Kant’s strategy is to show that the Categories serve as conditions of the possibility of a sort of consciousness without abandoning the claim that his strategy is to show that the Categories serve as conditions of the possibility of thinking. As will also become clear in the course of §3, Kant identifies our faculty of or understanding with our ability to achieve a special sort of consciousness that he calls ‘transcendental apperception’ or ‘original apperception’. And the simplest way to reconcile Kant’s claim in the key sentence quoted above with the evidence about his strategy in the Transition to the Transcendental Deduction of the Categories is to accept that when he writes that ‘only by means of [the Categories] can appearances belong to... our consciousness’, what he really means is that it is only by means of the Categories that appearances can stand in a certain relation to our original apperception – and, therefore, to our faculty of understanding. This is a reasonable move to make in view of the fact that, first, Kant often refers to our faculty of transcendental or original apperception simply as our faculty of apperception (e.g. B153), and second, Kant tends to use the terms ‘apperception’ and ‘consciousness’ interchangeably (e.g. B139-140).

So much for the claim that Kant's strategy in his Deduction of the Categories is to investigate the conditions of the possibility of consciousness in general, as opposed to the conditions of the possibility of thinking. Some other commentators have maintained that the strategy that Kant adopts in his Deduction of the Categories is to show that the Categories serve as conditions of the possibility of empirical intuition. A prominent example here is Allison (2004). Allison thinks that the basic worry that Kant wants address in his Deduction of the Categories is that the concepts that we cannot avoid using if we want to form thoughts or judgments about objects – that is to say, the Categories or Pure Concepts of the Understanding – might turn out not to be instantiated by any of the objects of sensible intuition – or, as he puts it, 'that the deliverances of sensibility might not correspond to the *a priori* rules of thought' (2004, p.160). And Allison writes that what Kant tries to do in the part of his Deduction of the Categories that addresses this worry is to 'establish the applicability of the categories to whatever is given under the conditions of human sensibility', to demonstrate 'that the Categories... have a non-discursive function as conditions under which whatever is given... can enter empirical consciousness', 'to link the Categories... to the perception rather than merely the thought of objects' (2004, p.162).

Does Kant really want to show that the Categories serve as conditions of the possibility of the intuition of objects by means of the senses? This seems wildly implausible in view of some remarks that I have already quoted in this section:

[the] categories of the understanding... do not represent to us the conditions under which objects are given in intuition at all, hence objects can indeed appear to us without necessarily having to be related to

functions of the understanding... (A89/B122)

Kant also writes that even if appearances did not conform to the Categories, they would 'nevertheless present objects to our intuition, for intuition in no way requires functions of thinking' (A90-91/B123).

But there are also some passages of text that can seem at first glance strongly to support the claim that Kant's strategy in his Deduction of the Categories is to show that the Categories serve as conditions of the possibility of empirical intuition. In §26 of the second edition Deduction of the Categories, Kant draws attention to a special act of the mind, which he calls 'the synthesis of apprehension', which he defines as

the composition of the manifold in an empirical intuition, through which perception, i.e. empirical consciousness of it... becomes possible. (B160)

Suppose we assume that what Kant means by perception in this context is empirical intuition. In that case, Kant's claim here implies that the synthesis of apprehension is an act of the mind that serves as a condition of the possibility of empirical intuition. Now, Kant goes on to say in this passage that 'a *combination* with which everything that is to be represented determinately in space or time must accord, is already at the same time given *a priori*... as [a] condition of the synthesis of all *apprehension*' (B161), and he concludes on this basis that 'all synthesis, through which perception itself becomes possible, stands under the Categories' (B162). Kant seems to claim at this point that the Categories serve as conditions of the possibility of the synthesis of apprehension. And if the Categories serve as conditions of the possibility of the synthesis of

apprehension then it follows, presumably, that they serve as conditions of the possibility of whatever the synthesis of apprehension serves as a condition of the possibility of. But the synthesis of apprehension serves as a condition of the possibility of empirical intuition. So it follows that the Categories serve as conditions of the possibility of empirical intuition. As Kant puts it later on in §26:

since all possible perception depends on the synthesis of apprehension, but the latter itself, this empirical synthesis, depends on the transcendental one, thus on the categories, all possible perceptions, hence everything that can ever reach empirical consciousness, i.e. all appearances of nature, as far as their combination is concerned, stand under the Categories... (B164- 165)

Of course, these pieces of text are not at all clear at first glance, and they require careful and detailed analysis if they are not to be misunderstood. However, I do think that there is good reason to doubt that they provide obvious support for ascription to Kant of the claim that the Categories serve as conditions of the possibility of empirical intuition. Take the last few lines that I quoted. Kant's claim here is that 'all possible perceptions... as far as their combination is concerned, stand under the Categories' (B164-165). One important point here concerns the qualification 'as far as their combination is concerned'. In §15 of the second edition Deduction of the Categories, Kant ascribes the act of combination to the faculty of understanding. He writes that 'the combination... of a manifold in general.... is an act of the understanding...' (B130). This makes it reasonable to interpret Kant's claim as follows: All possible perceptions stand under the Categories insofar as we exercise our faculty of understanding in

relation to them. We can also put pressure on the assumption that what Kant means by a perception is an empirical intuition. After all, Kant tells us that a perception becomes possible by means of ‘the composition of the manifold in an empirical intuition’. If this composition is an act of the understanding, then perhaps what Kant means by a perception is a sort of representation that comes about when we exercise our faculty of understanding in relation to an empirical intuition.

In the first half of this section I argued on the basis of what Kant claims in the Transition to the Transcendental Deduction of the Categories that what he thinks we need to do if we want to show that the Categories are objectively valid is to show that an object must conform to the Categories if it is to count as an object of thought – that is to say, if we are to think of, or form a concept of, this object. My aim in the second half of this section has been to examine some of the textual considerations that have been adduced in support of the claim that this is not what Kant actually does in his Deduction of the Categories. I have considered some evidence that Kant’s strategy is to show that the Categories serve as conditions of the possibility of consciousness in general, and I have considered some evidence that his strategy is to show that the Categories serve as conditions of the possibility of empirical intuition. What I have tried to show is that this evidence is not strong enough to overturn the presumption in favour of the interpretation that I defended in the first half of the section. In §2.2, I shall address another important question: If Kant’s strategy in his Deduction of the Categories to show that an object must conform to the Categories if it is to count as an object of thought or understanding, what exactly does he mean by an object of thought or understanding?

§2.2 Empirical cognition and the act of judgment

What exactly does Kant mean when he refers to the thinking that is an essential ingredient in a state of empirical cognition? What exactly does he mean by the use of our faculty of understanding? It is crucial for us to get clear about the answer to these questions before we look in any detail at the substance of Kant's Deduction of the Categories. We cannot hope to make sense of Kant's analysis of the faculty of thinking or understanding if we aren't sure exactly what this analysis is an analysis of.

Here is one way to understand Kant's concept of thinking. When Kant refers to the faculty of understanding, what he has in mind is just our ability to think that such and such is the case. What he means by the faculty of understanding is our ability to think that, for example, there are clouds over Oxford, or that the sun will rise tomorrow in the East, or that England will retain the Ashes. There is a straightforward argument for this way of understanding Kant's references to thinking and understanding in the Deduction of the Categories. Another claim that Kant makes in the first chapter of the *Analytic of Concepts* is this:

We can... trace all actions of the understanding back to judgments, so that the understanding in general can be represented as a faculty for judging.
(A69/B94)

What this suggests is that according to Kant every act of thinking is an act of judging. The understanding is, Kant claims, 'a faculty for thinking' (A69/B94), and '[thinking] is cognition through concepts' (A69/B94), yet 'the understanding can make no other use of... concepts than that of judging by means of them' (A68/B93). Kant also writes in a

later work that '[the] understanding shows its power solely in judgments...' (Ak XX:271). So someone engages in an act of thinking, which is to say, she makes use of her faculty of understanding, if and only if she engages in an act of judging. But to engage in an act of judging is surely just to think that such and such is the case. I can wonder whether England will retain the Ashes, I can entertain the thought that England will retain the Ashes. But I can also judge that England will retain the Ashes. I can also judge that they might retain the Ashes. In making these judgments, I come to a decision about the way things are. I think to myself that such and such is the case.

Two points should be noted about this reading of Kant's concept of thinking or understanding. The first is that if it is correct then not every act of the mind that we would be prepared to call an act of thinking counts as thinking in the sense that matters to Kant in his Deduction of the Categories. I have just contrasted a case where someone judges that England will retain the Ashes with a case where someone wonders whether England will retain the Ashes or entertains the thought that England will win the Ashes. Now, it is entirely reasonable to say that wondering whether such and such is the case or entertaining the thought that such and such is the case is a way of thinking. If I have been wondering whether or not England will win the Ashes, I have been thinking about something. And this is true even if I have not reached any decision - that is to say, even if I have not arrived at the judgment that England will, or will not, win the Ashes. But I have not engaged in an act of thinking on the reading of Kant's concept of thinking that I have just introduced. For it is precisely a judgment or decision about the way things are that is required for thinking on this reading. And wondering whether or not England will win the Ashes is not the same as making a judgment about the way things are.

The second point that should be noted about this reading is as follows: If I make the judgment that such and such is the case, that is to say, if I decide that such and such is the case, it is quite possible that my decision will turn out to be the wrong one. My judgments can be – and very often are – incorrect. I can think that such and such is the case even when such and such is not in fact the case. So if this first reading of Kant's concept of thinking or understanding is correct then thinking in the sense that Kant has in mind, is not factive. There is no requirement that what we think or understand should turn out to be the case.

There is substantial textual evidence to support the claim that a judgment in the sense that Kant has in mind in the *Critique* can be false or incorrect. One of the claims that Kant makes in both the *Critique* and the *Prolegomena* is that it is only insofar as we make a judgment about objects that our state of mind can be assessed for truth or falsehood. At the start of the Transcendental Dialectic, Kant writes that

truth and illusion are not in the object insofar as it is intuited, but in the judgment about it insofar as it is thought. Thus it is correctly said that the senses do not err, yet not because they always judge correctly, but because they do not judge at all. Hence truth, as much as error... are to be found only in judgments, i.e. only in the relation of the object to our understanding. (A293/B350)

Here Kant seems to draw a contrast between the intuitions that we acquire by means of our senses and the judgments that we make by means of our understanding. The contrast is that whereas our judgments can be true or false – that is to say, correct or

incorrect – the intuitions of our senses can be neither true nor false. Crucially, Kant seems to allow here for incorrect or erroneous judgments made by means of our understanding. And this seems to count in favour of our first reading of Kant's concept of thinking. Kant makes the same point at the end of the first section of the *Prolegomena* (Ak IV:290-291), where he writes that there is 'neither falsehood nor truth' in the representation of the senses, and that there can be falsehood only when we 'render an objective judgment from the appearance' (Ak IV:291).

One commentator who understands Kant's concept of thinking or understanding in this way is Van Cleve. According to Van Cleve, Kant adopts two very different strategies in the course of his Deduction of the Categories. Kant's eventual strategy is to construct an argument for the objective validity of the Categories that starts from the thin conception of experience as consciousness that I mentioned in §2.1. But before Kant realizes that this is the only way to deduce the Categories, he tries out a different strategy that involves arguing for the objective validity of the Categories on the basis of a conception of experience as involving two elements, viz. 'an intuitional component (e.g. sensing an expanse of red)' and also 'a conceptual component (e.g., believing that the concept apple applies to what one is experiencing)' (1999, p.77). When Kant sets out to deduce the Categories in this way, according to Van Cleve, his aim is to show that they serve as conditions of the possibility of the 'conceptual component' of experience. Now, Van Cleve reasons that since, according to Kant 'to use any concepts is to judge' (1999, p.77), the conceptual component of a state of experience must be a state of judging, or believing, that such and such is the case (1999, p.76). He remarks that '[it] does not seem to matter to the argument whether the belief is either justified or true' (1999, p.77-78).

Allison is another commentator who interprets Kant's concept of thinking or understanding in this way. Allison takes Kant to identify the faculty of thinking or understanding – or, as Allison likes to say, 'discursive cognition' – with the faculty of judgment (2004, p.82). And he clearly thinks that a judgment in the sense that Kant has in mind in the *Critique* is something that can be either true or false. In relation to Kant's claim in the second edition Deduction of the Categories that all judgments have the property of objective validity, he writes:

If this claim is to make any sense, objective validity cannot be equated with truth; otherwise, Kant would be committed to the absurdity that every judgment is true, simply in virtue of being a judgment. (2004, p.88)

Allison reads Kant's claim that all judgments are objectively valid as the claim that 'every judgment makes a claim to truth' (2004, p.88). And this seems to confirm that it is the first reading of Kant's concept of thinking that Allison has in mind. For what makes a case in which I think that England will retain the Ashes different from a case in which I wonder whether England will retain the Ashes or just entertain the thought that England will retain the Ashes is that only in the first case is there a claim to truth.

However, there is also an objection to this interpretation of Kant's concept of the thought of an object. This objection has to do with Kant's concept of a state of cognition. There is good evidence that in order for someone to achieve cognition of an object in the sense that Kant has in mind in the *Critique*, not only must he make a judgment about an object, but this judgment must also be true. Consider, for example, Kant's famous Refutation of Idealism. Kant's aim in the Refutation of Idealism is to

prove that we really do have experience – that is to say, empirical cognition – of objects in space. As Kant puts it,

The required proof must... show that we have experience, and not merely imagination of outer things... (B275),

What Kant means here by ‘outer things’ is just objects in space. Now, the idealist that Kant has set out to refute in this section of the *Critique* is someone who either doubts or denies the existence of objects in space. What this idealist doubts or denies is that any of the judgments that commit us to the existence of objects in space – say, the judgment that there is a book on the table in front of me, or the judgment that there are five coins in my pocket – are true. What the idealist does not doubt or deny is that we really do make judgments that commit us to the existence of objects in space. So when Kant writes that his Refutation of Idealism must show that we have experience of objects in space – that is to say, empirical cognition of objects in space – he cannot mean simply that it must show that we really do make judgments that commit us to the existence of objects in space. This would not suffice to refute the idealist. It would amount to proving something that the idealist never doubted. What Kant means when he refers to a state of experience or empirical cognition of objects in space must be a state that involves judging, and, moreover, judging correctly, that there are objects in space. And so it seems that what Kant means by the cognition of an object in the Refutation of Idealism is a factive state of mind.

There is also what appears to be direct textual evidence for this claim. Kant makes it clear at various points in the *Critique* that he thinks that a state of cognition is

as such objectively valid. For example, he writes in the second edition Deduction of the Categories that ‘the unity of consciousness is that which alone constitutes the relation of representations to an object, thus their objective validity, and consequently that which makes them into cognitions’ (B137). This implication here is that it is the fact that a representation has objective validity that makes it count as a state of cognition. However, there are also places in Kant’s Deduction of the Categories where Kant clearly identify the concept of objective validity with the concept of truth. In a passage that I have already quoted, he writes:

Actual experience... contains in the last and highest (of the merely empirical elements of experience) concepts that make possible the formal unity of experience and with it all objective validity (truth) of empirical cognition. (A124-125)

And he writes in the Transcendental Doctrine of Method that

only through the fact that an object is determined for the concept [of that which happens in general] by means of the law of causality does the represented occurrence have objective validity, i.e. truth. (A788/B816)

If Kant thinks that a representation counts as objectively valid insofar as it is true then it seems that what he has in mind when he tells us that cognition is as such objectively valid is the idea that cognition is as such true.

However, as I mentioned in §2.1, Kant thinks that a state of cognition in general

has two components. It involves, first, the intuition of an object, and second, the thought or concept of an object. As Kant writes,

cognition has two parts: first, the concept, through which in general an object is thought... and second, the intuition, through which it is given...
(B146, also A92-93)

Now, I have said very little so far in this thesis about Kant's concept of an intuition. But it is reasonable to assume that according to Kant someone achieves an empirical intuition of an object – that is to say, the sort of intuition of an object that is required for empirical cognition – insofar as he sees or touches or in some other way perceives an object by means of his senses. Suppose that this is correct. And suppose that the interpretation of Kant's concept of thinking or understanding that I defended in the first half of this section is also correct. In that case, so it may be argued, Kant must think that a state of cognition is the state that a person is in when he both sees or touches or in some other way perceives an object by means of his senses and also makes a judgment about this object – say, the judgment that this object is round or that it is sitting on top of a desk. But in that case a state of cognition cannot be a factive state of mind. For it is entirely possible that someone should make a false judgment about an object that he sees or touches etc. So since Kant thinks that cognition is a factive state of mind, my interpretation of Kant's concept of thinking or judging must be false.

In support of this objection we can consider a famous example that Kant uses in his logic lectures to illustrate the difference between the mere intuition of an object and the concept or thought of an object. Kant writes:

E.g. [if] a wild man sees a house far away, the use of which he does not know, he has before him in his representation, to be sure, exactly the same object as another [man], who cognizes it determinately as a habitation established for men. But as to form this cognition of one and the same object differs in each case. In the one it is mere intuition, in the other intuition and concept at the same time. (Ak IX:33)

Both the wild man and the civilized man have achieved an intuition of the house. For they can both see this house. However, the civilized man has achieved something more than the wild man. He has also formed a concept of this house. And the result of this is that he has achieved cognition of the house. However, what Kant seems to mean by this is that the civilized man has recognized that the object he can see is a house, which is to say, that it is 'a habitation established for men'. The wild man's representation counts as a mere intuition because although this man can see the house, he doesn't know what it is. It seems to be crucial to Kant's example that the civilized man has not only judged that the object he can see is a house, but has judged correctly that the object he can see is a house. And so it is tempting to conclude that what it means for someone to form a thought or concept of an object is not just for him to make a judgment about this object. This judgment must also be true.

The trouble with this objection is that it rests on a piece of fallacious reasoning. Kant certainly thinks that the intuition of an object and the thought of an object are the two parts or components of a state of cognition. And there is good evidence that Kant thinks that a state of cognition is a factive state of mind. But it doesn't follow

from this that one of the two states of mind out of which a state of cognition is composed must itself be a factive state of mind. For it is entirely possible that a state of cognition is the state of the mind that arises when two parts or components, viz. the intuition of the object and the thought of the object, are combined in the right way. To put this reply a little differently, the objection rests on a misinterpretation of Kant's claims about the state of experience. Kant certainly claims that a state of cognition has two parts - namely, the intuition of an object and the thought of an object. But it doesn't follow from this claim that if someone has achieved both the intuition of an object and the thought of an object - even the thought of the very same object - then he has achieved a state of cognition.

There is some direct textual evidence that according to Kant more is required for us to achieve a state of empirical cognition than empirical intuition and the thought of an object. In his letter of 20th January 1792 to Jacob Sigismund Beck he writes:

I begin [my lectures] by defining "experience" in terms of empirical cognition.... Two sorts of representations are needed for cognition: 1) intuition, by means of which an object is given, 2) concept, by means of which it is thought. To make a single cognition out of these two pieces of cognition a further activity is required: the composition of the manifold given in intuition in conformity with the synthetic unity of consciousness, which is expressed by the concept. (Ak XI:315-316)

Evidently, Kant thinks that the empirical intuition of an object and the thought or concept of this object are individually necessary but not jointly sufficient for a state of

empirical cognition. He thinks that more is required for a state of cognition than empirical intuition and the thought or concept of an object. What is required is an additional act of the mind by means of which these two 'pieces of cognition' are combined together. I conclude that even if Kant defines a state of cognition as a factive state of mind, this poses no problems for the claim that what Kant means by a case in which we form the thought of an object is a case in which we make a judgment about an object.

§2.3 Kitcher on Kant's concept of empirical cognition

So far in this chapter, I have argued that Kant's strategy in his Deduction of the Categories is to show that there is a connection between the Categories and the use of our faculty of understanding – that is to say, a connection between the Categories and the thought of an object – and I have defended an interpretation of Kant's concept of the thought of an object. I have argued that someone makes use of his faculty of understanding in the sense that Kant has in mind in his Deduction of the Categories insofar as he makes a judgment about an object – say, the judgment that there are clouds over Oxford or that the sun will rise tomorrow. My aim in this section is to discuss a different interpretation of Kant's strategy in his Deduction of the Categories, recently defended by Kitcher (2011).

Kitcher writes:

The cognition that is the topic of [Kant's Deduction of the Categories] is not merely empirical... It is also conceptual and for Kant that means that it

is a matter of 'marks' or grounds of cognition that are recognised as such.

(2011, p.121)

Kitcher calls the sort of cognition that Kant has in mind in his Deduction of the Categories 'rational empirical cognition' (2011, p.121). The key to her reading is the concept of a mark. According to Kitcher, a concept F serves as a mark for another concept G insofar as two conditions are satisfied. The first condition is that the concept G contains the concept F. F is, as Kitcher writes, 'part of the content of' G (2011, p.120). To take some of Kitcher's own examples, the concept of a body contains the concept of what is impenetrable and the concept of what is extended, and the concept of copper contains the concept of what is metallic. The second condition that must be satisfied if a concept F is to serve as a mark for another concept G is that someone takes the fact that an object falls under the F as a reason to believe that this object falls under G:

for a representation to be a mark, it must be considered as such...
[cognizers] must recognize that a mark - say, 'impenetrable' - is a partial ground for their application of the concept 'body'. It is part of why they call something a 'body' or part of what they presuppose in calling something a 'body'. (2011, p.120)

Suppose I arrive at the judgment that the object I can see in the distance is a pub. And suppose that I make this judgment on the grounds that the object I can see in the distance has a door above which hangs a colourful sign. That is to say, I judge that the

object I can see falls under the concept of what has a door above which hangs a colourful sign, and I infer from this that the object I can see also falls under the concept of a pub. I draw this conclusion because part of what I think of when I think of a pub is an object that has a door above which hangs a colourful sign. In that case, my judgment counts as a rational judgment in the sense that Kitcher has in mind. And if my judgment is also empirical – as it is in this case – then I have achieved what Kitcher calls ‘rational empirical cognition’. If Kitcher is right then someone has achieved experience – that is to say, she has achieved rational empirical cognition – only if she has judged that an object falls under a concept, she has judged that the object falls under the concept on the grounds that it falls under some other concept, and she is aware that part of the reason why she has judged that the object falls under the first concept is that she has judged that it falls under the second concept.

Kitcher thinks that Kant’s strategy in his *Deduction of the Categories* is to investigate the conditions that must be satisfied if someone is to achieve a state in which she is conscious of making a judgment about an object on the basis of reasons. It is precisely this consciousness of the grounds of our cognition that Kitcher thinks serves as the crucial connection between a state of empirical cognition and the *Categories*. Kitcher does not think that Kant’s strategy in his *Deduction of the Categories* is just to investigate the conditions that must be satisfied if someone is to make a judgment about an object – say, the judgment that there are clouds over Oxford or that the sun will rise tomorrow morning. If she is right then the interpretation of Kant’s strategy that I have defended in this chapter is mistaken.

Why does Kitcher ascribe this conception of experience as rational empirical cognition to Kant? She refers to the following passage from the Jaesche transcript of

Kant's logic lectures:

From the side of the understanding, human cognition is discursive, i.e. it takes place through representations which take as the ground of cognition that which is common to many things, hence through marks as such. Thus we cognize things through marks and that is called cognizing. A mark is that in a thing which constitutes a part of the cognition of it, or, what is the same, a partial representation, insofar as it is considered as ground of cognition of the whole representation. All our concepts are marks, accordingly, and all thought is nothing other than a representing through marks. (Ak IX:58)

It is clear from this passage that according to Kant discursive cognition – that is to say, the sort of cognition that involves the use of our faculty of thinking or understanding – is cognition by means of marks (Ak XVI:298, Ak XVI:300), where a mark is a representation that is a part of another representation. Kitcher also thinks that according to Kant a mark is something general – a representation that is a part of a number of distinct representations (cf. Smit 2000). And she takes it that what Kant means in this passage by a case in which I achieve cognition by means of a mark is a case in which I judge that a representation – say, the representation of the man I can see in front of me – falls under one concept – say, the concept of being a man – on the grounds that this representation has as a part a representation that is common to a number of other representations – say, the representation of having eyes.

There are, however, some other ways to understand Kant's claim that discursive

cognition is cognition by means of a mark. One possible reading is suggested by Kant's claims about the nature of judgment in the *Critique*. Here Kant writes:

Judgment is... the mediate cognition of an object, therefore the representation of a representation of it. In every judgment there is a concept that holds of many, and that among this many also comprehends a given representation, which is then related immediately to the object. So in the judgment, e.g. all bodies are divisible, the concept of the divisible is related to various other concepts; among these, however it is here particularly related to the concept of body, and this in turn is related to certain appearances that come before us. These objects are therefore mediately represented by the concept of divisibility. (A68-69/B93-94)

Kant's claim here seems to be that what makes a judgment count as a mediate or indirect form of cognition is the fact that in a judgment, instead of ascribing a concept - in Kant's example, the concept of divisibility - directly to an object - say, to the book on the desk in front of me - we ascribe it to another concept - in Kant's example, the concept of body - which concept is already related to the object. If this is what Kant means then one way of understanding what he means by cognition by means of a mark would be to say that it is cognition where instead of thinking of a particular object and thinking that this object falls under some concept, we think of a concept under which the particular object falls and then think that there is a connection between this concept and some other concept. The trouble with this reading is that it assumes that what may be an incidental feature of Kant's example is essential to the point he is

making, viz. that the concept of divisibility is related to another concept, rather than to, say, an intuition of an object. But there is reason to think that this is not Kant's intention. Just before claiming that judgment is 'therefore the mediate cognition of an object', Kant writes:

Since no representation pertains to the object immediately except intuition alone, a concept is thus never immediately related to an object, but is always related to some other representation of it. (A68/B93)

This suggests that Kant allows for a judgment where we start with the mere intuition of an object – say, the intuition of the book on the desk in front of me – and think of the connection of this intuition to a concept – say, the concept of a body. Such a judgment counts as a mediate cognition of an object in virtue of the fact that the object is cognized by means of the representation that there is a connection between one representation of this object – that is, the intuition of the book on the desk in front of me – and another representation – in this case, the concept of a body.

If this is right then we can construct a simpler interpretation of Kant's claim that discursive cognition is cognition by means of a mark. On this reading, what Kant means by a case in which I achieve cognition by means of a mark is just a case in which I judge that a representation – say, the representation of the man I can see in front of me – has as a part a representation that is common to several other representations – say, the representation of having eyes. Discursive cognition requires nothing more than the judgment that an object has some mark or characteristic – that is to say, the application to this object of some concept or other. Even an arbitrary judgment counts

as cognition by means of a mark on this reading, since it is cognition that involves a concept, and a concept is a mark – that is to say, a representation of what is common to many things. Since this interpretation fits far better than Kitcher’s interpretation with Kant’s claims about judgment in the *Critique*, there is reason to doubt her claim that thinking in the sense that Kant has in mind is a matter of forming a rational judgment. And so there is reason to doubt Kitcher’s claims about the strategy of Kant’s Deduction of the Categories.

§2.4 Empirical cognition and empirical knowledge

Some prominent commentators have assumed that what Kant means by a case in which someone has achieved empirical cognition of objects is a case in which he has achieved empirical knowledge of objects. Ameriks, for example, writes at the start of his famous article on the aim and structure of Kant’s Deduction of the Categories that

it is necessary and profitable to understand the deduction as moving from the assumption that there is empirical knowledge to a proof of the preconditions of that knowledge. (1978, p.273)

Similarly, Guyer remarks at one point in his book on the first *Critique* that, in the sense that Kant has in mind,

experience may be equated with knowledge of objects given by perception. (1987, p.85)

There are many passages of text that appear to count in favour of this interpretation. Consider once again the example of the wild man and the civilized man:

E.g. [if] a wild man sees a house far away, the use of which he does not know, he has before him in his representation, to be sure, exactly the same object as another [man], who cognizes it determinately as a habitation established for men. But as to form this cognition of one and the same object differs in each case. In the one it is mere intuition, in the other intuition and concept at the same time. (Ak IX:33)

Only the civilized man has achieved empirical cognition of the object that he can see. And this is because he has recognized what it is that he can see, viz. a house. The wild man has fallen short of a state of cognition. And that is because he doesn't know that what he can see is a house. So it seems that what Kant means when he claims that someone lacks empirical cognition of an object is that he lacks empirical knowledge of an object. Remember also my discussion of Kant's Refutation of Idealism in §2.2. Here Kant's aim is to respond to someone who doubts that it is possible for us to achieve a particular sort of cognition - namely, cognition of objects in space. And it is natural to interpret this doubt about the possibility of cognition of objects in space as a doubt about the possibility of knowledge of objects in space.

However, there are also some possible objections to the claim that what Kant means by empirical cognition is empirical knowledge. One of these objections has to

do with the connection between a state of knowledge and justification. If Ameriks and Guyer are right then Kant thinks that I have achieved a state of empirical cognition only if I am justified in thinking that such and such is the case. For I know that such and such is the case only if I am justified in thinking that such and such is the case. I must have reasons or grounds for thinking as I do. The trouble with this implication of the interpretation defended by Ameriks and Guyer is that there is evidence that considerations about my reasons for thinking as I do are irrelevant to the question whether or not I have achieved a state of cognition in the sense that Kant has in mind. These considerations are relevant, according to Kant, only if we are asking a very different question.

To see why this is true it is necessary to look at a passage of text towards the end of the *Critique* in the Transcendental Doctrine of Method. Here Kant distinguishes three states of mind in which we can find ourselves. He writes:

Taking something to be true... has the following three stages... having an opinion, believing, and knowing. Having an opinion is taking something to be true with the consciousness that it is subjectively as well as objectively insufficient. If taking something to be true is only subjectively sufficient and is at the same time held to be objectively insufficient, then it is called believing. Finally, when taking something to be true is both subjectively and objectively sufficient it is called knowing. Subjective sufficiency is called conviction (for myself), objective sufficiency, certainty (for everyone).
(A822/B852)

I take it that what Kant has in mind here is the following distinction: If I think that such and such is the case – if, say, I think that there are clouds over Oxford – one question that I can ask is why I think that this is the case. As Kant puts it at one point, I can ask what are the ‘subjective causes’ of this ‘occurrence in [my] understanding’ (A820/B848). But I can also ask whether the reasons why I think that this is the case are also reasons for thinking that this is the case. I can ask whether, in light of the considerations that have in fact caused me to think that this is the case, I really ought to think that this is the case. Now suppose that I think that, say, the Birmingham accent is nicer than the Newcastle accent. And suppose that I am not even conscious of the reasons why I think that this is true, let alone conscious that these considerations justify me in thinking that this is true. In that case, my taking this to be true is a case of mere opinion in Kant’s sense. If, however, I am perfectly conscious of the reasons why I think that the Birmingham accent is nicer than the Newcastle accent, but am still not conscious that these considerations justify me in thinking that this is true, my taking this to be true is a matter of believing. In Kant’s terms, my taking this to be true is subjectively sufficient yet objectively insufficient. Finally, if I am conscious both of the reasons why I think that the Birmingham accent is nicer than the Newcastle accent and that these reasons are sufficient to justify thinking that the Birmingham accent is nicer than the Newcastle accent, my taking this to be true is a case of knowing in Kant’s sense.

If my reading of Kant’s threefold distinction between knowing, believing, and having an opinion is correct, there is reason to question the claim that what Kant means by a state of empirical cognition is a state of empirical knowledge. For the concept that Kant picks out by the German word ‘*Wissen*’, and which I have translated

by the English word 'knowledge', seems to be very close to our ordinary concept of knowledge. Yet there is no evidence that Kant treats the word 'Wissen' as a synonym for 'Erkenntnis', which is to say, cognition. Kant appears to treat these as entirely distinct concepts. He seems to think that while considerations about our grounds or reasons for thinking that such and such is the case are relevant when addressing the question whether our state of mind counts as a state of knowing, believing, or merely having an opinion, they are not relevant when addressing the question whether our state of mind counts as a state of cognition.

Van Cleve has raised a different objection to the claim that what Kant means by a state of empirical cognition is a state of empirical knowledge (1999). This objection is based on what Kant tells us about the concept of truth in the introduction to the *Transcendental Logic*. Here Kant refers to 'the old and famous question with which the logicians were to be brought to such a pass that they must either fall into a miserable circle or else confess their ignorance' (A57-58/B82), viz. the question: What is truth? Kant then writes:

The nominal definition of truth, namely, that it is the agreement of cognition with its object, is here granted and presupposed, but one demands to know what is the general and certain criterion of the truth of any cognition. (A58/B82)

Most commentators assume that this nominal definition of truth is one to which Kant is himself committed. This isn't obvious on the basis of the lines I have just quoted. Kant's claim here is that the definition of truth as 'the agreement of cognition with its

object' is presupposed by those who try to embarrass logicians by asking them what truth is. Kant is similarly non-committal later on in the same section of the *Critique*, where he writes:

If truth consists in the agreement of a cognition with its object then this object must thereby be distinguished from others, for a cognition is false if it does not agree with the object to which it is related even if it contains something that could well be valid of other objects. (A58/B82)

However, there is sufficient evidence to ascribe this nominal definition of truth to Kant himself. In the Second Analogy of Experience he writes:

One soon sees that, since agreement of the cognition with the object is truth, only the formal conditions of empirical truth can be enquired into here (A191/B236, see also A196-197/B158).

And in the Appendix to the Transcendental Dialectic he writes:

just as the latter [the categories] lead to truth, i.e. to the agreement of our concepts with the object [check number agreement], the former [the transcendental ideas] effect a mere but unavoidable illusion (A642/B670).

Van Cleve argues that Kant's acceptance of this definition of truth creates a problem for those who interpret his concept of empirical cognition as the concept of empirical

knowledge. He writes that ‘if by [cognition in his definition of truth] he [i.e. Kant] meant justified true belief, this definition would be circular’ (1999, p.77). The idea here is this: If what Kant means by cognition is knowledge then cognition is a state of the mind that requires a true and justified belief. But in that case Kant has subscribed to a definition of truth as the agreement or conformity of a particular sort of true state of mind with its object. And so the concept of truth features in Kant’s definition of truth itself, which means that the definition is circular.

I propose that these objections give us good reason to reject the claim that what Kant means by a state of empirical cognition is a state of empirical knowledge. We should not assume that Kant’s strategy in his Deduction of the Categories is to investigate the conditions that must be satisfied if we are to achieve a state of empirical knowledge. On the contrary, what Kant thinks we need to do in order to carry out a deduction of the Categories is to show that the conformity of an object to the Categories is necessary if we are to exercise our faculty of understanding in relation to this object, which is to say, if we are to make a judgment about this object.

§3 Kant's Principle of the Original Synthetic Unity of Apperception

So far in this thesis I have discussed the aim and the strategy of Kant's Deduction of the Categories. In §1, I argued that what Kant is trying to do in his Deduction of the Categories is to explain how it is that the Categories or Pure Concepts of the Understanding are able to serve as representations of objects. I showed that a concept serves as a representation of an object in the sense that Kant has in mind insofar as it stands in a relation of necessary conformity to an object. Finally, I argued that Kant thinks that the Categories stand in a relation of necessary conformity to objects because he thinks that it is necessary for an object to conform to the Categories if it is to count as an object of experience. In §2, I argued that Kant thinks that it is necessary for an object to conform to the Categories if it is to count as an object of experience precisely because he thinks that it is necessary for an object to conform to the Categories if it is to count as an object of thought or understanding. And I explained that what Kant means by a case in which someone has achieved the thought of an object – that is to say, what he means by a case in which someone has exercised his faculty of understanding in relation to an object – is a case in which this person has made a judgment about an object – say, the judgment that there are clouds over Oxford or that England will win the forthcoming Ashes series. If what I claimed in §1 and §2 is correct, Kant's aim in his Deduction of the Categories is to establish that it is necessary for an object to conform to the Categories or Pure Concepts of the Understanding if we are to make a judgment about it. He thinks that if we investigate the conditions

that must be satisfied in order for someone to make a judgment about an object, we will find that these conditions include the conformity of this object to the Categories.

In this chapter, I focus on what is undoubtedly the key claim in Kant's argument – namely, the claim that he calls the Principle of the Original Synthetic Unity of Apperception (B137), or the Principle of the Necessary Unity of Apperception (B135), or the Principle of the Transcendental Unity of Apperception (B142). The first conclusion that I reach in this chapter is that the Principle of the Original Synthetic Unity of Apperception is a claim about the conditions that must be satisfied if we are to make a judgment about an object – that is to say, a claim about the conditions that must be satisfied if we are to make use of our faculty of understanding. The second conclusion that I reach in this chapter is that the Principle of the Original Synthetic Unity of Apperception consists in the claim that if we are to make a judgment about an object – if, that is, we are to make use of our faculty of understanding – then those conditions must be satisfied that must be satisfied if we are to achieve a very special sort of consciousness – in particular, the consciousness that we have connected a number of our representations in a way that it was necessary for us to connect them. My third and final conclusion in this chapter is that Kant identifies a Pure Concept of the Understanding with the consciousness that we have connected a number of our representations in a way that it was necessary for us to connect them.

If the interpretation that I defend in this chapter is correct then we can make very good sense of Kant's famous remark in the Preface to his *Metaphysical Foundations of Natural Science* that the solution of the problem 'how experience is... possible by means of [the] Categories' (Ak IV:475),

can almost be accomplished through a single inference from the precisely determined definition of a judgment in general... (Ak IV:475-476)

I believe that Kant's Principle of the Original Synthetic Unity of Apperception can be understood precisely as a claim about what the act of judgment consists in – that is to say, about the precisely determined definition of a judgment in general. It is the claim that, as Kant puts it in §19 of the second edition Deduction of the Categories,

a judgment is nothing other than the way to bring given cognitions to the objective [i.e. original synthetic] unity of apperception. (B141)

All that is required to complete the Deduction of the Categories once this claim has been established is to note that it is by means of the Categories or Pure Concepts of the understanding that 'given cognitions' are brought to 'the objective' – or, in other words, the original synthetic – unity of apperception.

At the start of §3.1, I use what Kant tells us in §17 and §19 of the second edition Deduction of the Categories to justify my conclusion that Kant's Principle of the Original Synthetic Unity of Apperception is a claim about the conditions that must be satisfied if we are to make a judgment about an object – if, that is, we are to exercise our faculty of understanding. I also show that this conclusion is supported by textual evidence in the first edition Deduction of the Categories. Finally, I show that this interpretation is consistent with Kant's claims in the passage of text that is more closely associated than any other with the Principle of the Original Synthetic Unity of Apperception – that is, the start of §16 of the second edition Deduction of the

Categories. In the rest of the chapter, I turn to the question how Kant's concept of the original synthetic unity of apperception should be understood. In §3.1, I focus on Kant's concept of apperception in general, and his concept of original apperception in particular. I argue that what Kant means by apperception is the consciousness of a state of mind as one that serves as a representation of an object, and that what he means by original apperception is the consciousness of a state of mind as one that serves as a representation of an object *a priori* – that is to say, in virtue of the fact that the object conforms to the state of mind. In §3.2 I turn to Kant's concept of an act of synthesis. I show that an act of synthesis is an operation of the mind by means of which it forms a connection between several of its representations. And I explain how exactly this claim should be understood. Finally, in §3.4, I use what I have established in §3.1 and §3.2 to complete my interpretation of the Principle of the Original Synthetic Unity of Apperception. And I show that there is a short path between this principle and the claim that there is a connection between the Categories or Pure Concepts of the Understanding and the use of our faculty of understanding.

§3.1 The original synthetic unity of apperception and the act of judgment

The title of §17 of the second edition Deduction of the Categories is:

The principle of the synthetic unity of apperception is the supreme principle of all use of the understanding. (B136)

The section begins:

The supreme principle of the possibility of all intuition in relation to sensibility was... that all of the manifold of the same [i.e. the manifold of intuition] stand under the formal conditions of space and time. The supreme principle of the very same [that is to say, the supreme principle of the possibility of all intuition] in relation to understanding is that all of the manifold of intuition stand under conditions of the original synthetic unity of apperception. The manifold representations of intuition stand under the first [principle] insofar as they are given to us, [and] under second [principle] insofar as they must be able to be combined in one consciousness, for without that nothing can be thought or cognized thereby, because the given representations do not have the act of apperception *I think* in common... (B136-137)

Remember that in the Transition to the Transcendental Deduction of the Categories Kant specifies two conditions that must be satisfied if someone is to achieve a state of empirical cognition. First, this person must acquire an empirical intuition of an object, and second, she must think of, or form a concept of, the object of this empirical intuition (A92-93/B125, A93/B126). To put the point another way, this person must make use of her capacity for sensibility, and she must make use of her faculty of understanding. Now, the principle that Kant mentions in the first sentence of §17 is a claim about the conditions that must be satisfied by an intuition if it is to be one that is acquired by means of our capacity for sensibility. Specifically, it is the claim that this

intuition must be an intuition of an object in space or in time. The principle that Kant mentions in the second sentence of §17, by contrast, is a claim about the conditions that must be satisfied by an intuition if someone is to think of, or form a concept of, the object of this intuition. Specifically, it is the claim that this intuition must stand under, which I take to mean: it must satisfy, the conditions of the original synthetic unity of apperception. If intuitions fail to satisfy the conditions of the original synthetic unity of apperception then they do not – or, rather, they cannot – ‘have the act of apperception *I think* in common’, from which it follows that ‘nothing can be thought or cognized’ by means of them (B137). To put this point another way, Kant’s claim in the second sentence of §17 is that if someone is to make use of her faculty of understanding, she must acquire an intuition of an object that satisfies the conditions of the original synthetic unity of apperception. And this is the Principle of the Synthetic Unity of Apperception – or, as Kant calls it a few lines later, the Principle of the Original Synthetic Unity of Apperception (B137). We can conclude on this basis that Kant’s Principle of the Original Synthetic Unity of Apperception is a claim about the conditions that must be satisfied by an intuition if we are to make use of our faculty of understanding in relation to this intuition of an object.

This conclusion is corroborated by what Kant tells us in §19 of the second edition Deduction of the Categories. This is the section of the Deduction of the Categories where Kant comes closest to making good on his remark in the *Metaphysical Foundations of Natural Science* that the problem ‘how experience is... possible by means of [the] Categories’

can almost be accompanied through a single inference from the precisely determined definition of a judgment in general (an action through which given representations first become cognitions of an object). (Ak IV:475-476)

At the start of §19, Kant makes two criticisms of what he calls the standard logicians' definition of judging:

I have never been able to satisfy myself with the explanation that logicians give of a judgment in general – it is, they say, the representation of a relation between two concepts. Without quarreling here about the incorrectness of this explanation – that, at best, it fits only *categorical* [judgments], but not hypothetical and disjunctive judgments – I only remark that in what this *relation* consists is not here determined. (B140-141)

Kant's first criticism is that the logicians' definition of judgment is too narrow, since it fails to cover hypothetical judgments and disjunctive judgments. What Kant means by a categorical judgment is a judgment of the form: A is B. Judgments of this sort include the judgment that Socrates is mortal and the judgment that all men are mortal. What Kant means by a hypothetical judgment is a judgment of the form: If A then B. One example of a hypothetical judgment that Kant provides in the *Critique* is: If there is perfect justice then the persistently wicked will be punished (A73/B98). Finally, what Kant means by a disjunctive judgment is a judgment of the form: Either A or B or C etc. (and not A and B and C etc.). Kant's example of a disjunctive judgment is: The world exists either by sheer chance or by inner necessity or because of an external cause

(A74/B99). Now, Kant thinks that a hypothetical judgment consists in the representation of a relation between two judgments – say, between the judgment that there is perfect justice and the judgment that the persistently wicked will be punished. He thinks that a disjunctive judgment consists in the representation of a relation between two or more judgments. And so he rejects the definition of a judgment as the representation of a relation between two concepts.

Now, there is a natural objection to what Kant has said so far in §19. He has claimed that a hypothetical judgment consists in the representation of a relation between two judgments. If he is correct then what it is for me to judge that if there is perfect justice then the persistently wicked will be punished is for me to represent a relation between the judgment that there is a perfect justice and the judgment that the persistently wicked will be punished. But this suggests that I cannot judge that the persistently wicked will be punished unless I judge both that there is a perfect justice and that the persistently wicked will be punished. For it seems that I can represent a relation between these judgments only if I can represent the judgments themselves. And I can represent the judgments themselves only if I actually make these judgments.

To assess this objection it is necessary for us to bear in mind Kant's concept of the modality of judging. Kant thinks that a judgment can have problematic modality, assertoric modality, or apodictic modality (A70/B95). He writes:

Problematic judgments are those in which one regards the assertion or denial as merely possible (arbitrary). Assertoric judgments are those in which it is considered actual (true). Apodictic judgments are those in which it is seen as necessary. (A74-75/B100)

And he clearly thinks that it is crucial that when we form the hypothetical judgment that if there is perfect justice then the persistently wicked will be punished, the judgments that we represent as related to one another are problematic judgments. As he puts it himself,

the two judgments whose relation constitutes the hypothetical judgment... as well as those in whose reciprocal relation the disjunctive judgment consists... are all merely problematic. In the above example the proposition: there is a perfect justice is not said assertorically, but is only thought of as an arbitrary judgment that it is possible someone might assume, and only the implication is assertoric. (A75/B100)

It is not immediately clear how this response to the objection should be understood. Since the objection has little bearing on my aims in this chapter, I shall set it to one side.

The more important aspect of Kant's discussion in §19 as far as this thesis is concerned is his second criticism of the standard logicians' definition of judging. Kant remarks that 'in what this *relation* consists', which is to say, in what consists the relation between two concepts or two or more judgments that is represented in a judgment, 'is not here determined' (B141). This time Kant's objection is that the standard logicians' definition is too broad. Kant thinks that there is something distinctive about the relation between concepts or judgments that is represented in a judgment. And he thinks that an adequate definition of an act of judging must include an answer to the

question what this relation is. Such a definition must distinguish the sort of relation that is represented in a judgment from the sort of relation that is represented in other states of the mind. It is in precisely this sense that an adequate definition of judgment must determine the relation that is represented in an act of judgment.

Kant's own answer to the question 'in what this *relation* consists' is clear from the next few lines of §19. Here he writes:

If... I investigate more precisely the relation of given cognitions in every judgment, and distinguish it as belonging to the understanding from the relation according to laws of the reproductive imagination (which has only subjective validity), I find that a judgment is nothing other than the way to bring given cognitions to the *objective* unity of apperception. This is the point of the copula *is* in them... For this signifies the relation of them to the original apperception and the *necessary unity* of [this apperception]...
(B141-142)

Kant's claim in these lines is that what is distinctive about an act of judging is the fact that the cognitions – by which Kant means the concepts or judgments – that are united in an act of judging are related in some way to the '*objective* unity of apperception', or, in other words, to 'original apperception' and its '*necessary unity*'. In the title of §19, Kant puts the point as follows:

The logical form of all judgments consists in the objective unity of the apperception of the concepts contained therein. (B140)

Now, it is reasonable to assume that what Kant means by the necessary unity of original apperception in §19 is the same as what he means by the original synthetic unity of apperception in §17. If this is correct then Kant's claim in §19 is that there is a crucial connection between the act of judging and the original synthetic unity of apperception. However, Kant thinks that it is only by means of the act of judgment that we make use of our faculty of understanding. 'The understanding in general', he writes, 'can be represented as a faculty for judging' (A69/B94). The faculty of understanding 'shows its power solely in judgments' (Ak XX:271). So Kant's claim in §19 of the second edition Deduction of the Categories can also be expressed as a claim about our use of our faculty of understanding. It is the claim that what is essential to the use of our faculty of understanding is a relation between our concepts or judgments and the original synthetic unity of apperception. So here as in §17, the concept of the original synthetic unity of apperception is introduced into Kant's argument in the context of a claim about the faculty of understanding. The claim is that if we look carefully at what is involved in making use of our faculty of understanding, we find that it involves the original synthetic unity of apperception. And it is a corollary of this claim that if we are to make use of our faculty of understanding in relation to our intuition of an object then this intuition must satisfy the conditions of the original synthetic unity of apperception.

This interpretation of the place of Kant's concept of the original synthetic unity of apperception in his Deduction of the Categories is controversial. Compare the interpretations defended by two recent commentators. Guyer thinks that Kant adopts a number of different strategies in the course of his Deduction of the Categories. One

of these strategies is to show that the Categories are objectively valid by investigating the conditions that must be satisfied if we are to make use of our faculty of understanding – that is to say, if we are to make judgments. Guyer claims that

the tactic for the deduction in which Kant obviously felt the most confidence in the years immediately following the publication of the *Critique*... is the argument that the very idea of judgment itself implies knowledge of a necessary and universal connection, which can be explained only by the supposition that we have *a priori* knowledge of the categories in terms of which such judgments are made. (1987, p.94)

Crucially, the necessary and universal connection that Guyer thinks is mentioned in this line of argument has nothing to do with the original synthetic unity of apperception. Guyer clearly states that ‘the unity of self-consciousness as such plays no role whatever in the argument’ (1987, p.95). On the other hand, Guyer writes that

in each of the two editions of the *Critique* Kant does also attempt to deduce our *a priori* knowledge of the objective validity of the categories from a conception of apperception as a cognitively significant form of self-consciousness which is not, at least immediately, simply identified with judgments about objects claiming necessary truth. (1987, p.131)

This time the line of argument that Guyer finds in Kant’s Deduction of the Categories starts from the claim that there is a special sort of consciousness that we do in fact

achieve – specifically, a consciousness of the necessary identity of the self. Kant’s strategy, Guyer thinks, is to show that the Categories are objectively valid by investigating the conditions that must be satisfied if we are to achieve this consciousness of the necessary identity of the self. So Guyer does not think that Kant introduces the original synthetic unity of apperception into his argument in the context of a claim about the use of our faculty of understanding, or the nature of an act of judgment. He does not think that the concept of the original synthetic unity of apperception plays any role in Kant’s attempt to show that the Categories are objectively valid by way of an analysis of our faculty of understanding.

Van Cleve interprets the role of the original synthetic unity of apperception in much the same way as Guyer. As I mentioned in §2.1, Van Cleve finds in Kant’s Deduction of the Categories two arguments for the objective validity of the Categories. The first argument, which Van Cleve calls ‘the objective deduction’ (1999, p.76), is designed to show that the Categories must be objectively valid if we are to achieve a state of empirical cognition. In particular, it is designed to show that there is a connection between the objective validity of the Categories or Pure Concepts of the Understanding and what Van Cleve calls the ‘conceptual component’ of a state of empirical cognition (1999, p.77). Crucially, there is no reference in this argument to the unity of apperception. Van Cleve thinks that the unity of apperception becomes relevant only when we turn to the argument that he calls ‘the subjective deduction’ (1999, p.79). This argument for the objective validity of the Categories begins from the premise that ‘[all] representations of which I am conscious have unity of apperception’ (1999, p.79), which is how Van Cleve interprets the Principle of the Original Synthetic Unity of Apperception.

Why do Guyer and Van Cleve fail to see that Kant introduces the concept of the original synthetic unity of apperception in the context of a claim about the conditions that must be satisfied if we are to make use of our faculty of understanding? Why do they take Kant's argument for the objective validity of the Categories – or, at least, the argument for the objective validity of the Categories that turns on a claim about the original synthetic unity of apperception – to start with the claim that the original synthetic unity of apperception is something that we do in fact achieve or that is in fact a feature of our representations? I take it that the reason has to do with what Kant says in a passage of the *Critique* that is almost always discussed in this context – namely, §16 of the second edition Deduction of the Categories: On the Original Synthetic Unity of Apperception (B131). This section begins:

The *I think* must be able to accompany all of my representations; for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation either would be impossible, or at least [that it] would be nothing for me. That representation which can be given before all thinking is called intuition. Thus all of the manifold of intuition has a necessary relation to *the I think* in the same subject in which this manifold is encountered. (B131-132)

What Kant has in mind when he refers to the *I think* in the first sentence of this passage is a special representation – and, in particular, a special sort of consciousness of the self, which he calls the original apperception. This is clear from the fourth sentence of §16:

But this representation [i.e. the *I think*] is an act of spontaneity, i.e. it cannot be regarded as belonging to sensibility. I name it the pure apperception in order to distinguish it from the empirical [apperception], or also the original apperception, because it is that self-consciousness which, as it yields the *I think*, which must be able to accompany all others, and which is one and the same in all consciousness, can be accompanied by no further [representation]. (B131)

I propose that Kant's claim in the famous first sentence of §16 should be understood as follows: Any representation that counts as one of mine must be able to be accompanied by the special representation or apperception *I think*. Van Cleve, however, rejects this interpretation. He writes:

The first thing that must be realized about the sentence just quoted [i.e. the first clause of the first sentence of §16] is that Kant means more than "it must be possible for each of my representations to be accompanied by an 'I think'." The 'I think' must attach to my representations conjointly. (1999, p.80)

Suppose that I have exactly three representations x , y and z . On Van Cleve's view, when Kant claims that the *I think* must be able to accompany all of my representations, what he means is not that the *I think* must be able to accompany x , and that it must be able to accompany y , and that it must be able to accompany z . What he means is that it must be possible for x , y and z to be 'compresent to consciousness' (1999, p.81).

What reason is there to accept Van Cleve's interpretation? He draws attention to the following lines from the same section of the second edition *Deduction of the Categories*:

Thus this relation [relation to the identity of the subject] does not yet come about [check] in virtue of the fact that I accompany each representation with consciousness, but [rather in virtue of the fact that] I add one to the other and am conscious of the synthesis of [these representations]. (B133)

On Van Cleve's reading, what Kant claims in the first clause of this sentence is that it is insufficient for the original synthetic unity of apperception that the *I think* accompanies each one of my representations. But this interpretation of the sentence turns on the assumption that when Kant refers to the fact that 'I accompany each representation with consciousness', what he has in mind is the fact that I accompany each representation with the *I think* - that is to say, with the representation that Kant also calls the pure apperception or original apperception (B131). But this is not the case. Consider the previous two sentences of §16:

Namely, this thoroughgoing [check] identity of the apperception of a manifold given in intuition contains a synthesis of representations, and is possible only by means of the consciousness of this synthesis. For the empirical consciousness that accompanies different representations is in itself scattered and without relation to the identity of the subject. (B133)

What Kant claims here is that if we only take into account the empirical consciousness that accompanies different representations, we will find no unity of consciousness, which is to say, no identity of apperception. This makes it reasonable to interpret the first part of Kant's claim in the next sentence of §16 as follows: The identity of apperception does not come about in virtue of the fact that each representation is accompanied by empirical consciousness. And if this is correct then what Kant tells us at this point in the text does not support Van Cleve's interpretation of the first clause of the first sentence of §16. Kant's claim is not that that it is insufficient for the original synthetic unity of apperception that each of my representations is accompanied by the original apperception – that is to say, by the *I think* – but rather that it is insufficient for the original synthetic unity of apperception that each of my representations is accompanied by empirical consciousness. So it is still reasonable to interpret the first clause of the first sentence of §16 as the claim that the special representation *I think* must be able to accompany each one of my representations.

Now, one of the claims that Kant makes about this claim is that it is analytic. He writes in the course of §16 that

this principle of the necessary unity of apperception is, to be sure, itself identical, thus an analytic proposition. (B135)

And he adds in §17 that 'this last proposition', by which he means 'the principle of the original synthetic unity of apperception',

is, as we said, analytic... for it says nothing more than that all my

representations in any given intuition must stand under the condition under which alone I can ascribe them to the identical self as my representations... (B138)

Now Kant tells us that what he means by an analytic proposition is a proposition where 'the predicate B belongs to the subject A as something that is (covertly) contained in this concept A' (A6/B10), or, in other words, a proposition where 'the connection of the predicate is thought through identity' (A7/B10). In the introduction to this thesis I explained this concept of analyticity with reference to the distinction between a case in which someone makes a new discovery, that is to say, comes to know something that she did not know before, and a case in which she becomes conscious of something that she already knew. In establishing an analytic proposition – say, the proposition that all bodies are extended – I become conscious of something that was already contained, if covertly, in one of my concepts.

So what does Kant mean when he claims that the principle he has introduced at the start of §16 is analytic or identical? The answer is as follows: Part of what it is to think of a representation that counts as one of mine is to think of a representation that can be accompanied by the special representation *I think*. The concept of that which can be accompanied by the *I think* is contained – covertly, perhaps – in the concept of a representation that counts as one of mine. So Kant writes, for example, that

the manifold representations that are given in a certain intuition would not all together be my representations if they did not all together belong to a self-consciousness, i.e. as my representations... they must yet necessarily be

in accord with the condition under which alone they can stand together in a universal self-consciousness, because otherwise they would not throughout belong to me. (B132-133)

Kant's emphasis on the pronoun - and also the final clause of the sentence - makes it clear that on his view, unless a representation can be accompanied by the *I think*, it does not count as one of mine. Elsewhere in §16 Kant writes:

The thought that these representations given in intuition all together belong to me means, accordingly, the same as that I unite them in a self-consciousness, or at least can unite them therein... (B134)

With these points in mind, we can start to see why commentators such as Guyer and Van Cleve have failed to see that the concept of the original synthetic unity of apperception is introduced into Kant's argument in the context of a claim about the use of our faculty of understanding. The Principle of the Original Synthetic Unity of Apperception is the claim that 'all my representations in any given intuition must stand under the condition under which alone I can ascribe them to the identical self as my representations...' (B138). It is the claim that if a representation is one of mine then I must be able to accompany it with the original apperception or *I think*. And this claim is supposed to be analytic. It is part of the concept of a representation that counts as one of mine that I am able to accompany it with the original apperception or *I think*. However, there is no obvious mention of the faculty of understanding in this claim. Kant's Principle of the Original Synthetic Unity of Apperception doesn't seem on the

face of it to be a claim about the conditions that must be satisfied in order for someone to make use of his faculty of understanding, or make a judgment about an object. It seems instead to be a claim about what must be true of a representation in order for it to count as one of mine.

Yet there is a way to avoid this conclusion. The crucial question is this: What is it that makes Kant think that the Principle of the Original Synthetic Unity of Apperception is true? We can answer this question if we look again at the second half of the first sentence of §16. Here is the sentence in full:

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

What Kant does in the second half of this sentence is to describe the consequences if there is a representation in me that cannot be accompanied by the *I think*. The first claim he makes is that if there is such a representation then there is a representation in me – presumably, the very same representation – that cannot be thought at all. The second claim is a little harder to understand, since it seems as if Kant is unsure what exactly he should say. He seems unsure whether to say that if a representation of mine cannot be thought at all then that is as much as to say that it is impossible, or whether to say that if a representation of mine cannot be thought at all then that is as much as to say that it is not a representation of mine after all. I propose that what Kant has done here is simply to confuse two equivalent claims. The first is that if a

representation of mine cannot be thought at all then that is as much as to say that it is impossible. And the second is that if a representation cannot be thought at all then that is as much as to say that is not one of my representations. If this is correct then Kant's line of reasoning in the first sentence of §16 can be reconstructed as follows: If a representation cannot be thought at all then it does not count as one of my representations. But if a representation cannot be accompanied by the *I think* or original apperception then it cannot be thought at all. So if a representation counts as one of my representations then it is one that can be accompanied by the *I think*.

Now, what exactly does Kant mean by a representation that cannot be thought at all? In the very next sentence of §16, he writes:

That representation which can be given before all thinking is called intuition. (B132)

What Kant means by 'thinking' in this sentence is the activity of the faculty of understanding, as contrasted with 'intuition', which is the activity of the capacity for sensibility (e.g. A51/B75). So it is reasonable to assume that what Kant means by the word 'thought' in the first sentence of §16 is the activity of the faculty of understanding. In that case, Kant's claim that 'otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation either would be impossible', should be understood as follows: If a representation of mine cannot be accompanied by the *I think* then it is not one in relation to which I can make use of my faculty of understanding. For part of what it is to make use of my faculty of understanding in relation to a representation is to

accompany this representation with the I think or original apperception. But if a representation is not one that I can accompany with the I think or original apperception then it does not count as one of my representations at all. For part of what it is for a representation to count as one of mine is for it to be one in relation to which I can make use of my faculty of understanding. So a representation of mine that cannot be accompanied by the I think or original apperception is impossible.

If this interpretation of the first sentence of §16 is correct then Kant's claim that the *I think* or original apperception must be able to accompany all of my representations is in fact a consequence of two claims. The first is a claim about what it is for a representation to count as one of mine and the second is a claim about what must be true of a representation if it is to be a representation in relation to which I can make use of my faculty of understanding. Crucially, the second claim is that I can make use of my faculty of understanding in relation to a representation only if I can accompany it with the *I think* or original apperception. It is the claim that Kant expresses in the Transcendental Dialectic as follows:

the proposition *I think...* contains the form of every judgment of understanding whatsoever... (A348/B406)

And so it appears that what Kant tells us at the start of §16 is not so different from what he tells us in §17 and §19. As in §17 and §19, the connection between original apperception and the use of our faculty of understanding is absolutely crucial. In §19, the claim is that the relation of representations to original apperception is essential to the use of our faculty of understanding, which is to say, it is essential to the act of

judgment. In §17, the claim that Kant identifies as the Principle of the Original Synthetic Unity of Apperception is that if an intuition is to be one in relation to which we make use of our faculty of understanding, it must satisfy the conditions of the original synthetic unity of apperception. And in §16, the claim is that since a representation is one in relation to which we can make use of our faculty of understanding only if it is one that we can accompany with the original apperception, and since a representation is one of mine only if it is one in relation to which I can make use of my faculty of understanding, a representation is one of mine only if it is one that I can accompany with the original apperception. So Guyer and Van Cleve are mistaken when they assume that the concept of original apperception or the original synthetic unity of apperception is not introduced in the context of a claim about what is required by the use of our faculty of understanding.

§3.2 Kant's concept of original apperception

So far in this chapter, I have reached the conclusion that Kant's Principle of the Original Synthetic Unity of Apperception is a claim about the conditions that must be satisfied if we are to make a judgment about an object – or, in other words, a claim about the conditions that must be satisfied if we are to make use of our faculty of understanding. My aim in the rest of this chapter is to explain what exactly this claim is supposed to be. The first step that I need to take is to clarify Kant's concept of apperception in general, and his concept of original apperception in particular.

Kant's earliest reference to apperception in the first edition of the *Critique* comes in the Transition to the Transcendental Deduction of the Categories:

There are... three original sources... of the possibility of all experience, [which] cannot themselves be derived from any other faculty of the mind, namely, sense, imagination and apperception. (A94)

Kant claims that

[on] these [faculties] are grounded... the synopsis of the manifold *a priori* through sense, the synthesis of this manifold through the imagination, [and] finally, the unity of this synthesis through original apperception. (A94)

He then remarks that

[in] addition to their empirical use, all of these faculties have a transcendental [use], which is concerned solely with form and which is possible *a priori*. (A94)

Kant clearly thinks that apperception is a faculty of the mind that has both an empirical use and a transcendental or *a priori* use. And he clearly thinks that original apperception – I take this to mean the transcendental or *a priori* use of the faculty of apperception – is responsible for the unity of synthesis.

Kant's next reference to apperception in the first edition of the *Critique* comes right in the middle of the Deduction of the Categories itself. Here Kant refers to a

certain necessity, and claims that this necessity requires a transcendental condition. He then writes that

this original and transcendental condition is nothing other than transcendental apperception. (A106-107)

He continues:

The consciousness of oneself in accordance with the determinations of our state in inner perception is merely empirical, forever variable; it can provide no standing or abiding self in this stream of inner appearances, and is customarily called inner sense or empirical apperception. That which should necessarily be represented as numerically identical cannot be thought of as such through empirical data. (A107)

I take it that the necessity that Kant has in mind in this passage is the unity of synthesis to which he referred in the Transition to the Transcendental Deduction of the Categories. For Kant often describes the unity of synthesis that he thinks is essential to experience as a necessary unity of synthesis (e.g. A108). And I take it that what Kant means by transcendental apperception is the transcendental or *a priori* use of the faculty of apperception. If this is correct then Kant has here repeated his claim that the *a priori* use of the faculty of apperception – that is to say, the use of this faculty in original apperception – is responsible for the unity of synthesis. He has also made it clear that the faculty of apperception can be used both empirically and transcendently. And he

has indicated that the empirical use of the faculty of apperception is what is usually called 'inner sense' (A107). Finally, he has made a third claim about apperception – namely, that whether our faculty of apperception is used empirically or transcendently, it constitutes an ability to achieve some sort of consciousness of the self. A key difference between empirical and transcendental apperception is that it is only by means of the transcendental use of our faculty of apperception that we are able to become consciousness of the self as something that is numerically identical.

One criterion that should be satisfied by an interpretation of Kant's concept of apperception is that it should help us to understand the claims about apperception that I have extracted from the first edition of the *Critique*. It should help us to understand the distinction between the empirical and the transcendental or *a priori* use of this faculty, and it should help us to understand how apperception counts as a sort of consciousness of the self. Another criterion that should be satisfied by an interpretation of Kant's concept of apperception is that it should allow us to see this concept as closely related to the concept of apperception that we find in the work of other 18th Century German philosophers – in particular, Leibniz and Wolff. The reason for this is as follows: Kant fails to provide any clear explanation of what he means by the word 'apperception'. While he is careful to define 'synthesis' (A77/B103), he never defines 'apperception'. The most charitable explanation of this failure is that Kant has set out to use the word 'apperception' in a sense that would have been familiar to his intended audience – that is to say, readers of 18th Century German philosophy. And the likeliest candidate is the sense of the word 'apperception' in the work of Leibniz and Wolff.

‘Apperception’ had been introduced as a philosophical term of art with the publication in 1720 of a German translation of Leibniz’s *Monadology*. This contains the lines:

The changing state, which grasps in itself and represents a manifold in [the monad], is nothing other than that which one calls *sensation* or *perception*, which one must distinguish from *apperception* or *consciousness*... (Lamarra et al., 2001, p.50)

Twelve years later, Wolff used the word ‘apperception’ in his influential *Empirical Psychology*. Wolff defines apperception as that which ‘is attributed to the mind insofar as it is conscious of its perceptions’ (1732, p.17), where perception is ‘the act of the mind [whereby it] represents some object or other to itself’ (1732, p.17). This definition of apperception is similar to the one that we find in the edition of Leibniz’s *Principles of Nature and Grace* that was published in 1740:

...it is well to make a distinction between the *perception* which is the inner state of the monad representing outer things, and the *apperception* which is the consciousness or the reflexive cognition of this inner state...

There were undoubtedly variations in the use of the word ‘apperception’ in the middle of the 18th Century (cf. Thiel 2011, Kitcher 2011). However, there is no other sense of the word ‘apperception’ that would have been as familiar to Kant’s readers as the sense in which it was used by Leibniz and Wolff. There is also reason to think that this was

the sense of the word 'apperception' at the forefront of Kant's mind in the years when he was working on the argument of his *Deduction of the Categories*. We know that Kant read Leibniz's *New Essays Concerning Human Understanding* after they were finally published in 1765. And this is the work of Leibniz's in which the word 'apperception' occurs most frequently.

One commentator, Kitcher (2011), has recently argued that Leibniz's concept of apperception is very different to Kant's. Kitcher thinks that what Leibniz means by the word 'apperception' is a special sort of cognition of the self – specifically, the cognition of the self as an object that exists and that is a substance and that is simple. She thinks that apperception in Leibniz's sense is 'no mere perception, but a highly sophisticated cognition achievement' (2011, p.57). She claims that 'in apperceiving [in Leibniz's sense], thinkers are aware of themselves as simple substances' (2011, p.57). If Kitcher is right that this is what Leibniz means by the word 'apperception' then she is right to conclude that Leibniz's concept of apperception is very different to Kant's. For Kant clearly denies that transcendental apperception – that is to say, original apperception – is any sort of cognition. He insists in the *Paralogisms of Pure Reason* that the representation of the I – by which he means original apperception (cf. B132) – 'is no more an intuition than it is a concept' (A382). He adds that this representation is 'the mere form of consciousness, which accompanies both sorts of representations [i.e. both intuitions and concepts]' (A382). Now, Kant thinks that a representation counts as a cognition only if it is either an intuition or a concept (A320/B376/B377). So he cannot think that transcendental apperception is any sort of cognition. But if Kant doesn't think that transcendental apperception is any sort of cognition then he cannot think that apperception in general is a sort of cognition of the self.

The trouble with Kitcher's argument is that it is based on a flawed interpretation of Leibniz's concept of apperception. The evidence that Kitcher uses to justify her interpretation is a passage in the *Monadology* where Leibniz writes:

it is also through the knowledge of necessary truths, and through their abstract expression, that we rise to acts of reflexion, which make us think of what is called I, and observe that this or that is within us: and thus, thinking of ourselves, we think of being, of substance, of the simple and the compound, of the immaterial, and of God himself...

Kitcher takes this to mean the following: It is possible for us to reflect upon ourselves – that is to say, it is possible for us to achieve apperception – only if we have first achieved a certain sort of cognition – specifically the sort of cognition that consists in understanding what it is for something to be a substance, what it is for something to be immaterial, to be simple etc. And this is because it is only if we understand what it is for something to be a substance, what it is for something to be immaterial, to be simple etc. that we can achieve cognition of ourselves as substances that are simple and immaterial. Kitcher concludes on this basis that what Leibniz means by achieving apperception is just achieving cognition of ourselves as substances that exist and that are simple and immaterial etc. As she puts it,

Prior to understanding what simple substances are, people lack the ability to reflect on themselves, because they are not equipped to think about what are, in fact, themselves. (2011, p.57)

But there is reason to be sceptical about Kitcher's reading of the key passage. Take the claim that 'thinking of ourselves, we think of being, of substance, of the simple...' etc. Leibniz's point here seems to be that it is only when we think of ourselves – that is to say, it is only when we think of the object that we call 'I' – that we are in a position to think of existence, substance, simplicity, etc. Compare here the claim in Leibniz's *New Essays* that

intellectual ideas, or ideas of reflection, are drawn from our mind. I would like to know how we could have the idea of being if we did not, as beings ourselves, find being within us.

Leibniz thinks that we perform certain acts of reflection and that this allows us to form the concepts of existence, simplicity, substance etc. Far from it being the case that apperception consists in the cognition of the self as something that exists and that is a simple substance etc., the claim seems rather to be that the thoughts of existence, simplicity etc. – and therefore the cognition of the self as something that exists and that is a simple substance etc. – only become possible once we have already achieved apperception.

Both Leibniz and Wolff define apperception as the consciousness of a state of the mind that serves as a representation of an object. And I suggest that this is exactly what Kant means by the word 'apperception'. Or, to be a little more precise, I suggest that what Kant means by the word 'apperception' is the consciousness of a state of mind as one that serves as a representation of an object. One reason to think that this

interpretation is correct is that it allows us to make good sense of Kant's distinction between empirical apperception and transcendental apperception. Remember that Kant thinks that there are two ways in which a state of mind can serve as a representation of an object – that is to say, two ways in which a state of mind can stand in a relation of necessary agreement or conformity with an object. On the one hand, Kant thinks that a state of mind serves as a representation of an object if it conforms to an object – that is to say, if it has the form that it does precisely because the object itself has this form. If a state of mind serves as a representation of an object in this way, it counts as an empirical representation. On the other hand, Kant thinks that a state of mind serves as a representation of an object if the object conforms to the state of mind – that is to say, if the object has the form that it does precisely because the state of mind has this form. And if a state of mind serves as a representation of an object in this way, it is an *a priori* representation. Now, if Kant thinks that apperception consists in the consciousness of a state of mind as one that serves as a representation of an object, it is natural to make a distinction between the consciousness of a state of mind as one that serves as a representation of an object in virtue of the fact that it conforms to this object and the consciousness of a state of mind as one that serves as a representation of an object in virtue of the fact that the object conforms to the state of mind. The first sort of apperception would be the consciousness of a state of mind as an empirical representation of an object, and the second sort of apperception would be the consciousness of a state of mind as an *a priori* representation of an object. And so it would be natural to describe the first sort of apperception as empirical apperception and the second sort of apperception as a non-empirical or transcendental apperception.

There is textual evidence to corroborate this interpretation of the distinction between empirical apperception and transcendental apperception. In a famous footnote to his *Anthropology from a Pragmatic Point of View*, Kant writes:

If we consciously represent two acts: inner activity (spontaneity), by means of which a concept (a thought) becomes possible, or reflection; and receptiveness (receptivity), by means of which a perception (perceptio) becomes possible, or apprehension; then consciousness of oneself can be divided into that of reflection and that of apprehension. The first is a consciousness of understanding, pure apperception; the second a consciousness of inner sense, empirical apperception. In this case, the former is falsely named inner sense. (Ak VII:134)

Much of what Kant says in this footnote is easily misunderstood. Take the last sentence but one. It looks at first glance as if Kant's claim is that empirical apperception consists in the consciousness of the use of our inner sense whereas pure apperception consists in the consciousness of the use of our faculty of understanding. But I don't think this can be what Kant means to say. One reason for this is that there is evidence that Kant identifies our faculty of empirical apperception with our capacity for inner sense. Even the very last sentence of the passage from the footnote supports this conclusion: Kant's claim here is only that transcendental apperception is falsely named inner sense. But if inner sense is the same as empirical apperception then surely empirical apperception cannot consist in the consciousness of the use of our inner sense. Second, if empirical apperception consists in the consciousness of the use of our inner sense whereas pure

apperception consists in the consciousness of the use of our faculty of understanding, we can ask: What about outer sense? Why doesn't Kant mention a third sort of apperception that consists in the consciousness of the use of our outer sense? Why doesn't he define empirical apperception as the consciousness of the use of our capacity for sensibility, whether this is inner sense or outer sense? Third, if what I claimed in §3.1 is correct, this claim misrepresents the connection between the faculty of understanding and the faculty of transcendental apperception. Kant doesn't think that transcendental apperception consists in the consciousness of the use of our faculty of understanding. On the contrary, he thinks that transcendental apperception is an essential ingredient in the use of our faculty of understanding. Part of what it is to make use of our faculty of understanding is to achieve transcendental apperception.

In light of these considerations, I conclude that we should read Kant's claim in the penultimate sentence of the footnote in a very different way. When Kant describes pure apperception as the consciousness of understanding, what he means is not that it is the consciousness the object of which is the use of our faculty of understanding, but rather that it is the consciousness that is the essential ingredient in an exercise of our faculty of understanding. It counts as the consciousness of understanding in the sense that it is the consciousness that is a function of our faculty of understanding. If this is right then we should take Kant's claim that empirical apperception is the consciousness of inner sense as the claim that empirical apperception is the consciousness that stands to inner sense as pure apperception stands to the understanding, which is to say, as the essential ingredient in inner sense.

With this point in mind, let's look at the first claim in the passage from the footnote. Kant's claim here is that one sort of consciousness of the self – namely, pure

apperception – consists in the consciousness of the spontaneous activity of the mind, whereas the other sort of consciousness of the self – namely, empirical apperception – consists in the consciousness of the merely receptive activity of the mind. Now, consciousness of the receptive activity of the mind is naturally understood as consciousness of the mind as having the form that it does precisely because an object has this form – that is to say, consciousness of the mind as conforming to an object or being determined by an object. But that is exactly how empirical apperception is understood on the interpretation that I have introduced in this section. And the claim that transcendental apperception consists in the consciousness of a state of mind as one that determines an object – as a state of mind to which an object conforms – provides us with a plausible way to understand the claim that transcendental apperception consists in the consciousness of the spontaneous activity of the mind.

It is also easy to see how apperception counts as a sort of consciousness of the self on this interpretation. It is a consciousness of the state of our mind – that is to say, it is a consciousness the state of the self. This interpretation is also at least consistent with Kant’s claim about the sort of consciousness of the self that results from transcendental apperception as opposed to empirical apperception. Remember that according to Kant only transcendental apperception amounts to a consciousness of the self as something that is necessarily identical. The self of which we are conscious in empirical apperception is always changing. On the interpretation that I have defended so far, we can read this as follows: Those aspects of our states of mind that count as representations of objects in virtue of the fact that they conform to objects – that is to say, those aspects of our states of mind that represent objects empirically – are such that there is no part of them that is always the same. The way we are determined by objects

is always changing. So if we are to become conscious of an aspect of our states of mind that represents an object and that is common to all such states of mind, it must represent an object in a non-empirical way, that is to say, *a priori* or transcendently. It will become clear in §3.4 that this is precisely what Kant thinks is true of transcendental or original apperception.

The interpretation of Kant's concept of apperception that I have defended in this section is broadly similar to the one recently proposed by Dickerson (2004). Dickerson writes that 'apperception is the reflexive act whereby the mind grasps its own representations as representing, and is thus an essential part of all thought and cognition' (2004, p.81). The crucial point is that he thinks that apperception consists in an act of 'reflexively grasping' those of our states of mind that count as representations of objects as states of our mind that count as representations of objects. What is apperceived is precisely the relation of the states of our mind to objects. Dickerson also expresses the idea as follows: It is one thing for us to be in a state of mind that counts as a representation of an object, and another thing for a state of mind to represent an object to us - that is to say, for this state of mind to be part of our consciousness of objects. And what is added in the second case is apperception:

it is apperception that is the difference between simply being in a state that represents, 'expresses' or 'mirrors' external things - a state that, for example, a computer, a camera, or a thermometer can also be in - and genuine conscious experience - in which the state of the subject does not simply represent something, but represents something to the subject.
(2004, p.85)

There are, however, some differences between Dickerson's interpretation of Kant's concept of apperception and the one that I have defended in this section. The most important difference is that Dickerson fails to draw the conclusions that I have defended about the distinction between empirical apperception and transcendental apperception. Indeed, Dickerson entirely fails to consider the question how it is that there can be two sorts of apperception: one empirical and one transcendental. He interprets Kant's claims about 'the *I think*' in the Deduction of the Categories not as claims about a particular sort of apperception – namely, pure apperception or original apperception – but rather as claims about apperception in general. The reason why Dickerson fails to consider the way in which his interpretation can be used to understand the difference between empirical apperception and transcendental apperception is that while he does think that apperception consists in the 'reflexive grasping' of a state of mind as a representation of an object, he doesn't think that Kant understands the relation of representation between a state of mind and an object in the way that I proposed in §1.1. He doesn't appreciate Kant's distinction between the two ways in which a state of mind can serve as a representation of an object, and so he fails to see how this distinction can serve as the basis for the distinction between empirical apperception and transcendental apperception.

Another difference between Dickerson's interpretation and the one that I have defended in this section is that Dickerson does not define apperception as the 'consciousness' of a state of mind as one that counts as a representation of an object. He uses the vague language of 'reflexively grasping' instead. He defines apperception in this way because he thinks that if we achieve apperception in relation to a state of

mind, the object of our consciousness is not the state of mind, but rather the object that is represented to us by this state of mind. In apperceiving, or reflexively grasping the state of mind as one that represents an object, we become conscious of this object. Dickerson thinks that this is a crucial point. He insists that Kant's concept of apperception 'should not be assimilated to modern notions of self-awareness, self-consciousness, self-knowledge, or self-reference' (2004, p.86).

I suggest that the point that Dickerson has in mind here would be better expressed as follows: Apperception is the consciousness of a state of mind as one that serves as a representation of an object. The state of mind is itself an object of consciousness. But it is not the object of cognition. What is the object of cognition is the object that is represented by the state of mind. So Dickerson is right to say that apperception should not be understood as a sort of cognition of the self. But he goes too far when he says that apperception should not be understood as a consciousness of the self. According to Kant, apperception consists in a consciousness of the self that is an essential ingredient in all cognition – that is to say, a consciousness of the self that is as much an ingredient in the cognition of books, trees and cups of tea as it is in the cognition of our selves.

In this section I have defended an interpretation of Kant's concept of apperception in general, and an interpretation of his concept of original apperception in particular. I have claimed that what Kant understands by apperception in general is the consciousness of a state of mind as one that serves as a representation of an object. And I have claimed that Kant's distinction between empirical apperception and transcendental or original apperception is the distinction between the consciousness of a state of mind as one that represents an object empirically and the consciousness of a

state of mind as one that represents an object *a priori*. Original apperception is the consciousness of a state of mind as one that serves as a representation of an object in virtue of the fact that the object conforms to the state of mind. This interpretation allows us to see Kant's concept of apperception as continuous with the concept of apperception that we find in the work of Leibniz and Wolff. And it allows us to see Kant's concept of apperception as the concept of a consciousness of the self, but not a sort of cognition of the self. At the start of this section I mentioned another claim that Kant makes about original apperception in particular, viz. that the exercise of this faculty of the mind results in a unity of synthesis. It should count in favour of an interpretation of Kant's concept of apperception that it allows us to make sense of this claim. One of my aims in §3.4 will be to show that this is true of the interpretation that I have defended in this section. First, however, I must explain what Kant means by an act of synthesis.

§3.3 Kant's concept of an act of synthesis

Here is Kant's definition of the word 'synthesis' in the first chapter of the *Analytic of Concepts*:

By *synthesis* in the most general sense I understand the action of putting different representations together with each other and grasping their manifoldness in one cognition. (A77/B103)

The reason why Kant writes that this is what he means by synthesis ‘in the most general sense’ becomes clear if we look at the previous two sentences:

Now, space and time contain a manifold of pure *a priori* intuition... Only the spontaneity of our thought requires that this manifold [viz. the ‘manifold of pure *a priori* intuition’] is first in a certain way gone through, taken up, and combined, in order for a cognition to be made out of it. I call this action synthesis. (A77/B102)

Here Kant uses the word ‘synthesis’ to refer to the action by means of which a particular sort of manifold, viz. a manifold of ‘pure *a priori* intuition’, is ‘gone through, taken up, and combined’. But he doesn’t mean to define synthesis in the most general sense as an action that is performed with respect to a manifold of this sort. He allows for the possibility of synthesis that is performed with respect to a manifold that is not pure *a priori*, but rather empirical. Consider, for example, the sentence that comes immediately after his definition of synthesis:

Such a synthesis is *pure* if the manifold is given not empirically but *a priori* (as is that in space and time). (A77/B103)

Pure synthesis is the action by means of which *a priori* representations are ‘gone through, taken up, and combined’. Empirical synthesis is the action by means of which empirical representations are ‘gone through, taken up, and combined’. In general, synthesis is the action by means of which different representations are ‘gone through,

taken up, and combined'. We should bear in mind, however, that if this is what Kant means by the word 'synthesis' in 'the most general sense', there may be a narrower sense in which he uses the word 'synthesis', in which it only refers to an action that is performed with respect to a 'manifold of pure *a priori* intuition', and which is required by 'the spontaneity of our thought'. It may be that Kant occasionally restricts the meaning of 'synthesis' to what he also calls 'pure synthesis', just as he occasionally restricts the meaning of 'apperception' to what he also calls 'pure apperception', or 'original apperception' (cf. §3.2).

Clearly, Kant thinks that synthesis is an act of the mind that is performed with respect to a manifold of representation – that is to say, with respect to a number of representations. If the mind has not acquired several representations, it cannot perform an act of synthesis. In the second edition Deduction of the Categories, Kant mentions some of the different sorts of manifold of representation with respect to which an act of synthesis can be performed:

the combination (*conjunctio*) of a manifold in general... is an act of the spontaneity of the power of representation, and, since one must call the latter understanding in distinction from sensibility, all combination, whether we are conscious of it or not, whether it is a combination of the manifold of intuition or of several concepts, and in the first case either of sensible or non-sensible intuition, is an action of the understanding, which we would designate with the general title synthesis. (B129-130)

Here it seems that an act of synthesis can be performed with respect to either several intuitions or several concepts. Now, intuitions and concepts are states of the mind that are related to objects in the sense that they serve as representations of objects. So it seems that Kant conceives of an act of synthesis as an act of the mind that is performed with respect to states of mind that are related to objects.

It may be objected at this point that Kant also thinks that there is a sort of synthesis that it is necessary for us to perform if we are to have an intuition in the first place. What I have in mind here is the 'synthesis of apprehension in the intuition', which is the first element of the 'threefold synthesis' that Kant claims is 'necessarily found in all cognition' (A97). Here is how Kant introduces the synthesis of apprehension in the second section of the first edition Deduction of the Categories:

Every intuition contains a manifold in itself, which however would not be represented as such if the mind did not distinguish the time in the succession of impressions on one another... Now in order for unity of intuition to come from this manifold... it is necessary first to run through and then take together this manifoldness, which action I call the synthesis of apprehension, since it is aimed directly at the intuition, which to be sure provides a manifold, but can never effect this as such, and indeed as contained in one representation, without the occurrence of such a synthesis. (A99)

Here is one fairly natural interpretation of these lines: Kant thinks that an intuition is a representation that contains several representations – that is to say, a representation

that has a number of other representations as its parts. He thinks that in order for us to acquire a representation of this sort it is necessary for us to acquire several representations – specifically, several impressions – and then to combine them in an act of synthesis. What is produced by this act of synthesis – that is to say, the synthesis of apprehension – is one representation – specifically, one intuition. The synthesis of apprehension can be said to be ‘aimed directly at the intuition’ in the sense that an intuition is the immediate result of this synthesis. The representations that are combined in the synthesis of apprehension are not intuitions. Rather, they are impressions.

Further support for this last point can be extracted from Kant’s discussion of the synthesis of apprehension towards the end of the first edition Deduction of the Categories:

since... different perceptions by themselves are encountered dispersed and separate in the mind, a combination of them, which they cannot have in sense itself, its therefore necessary. There is thus an active faculty of the synthesis of this manifold in us, which we call imagination, and whose action exercised immediately upon perceptions I call apprehension. For the imagination is to bring the manifold of intuition into an image; it must therefore... take up the impressions into its activity, i.e. apprehend them.
(A120)

Here once again Kant indicates that the representations that are combined in the synthesis of apprehension are impressions – or, as he also writes, perceptions.

We should, however, resist the conclusion that an intuition is the result of the synthesis of apprehension. Notice first that Kant claims that in the synthesis of apprehension the ‘manifold of intuition’ is made into an ‘image’ (A120). But that is just to say that several intuitions are combined into one representation. Second, when Kant writes that the synthesis of apprehension is ‘aimed directly at the intuition’, he need not mean that an intuition is the immediate result of the synthesis. He could also mean the same as when he writes in the later passage that the synthesis of apprehension is the act of the imagination ‘exercised immediately upon perceptions’. He could mean that it is a manifold of intuition with respect to which the synthesis of apprehension is performed. The synthesis is aimed at the intuition in the sense that it is intuition to which it is applied. Third, there is evidence that when Kant refers to the ‘perceptions’ that are combined in the synthesis of apprehension, what he has in mind are intuitions – or, rather, intuitions of which the mind is conscious. In the *Prolegomena* Kant writes:

At [the] bottom [of experience] lies the intuition of which I am conscious,
i.e. perception (perceptio), which belongs solely to the senses. (Ak IV:300).

Further, in the lines just before his reference to the connection of different perceptions in the first edition Deduction of the Categories Kant writes:

The first thing that is given to us is appearance, which, if it is combined
with consciousness, is called perception... (A119-120)

But what Kant means by an appearance is just an object of empirical intuition (cf. A20/B34). So if someone has appearance combined with consciousness then she has empirical intuition combined with consciousness. I can conclude that the synthesis of apprehension is not a sort of synthesis that is required in order for us to have an intuition. Rather, it is a sort of synthesis that is performed with respect to a particular sort of manifold, viz. a collection of empirical intuitions. And so there is no need for us to reject the claim that Kant thinks of synthesis as an act of the mind that is performed with respect to several intuitions or several concepts – that is to say, with respect to several states of our mind that are related to objects either directly or indirectly.

The next point to make about an act of synthesis is that it involves the representation of a connection between the objects of several representations. Consider first Kant's discussion of the synthesis of apprehension – that is to say, the synthesis of several empirical intuitions – towards the end of the second edition Deduction of the Categories. Here he writes:

by the synthesis of apprehension I understand the composition of the manifold in an empirical intuition, through which perception... becomes possible. (B160)

He then makes a claim about the connection between the synthesis of apprehension and another sort of synthesis. And he illustrates this claim with two examples:

If, for example, I make the empirical intuition of a house into perception through apprehension of its manifold, my ground is the necessary unity of space and of outer sensible intuition in general... (B162)

If (in another example), I perceive the freezing of water, I apprehend two states (of fluidity and solidity) as ones standing in a relation of time to each other.... (B162)

Consider the second example. The act of synthesis that Kant has in mind here is performed with respect to two empirical intuitions. The first is the empirical intuition of a patch of water as liquid and the second is the empirical intuition of the same patch of water as solid. One of the results of the act of synthesis seems to be a representation of a relation between the liquidity of the water and the solidity of the water - in particular, a relation of priority in time.

As for the first example, it's not immediately obvious in this case what the empirical intuitions that are combined in the act of synthesis are supposed to be. But the example of the house calls to mind some famous lines from the *Analytic of Principles*:

In the... example of a house my perceptions could have begun at its rooftop and ended at the ground, but also have begun below and ended above; likewise I could have apprehended the manifold of empirical intuition either from the right or from the left. (A192/B237-238)

I take it from this that the empirical intuitions in Kant's first example in the Deduction of the Categories are the intuitions of the parts of the house – for example, the intuition of the roof, the intuition of the door, the intuition of the floor, etc. One of the results of the synthesis of apprehension in this example is a representation of the whole house – that is to say, a representation of the roof, the door, the floor etc. as connected or related in some way.

Let's focus on this representation of a connection between the objects of several representations. Given his assumptions about what is required for a state of mind to serve as a representation of an object, Kant must think that there are two ways in which a state of mind is able to serve as a representation of a connection between the objects of several other representations. He thinks that a state of mind can represent such a connection either empirically or *a priori*. The state of mind represents the connection empirically if it is a consequence of the fact that the objects of the representations are connected in a particular way that the state of mind represents these objects as connected in this way. And the state of mind represents the connection *a priori* if it is a consequence of the fact that the state of mind represents the objects of the representations as connected in a particular way that the objects are in fact connected in this way. Now, I propose that Kant doesn't think that a state of our mind can represent a connection between the objects of several representations empirically. He doesn't think that a representation of the connection between the objects of several representations is the result of the connection that already exists between the objects of these representations. And so he thinks that what is involved in an act of synthesis must be a state of mind that represents a connection between the objects of several representations *a priori*. He must think that what results from an act of synthesis is not

simply a representation of the objects of several representations as connected in some way, but rather an actual connection between the objects of several representations.

To make this as clear as possible: If the interpretation of Kant's concept of synthesis that I have just introduced is correct then Kant thinks that an act of synthesis takes place in the following way. First, the mind acquires several intuitions or concepts of objects. So far, the objects of these several intuitions or concepts are not connected to one another. They may bear relations of similarity to one another, and they may stand in temporal relations to one another, but they are not yet connected or combined. Second, the mind forms for itself a representation of a connection between the objects of these several intuitions or concepts. Third, the objects of these several intuitions or concepts conform to this representation. That is to say, these representations become connected in a particular way as a result of the fact that the mind has formed a representation of this connection. If this interpretation is correct then if we become conscious of the representation of a connection between several of our representations then this consciousness must be a case of original apperception. For it consists in the consciousness of a state of mind as one that serves as a representation of an object – or, to be precise, that serves as a representation of a connection between several objects – in virtue of the fact that the object conforms to the state of mind.

What evidence is there to support this interpretation of Kant's concept of an act of synthesis? There is plenty of evidence that Kant doesn't think that our representations are connected to one another prior to the act of synthesis. He writes that

since... different perceptions are encountered dispersed and separate in the mind, a combination of them, which they cannot have in sense itself, is therefore necessary. There is therefore an active faculty of the synthesis of this manifold in us... (A120)

I take this to mean that if we consider the empirical intuitions in our mind prior to any act of synthesis, we find that these intuitions are entirely unconnected. It is only by means of the act of synthesis that they become connected to one another. Kant also writes that

[if] every individual representation were entirely foreign to the other, as it were isolated and separated from it, then there would never arise anything like cognition, which is a whole of compared and connected representations. If, therefore, I ascribe a synopsis to sense... a synthesis must always correspond to this... (A97)

Kant's claim here is that our representations must be connected to one another, and that they can become connected to one another only if we have a faculty of synthesis. But this implies that our representations are connected to one another only in virtue of the fact that we perform acts of synthesis.

This interpretation of Kant's concept of synthesis also allows us to make sense of Kant's claims about the active or spontaneous character of synthesis. He writes in the first edition Deduction of the Categories:

If... I ascribe a synopsis to sense... a synthesis must always correspond to this, and receptivity can make cognitions possible only if combined with spontaneity. (A97)

And in the second edition Deduction of the Categories he writes:

the combination... of a manifold in general can never come to us through the senses... for it is an act of the spontaneity of the power of representation, and, since one must call the latter understanding... all combination... is an action of the understanding, which we would designate with the general title synthesis in order at the same time to draw attention to the fact that we can represent nothing as combined in the object without having previously combined it ourselves and that among all representations combination is the only one that is not given through objects but can be executed only by the subject itself, since it is an act of its self-activity. (B129-130)

In §3.2, I proposed that what Kant has in mind when he contrasts the spontaneity of our power of representation to its receptivity is the idea that while some representations in us count as representations of objects in virtue of the fact they conform to objects, other representations in us count as representations in virtue of the fact that the representations conform to them. If this is correct then we can see how synthesis counts as an act of spontaneity. For an act of synthesis consists in the act by means of

which the mind forms a representation of a connection between the objects of several representations to which the objects themselves subsequently conform.

There is also evidence to support this interpretation of Kant's concept of an act of synthesis in a letter that Kant wrote to Jacob Sigismund Beck on 20th January 1792.

Here he writes:

one may still ask: How can union of representations, being complex, be represented? Not through the awareness that it is given to us; for a union requires uniting, (synthesis), of the manifold. It must thus, (since it is a union), be produced, and produced furthermore by an inner activity that is valid for a given manifold in general, and that precedes *a priori* the manner in which the manifold is given. (Ak XI:314)

Kant's claim here is about the representation of a connection between several representations. His first point is that this representation is not an empirical representation. A connection between several representations is not 'given to us'. And this means that the representation of a connection between several representations must be spontaneous. It must be a representation that relates to its object *a priori*. As Kant goes on to say,

Since composition... cannot be given, but must be produced, it must rest on the pure spontaneity of the understanding in concepts of objects in general (of the composition of the given manifold...)

It may be objected, however, that if an act of synthesis is understood in the way that I have proposed, there is no room left for Kant's distinction between empirical synthesis and pure synthesis. After all, if an act of synthesis is understood in this way, all synthesis involves an *a priori* representation of a connection between the objects of several representations. And it seems natural to assume that the distinction between empirical synthesis and pure synthesis turns on the distinction between an empirical representation of a connection between the objects of several representations and an *a priori* representation of a connection between the objects of several representations. To see why this objection fails, we must remember that Kant's distinction between empirical synthesis and pure synthesis has to do with the character of the manifold of representation that is to be combined. I have already quoted Kant's claim that

Such a synthesis is *pure* if the manifold is given not empirically but *a priori*
(as is that in space and time). (A77/B103)

Kant thinks that an act of synthesis is empirical if the representations that are connected to one another are empirical representations – say, empirical intuitions. He thinks that an act of synthesis is pure if the representations that are connected to one another are *a priori* representations. So Kant's distinction between empirical synthesis and *a priori* synthesis is entirely consistent with the interpretation of Kant's concept of synthesis that I have introduced in this section.

How does this interpretation of Kant's concept of synthesis compare to other recent interpretations? Kitcher claims that

A synthesis is an act, or to be more neutral, a process that produces a representation, by adding or combining diverse elements contained in different cognitive states in a further state that contains elements from these states. (1990, p.74)

Kitcher's explanation of this definition makes it clear that what she has in mind is an extremely abstract interpretation of Kant's concept of synthesis. She takes the concept of synthesis to be the concept of an act of the mind whereby we start with a number of representations x , y , z etc. and then form for ourselves a new representation r , where it is the fact that we had the particular representations x , y , z etc. that made it possible for us to form the new representation r . If Kitcher is correct then someone has performed an act of synthesis insofar as she has acquired a particular representation r , where she could not have acquired r had she not first acquired a particular set of representations x , y , z etc. As she puts it,

Given a set of input states, a synthesis produces a certain output state.
Thus, Kant's talk of rules of synthesis, or functions of synthesis, is...
pleonastic. (p.74-75)

Suppose that it is true of our mind that if we have acquired a particular set of representations x , y and z , there are two further representations that we are in a position to form. We can form representation r , or we can form representation s . Kitcher thinks that this is tantamount to the claim that there are two possible acts of synthesis that we are in a position to perform with respect to this set of representations.

One difference between this interpretation of Kant's concept of synthesis and the interpretation that I have defended in this chapter is that if Kitcher is correct then it is relatively uncontroversial that our mind has the capacity to perform acts of synthesis. Suppose that I have formed a cognition of the roundness of the apple in my hand. This is a representation that I could not have formed had I not already acquired a number of other representations - for example, the concept of roundness, the concept of the apple in my hand, etc. Given those other representations I was in a position to form the representation of the roundness of the apple in my hand. On Kitcher's interpretation of Kant's concept of synthesis, it follows from this that I have performed an act of synthesis.

Van Cleve defends a very different interpretation of Kant's concept of an act of synthesis. He starts from the following claim:

To say that a collection of representations has been synthesised is not merely to say that they are apprehended together in one act; it is to say how they got that way. And on Kant's view, they can get that way only as the result of a special three-step procedure... (1999, p.85)

When Van Cleve writes that a collection of representations has been 'apprehended in one act', what he means is that we are conscious of them all at once. So his first claim about synthesis is that the result of an act of synthesis is that we are conscious of several representations all at once. His next step is to clarify the procedure that he thinks is required to perform an act of synthesis:

First, there is the Synthesis of Apprehension in Intuition, which consists of apprehending the representations to be unified separately and successively. Second, there is the Synthesis of Reproduction in Imagination, which consists of forming and retaining memory-images of each of the representations apprehended in the first phase. Finally, there is the Synthesis of Recognition in a Concept, which consists of surveying the images generated during the second phase and seeing what they all add up to. (1999, p.85)

Here is an example that Van Cleve provides to illustrate this concept of an act of synthesis:

The whole procedure may be illustrated by the adventures of a curious flea on the back of an elephant. First, it makes a complete circuit of the beast, successively apprehending its various parts; all the while it forms and retains memory-images of the parts thus apprehended; finally it looks these over and exclaims, "Trunk, tusks, tail - by Jove, it's an elephant!" (1999, p.85)

If Van Cleve is correct then our mind has performed an act of synthesis insofar as it has taken several representations of the same object and becomes conscious of what sort of object these representations are all representations of. Take the example of the house that I have already discussed. Van Cleve would presumably say that the mind performs an act of synthesis insofar as it takes the intuition of the door, the intuition of the roof,

the intuition of the floor, etc., reflects on these intuitions, and becomes conscious that what these intuitions are all representations of is a house.

The trouble with these alternative interpretations of Kant's concept of an act of synthesis is that they make it hard to understand why an act of synthesis counts as a unification or combination of several representations. This is especially clear in the case of Kitcher's interpretation. Kant defines synthesis as 'the action of putting different representations together with each other and grasping their manifoldness in one cognition' (A77/B103). But if Kitcher is correct then our mind has performed an act of synthesis if it has formed a representation x where it could not have formed this representation had it not first acquired certain other representations y, z etc. There is no requirement here that the representations y, z etc. should be parts of the representation x , or that x should be a representation of a connection between y, z , etc. There is no sense in which this act of the mind requires a manifold of representation to be grasped 'in one cognition'. And so there is reason to reject Kitcher's interpretation of Kant's concept of an act of synthesis. As for Van Cleve's interpretation of Kant's concept of an act of synthesis, the process that Van Cleve thinks constitutes an act of synthesis would be more accurately described as the process of collecting evidence on the basis of which we make a judgment about the nature of an object. It is hard to see how this process counts as any sort of unification or combination of several representations.

§3.4 The unity of apperception and the Pure Concepts of the Understanding

In §3.1, I defended a claim about the place of Kant's concept of the original synthetic unity of apperception in his Deduction of the Categories. I argued that this concept is introduced in the context of a claim about the conditions that must be satisfied in order for us to make use of our faculty of understanding – that is to say, in order for us to form a thought of an object or make a judgment about an object. Kant's claim is that the original synthetic unity of apperception is an essential ingredient in the use of our faculty of understanding – and, therefore, that if we are to make use of our faculty of understanding in relation to an object then this object must satisfy the conditions of the original synthetic unity of apperception. In §3.2, I discussed Kant's concept of apperception in general, and his concept of original apperception in particular. I argued that what Kant means by apperception in general is the consciousness of a state of mind as one that serves as a representation of an object and that what he means by original apperception is the consciousness of a state of mind as one that serves as an *a priori* representation of an object, which is to say, the consciousness of a state of mind as one to which an object conforms. In §3.3, I turned to Kant's concept of an synthesis. I argued that what Kant means by an act of synthesis is an operation of the mind whereby several of its concepts or intuitions become connected in a particular way in virtue of the fact that the mind has formed a representation of them as connected in this way. And I mentioned that if an act of synthesis is understood in this way, the consciousness of an act of synthesis must be a function of original apperception. My aim in the last section of this chapter is to use the interpretations of Kant's concept of

original apperception and his concept of synthesis that I have defended in §3.2 and §3.3 to construct an interpretation of his concept of the original synthetic unity of apperception. Once I have explained how this concept should be understood, I will finally be in a position to explain exactly what Kant means by the Principle of the Original Synthetic Unity of Apperception.

I propose that when Kant refers to the original synthetic unity of apperception (B136), or to original apperception and its necessary unity (B142), what he has in mind is the consciousness of a particular unity. In particular, what he has in mind is the consciousness of the necessary unity of the synthesis of our representations. To clarify this proposal it will be helpful to start with a point about our faculty of apperception in general. The point is that while Kant defines our faculty of apperception as an ability to become conscious of our representations – or, to be a little more precise, an ability to become conscious of the states of our mind as representations of objects, he also ascribes to the faculty of apperception the consciousness of the numerical identity of our representations on different occasions.

At the start of the third section of the first edition Deduction of the Categories, Kant repeats some of the points that he made in the Transition to the Transcendental Deduction of the Categories. He writes that '[the] possibility of an experience in general... rests on three subjective sources of cognition: sense, imagination, and apperception' (A115, cf. A94). And he notes that each of these 'subjective sources of cognition' has both an empirical use and also an *a priori* or transcendental use that '[makes] this empirical use itself possible' (A115). He goes on to describe the empirical uses of sense, imagination and apperception:

Sense represents the appearances empirically in perception, the imagination in association (and reproduction), and apperception in the empirical consciousness of the identity of these reproductive representations with the appearances through which they were given, hence in recognition. (A115)

The crucial point for our purposes is this: Kant clearly thinks that the role of the faculty of empirical apperception in experience is to achieve a sort of recognition. What Kant means by recognition in this context is the consciousness of 'the identity of these reproductive representations with the appearances through which they were given' (A115). I take it that what Kant means here is this: Suppose that I acquire an empirical intuition on some occasion – say, the empirical intuition of a patch of water as liquid. One of the conditions of the possibility of experience is that I must be able to reproduce this empirical intuition. I must be able to put myself back into the state of mind that represents the patch of water as liquid. And another condition of the possibility of experience is that I must be able to become conscious that the state of mind into which I have put myself when I exercise my faculty of reproduction is the same state of mind that I was put into when I first acquired the empirical intuition. I must be able to become conscious that it is the same representation of an object that I am conscious of now that I was conscious of before. And Kant thinks that it is the task of my faculty of apperception – specifically, it is the task of my faculty of empirical apperception – to provide me with this consciousness.

This interpretation of Kant's concept of recognition is supported by his claims at the start of the section of the first edition *Deduction of the Categories* entitled: *On the Synthesis of Recognition in the Concept*. Here Kant writes:

Without consciousness that that which we think is the very same as what we thought a moment before, all reproduction in the series of representations would be in vain. (A103)

These lines make it clear that what Kant means by recognition is the consciousness 'that that which we think is the very same as what we thought a moment before', which is to say, the consciousness that the representation of which we are currently conscious is the very same representation of which we were conscious on a previous occasion.

If our faculty of empirical apperception is able to provide us with consciousness of the numerical identity of our empirical representations on several occasions, it is reasonable to suppose that our faculty of original apperception is able to provide us with consciousness of the numerical identity of our *a priori* representations on several occasions. For as I explained in §3.2, the basic difference between our faculty of empirical apperception and our faculty of original apperception is that whereas our faculty of empirical apperception is our ability to become conscious of a state of mind as one that serves as an empirical representation of an object, our faculty of original apperception is our ability to become conscious of a state of mind as one that serves as an *a priori* representation of an object. Now in the course of §3.3, I mentioned that if my interpretation of Kant's concept of an act of synthesis is correct then it is the faculty of original apperception by means of which we become conscious of our representations of connections between several other representations. If we take these thoughts together, we can conclude that one of the functions that Kant ascribes to the faculty of original apperception is the consciousness of the identity of our representations of connections

between several other representations – or, to put this point another way, the consciousness of the identity of our acts of synthesis.

At this point it may be helpful to provide some examples to illustrate the idea of a consciousness of the identity of an act of synthesis on different occasions. Suppose that I take the concept of the cup in front of me and the concept of roundness. And suppose that I connect these concepts in an act of synthesis. That is to say, suppose that these concepts become connected in virtue of the fact that I have formed a representation of a connection between them. To be a little more precise, suppose that I form the representation that the concept of the cup in front of me stands to the concept of roundness as subject to predicate. Now suppose that on some other occasion I connect the concept of the Radcliffe Camera to the concept of roundness as subject to predicate. There are obvious respects in which this act of synthesis differs from the first act of synthesis. It consists in the combination of the concept of the Radcliffe Camera with the concept of roundness, as opposed to the combination of the concept of the cup in front of me with the concept of roundness. However, there are also respects in which this act of synthesis is the same as the first act of synthesis. Both acts of synthesis consist in the combination of one concept with another concept as subject and predicate. Both acts of synthesis also consist in the combination of one concept with the concept of roundness as subject and predicate. I suggest that what Kant means by consciousness of the identity of synthesis is consciousness of the fact that several judgments involve the very same act of synthesis in this sense.

Now in the course his Deduction of the Categories Kant appeals to the concept of the unity of synthesis. And I propose that what Kant has in mind when he refers to the unity of synthesis this is precisely the identity of an act of synthesis on several

occasions – that is to say, the fact it is one and the same act of synthesis that is performed on several occasions. This may not be the most obvious way to interpret Kant’s concept of the unity of synthesis. It is tempting at first glance to assume that what Kant has in mind when he refers to the unity of synthesis is the unity that arises from an act of synthesis. Remember that when the mind performs an act of synthesis with respect to several intuitions or several concepts, these intuitions or concepts are combined or connected together (A78/B103, B129-130), and that is as much as to say that they come to have a sort of unity. It is not unreasonable to suppose that it is precisely this unity of several intuitions or several concepts that Kant has in mind when he refers to synthetic unity or the unity of synthesis. However, there are also good reasons to reject this interpretation of the concept of synthetic unity. Here are some lines from the first edition Deduction of the Categories that I discussed in §3.2:

there are... three original sources... of the possibility of all experience... namely, *sense*, *imagination*, and *apperception*. On these are grounded (1) the *synopsis* of the manifold *a priori* through sense, (2) the *synthesis* of this manifold through the imagination, [and] finally (3) the *unity* of this synthesis through original apperception. (A94)

Kant’s claim here seems to be that there is one faculty of the mind that is responsible for acts of synthesis and another faculty of the mind that is responsible for the unity of synthesis. Kant seems to think that if there is to be a unity of synthesis then the mind must do more than perform an act of synthesis. And in that case the unity of synthesis cannot be the unity that comes about in virtue of the performance of an act of

synthesis. Whatever Kant means by synthetic unity, it cannot be the unity for which an act of synthesis is sufficient – that is to say, the connection or combination of several intuitions or several concepts. Kant appears to make the same point in the first chapter of the *Analytic of Concepts*:

The first [thing] that must be given to us for the sake of the cognition of all objects *a priori* is the *manifold* of pure intuition; the *synthesis* of this manifold by means of the imagination is the second thing, but [this] still gives [us] no cognition. The concepts that give this pure synthesis *unity*... are the third thing... and they rest on the understanding. (A78-79/B104)

Here Kant distinguishes the contribution of the imagination to ‘the cognition of all objects *a priori*’, which is to say, the synthesis of a particular manifold, from the contribution of the understanding, which is the unity of this synthesis. He seems to think that there could be a synthesis of the manifold even if there were no unity of this synthesis. But if that is correct then the unity of synthesis cannot be the unity that is produced by an act of synthesis. And the obvious way to interpret the unity of synthesis in light of this point is as the fact that it is one and the same act of synthesis that is performed on several occasions – that is to say, the identity of an act of synthesis on several occasions.

Now when Kant refers to the unity of synthesis in the course of his Deduction of the Categories, what he has in mind is a unity of synthesis that counts as necessary. For example, in a key passage of the first edition Deduction of the Categories, he writes that

this unity of consciousness [i.e. the transcendental unity of apperception] would be impossible if in the cognition of the manifold the mind could not become conscious of the identity of the function by means of which the this manifold is synthetically combined into one cognition. Thus the original and necessary consciousness of the identity of oneself is at the same time a consciousness of an equally necessary unity of the synthesis of all appearances in accordance with concepts... (A108)

I take it that what Kant has in mind when he refers to the necessary unity of synthesis is the idea that one and the same act of synthesis is performed whenever our mind connects several intuitions or several concepts together. Or, to put this point another way, Kant thinks that there is an act of synthesis such that whenever we connect several intuitions or concepts together, we perform this act of synthesis. Consciousness of the necessary unity of synthesis is consciousness that whenever we connect several intuitions or concepts in an act of synthesis, we are always doing the same thing, that is to say, we are connecting several intuitions or concepts in the very same way.

Now, at last, I am in a position to explain how Kant's concept of the original synthetic unity of apperception – or, as he also puts it, the transcendental unity of self-consciousness (B132), or the necessary unity of apperception (B135) – should be understood. I take it that what Kant has in mind when he refers to this unity of apperception is precisely the consciousness of the necessary unity of synthesis – that is to say, the consciousness of an act of synthesis as one that is performed whenever several intuitions or concepts are connected together. What he is talking about is what he calls in the passage from the first edition Deduction of the Categories that I quoted

above the consciousness ‘of the identity of the function by means of which the manifold is synthetically combined’ (A108). And I take it that when Kant refers in §19 of the second edition Deduction of the Categories to the ‘conditions of the original synthetic unity of apperception’, what he has in mind are precisely the conditions that must be satisfied if we are to become conscious of an act of synthesis as one that is performed whenever several intuitions or concepts are connected together.

If this is correct then we can also put forward a more precise explanation of Kant’s Principle of the Original Synthetic Unity of Apperception. Here is Kant’s statement of this principle in §17 of the second edition Deduction of the Categories:

The supreme principle of all intuition in relation to the understanding is that all the manifold of intuition stands under conditions of the original synthetic unity of apperception. (B136)

As I explained in §3.1, what Kant has in mind here is the claim that if we are to make use of our faculty of understanding in relation to a manifold of intuition then this manifold of intuition must satisfy whatever conditions it has to satisfy if we are to be able to achieve the original synthetic unity of apperception. The necessary conditions of the original synthetic unity of apperception are the necessary conditions of the use of our faculty of understanding. I now propose that this claim can be stated more precisely as follows: If we are to make use of our faculty of understanding in relation to a manifold of intuition then this manifold of intuition must satisfy whatever conditions it must satisfy if we are to be able to become conscious of an act of synthesis that is necessary in the sense that we perform this act of synthesis whenever we perform any

act of synthesis whatsoever. These intuitions must be connected to one another in such a way that we can become conscious not only that they have been connected together, but also that one of the things that has gone on here is an act of synthesis that occurs whenever a number of intuitions are connected together.

It may be objected that this interpretation of Kant's concept of the original synthetic unity of apperception fails to account for the fact that what Kant often seems to have in mind when he refers to the necessary unity of apperception or consciousness is the fact that our consciousness or apperception itself is one and the same on different occasions – as opposed to the fact that we are conscious of an act of synthesis that is one and the same on different occasions. Consider for example Kant's claim in §16 of the second edition Deduction of the Categories that the original apperception

is that self-consciousness which, because it produces the representation I think, which must be able to accompany all others and which in all consciousness is one and the same, cannot be accompanied by any other representation. (B132)

This is followed by the remark:

I also call its unity [i.e. the unity of original apperception] the transcendental unity of self-consciousness... (B132)

What Kant seems to mean here by the unity of original apperception – that is to say, the transcendental unity of self-consciousness – is precisely the fact that original

apperception is 'in all consciousness one and the same'. How can this be reconciled with the interpretation of Kant's concept of the original synthetic unity of apperception that I have proposed in this chapter? To answer this objection it is worth bearing in mind the question: What are the conditions under which a consciousness on one occasion is one and the same as the consciousness on some other occasion? I propose that a consciousness on one occasion is the same as a consciousness on another occasion in the sense that Kant has in mind insofar as it is a consciousness of the same thing, e.g. the same act of synthesis. If this is correct then there is a clear connection between the consciousness of the unity of synthesis and the consciousness of the unity of consciousness itself.

Once we have understood Kant's Principle of the Original Synthetic Unity of Apperception in this way, it isn't hard to see how the rest of his Deduction of the Categories is supposed to work. What Kant thinks he has established so far in the argument is that if someone has made use of his faculty of understanding – if, that is, someone has made a judgment to the effect that the cup on the desk in front of him is round, or that there are clouds over Oxford – this person must be conscious of the necessary unity of synthesis. He must be conscious that he has performed an act of synthesis that is necessary in the sense that it is performed whenever any other act of synthesis is performed. He must be conscious that he has done something that he must do whenever he connects a number of concepts or judgments together.

Now, I propose that this last claim is equivalent for Kant to the claim that this person must have in mind a special sort of concept. To see why this is true we need to bear in mind what Kant means by a concept in general. Consider these lines from a footnote to §16 of the second edition Deduction of the Categories:

The analytical unity of consciousness pertains to all common concepts as such, e.g. if I think of red in general, I thereby represent to myself a feature that (as a mark) can be encountered in anything, or that can be combined with other representations; therefore only by means of an antecedently conceived possible synthetic unity can I represent to myself the analytical unity. A representation that is to be thought of as common to several must be regarded as belonging to those that in addition to it also have something different in themselves. (B133)

I shall not consider the claim that Kant is making in this passage. The key point for our purposes is that one of the things that Kant does here is to tell us a bit about what is involved in thinking of red in general, which is to say, what is involved in having in mind the concept of red. What Kant tells us is that when someone has the concept of red in mind, he has in mind a representation of what several other representations have – or at least could have – in common. However, the claim that Kant takes himself to established so far in his Deduction of the Categories is precisely that what is required for us to make use of our faculty of understanding, or to make a judgment about an object, is for us to become conscious of what several representations have in common – specifically, it is the claim that what is required for us to make use of our faculty of understanding is that we should become conscious of an act of synthesis that is involved in every representation of a connection between several intuitions or concepts. This consciousness of the unity of synthesis – that is to say, this consciousness of the

sameness of the synthesis that is involved in several states of our mind – clearly counts as a concept on Kant’s definition.

There is excellent textual evidence to support the claim that Kant identifies the consciousness of the unity of synthesis with a concept. First, remember that in the section of the first edition Deduction of the Categories Kant writes:

If, in counting, I forget that the units that now hover before my senses were successively added to each other by me, then I would not cognize the generation of the multitude through this successive addition of one to the other, and consequently I would not cognize the number, for this concept consists solely in the consciousness of this unity of synthesis. (A103)

Here Kant clearly states that the concept of a particular number consists in the consciousness of a unity of synthesis. Second, in the first chapter of the *Analytic of Concepts*, Kant writes:

The first thing that must be given to us *a priori* for the cognition of all objects is the manifold of pure intuition; the synthesis of this manifold by means of the imagination is the second thing, but it still does not yield cognition. The concepts that give this pure synthesis unity, and that consist solely in the representation of this necessary synthetic unity, are the third thing necessary for the cognition of an object that comes before us, and they depend on the understanding. (A78-79/B104)

This passage is surely decisive evidence for the claim that I have ascribed to Kant. He clearly thinks that the Categories or Pure Concepts of the Understanding consist in the consciousness of the necessary unity of the synthesis of our representations. So if he has shown that making use of our faculty of understanding involves being conscious of the necessary unity of the synthesis of our representations, he must think that he has shown that making use of our faculty of understanding involves a Category or Pure Concept of the Understanding.

§4 The Objective Validity of Judgment and the Original Synthetic Unity of Apperception

In §3, I defended an interpretation of Kant's Principle of the Original Synthetic Unity of Apperception. I argued that this principle is a claim about the conditions that must be satisfied if we are to make a judgment about an object – that is to say, a claim about the conditions that must be satisfied if we are to make use of our faculty of understanding. I argued that this principle consists in the claim that if we are to make a judgment about an object then those conditions must be satisfied that must be satisfied if we are to achieve a special sort of consciousness – in particular, the consciousness of the necessary unity of our synthesis of our representations. Finally, I argued that the condition that must be satisfied if we are to achieve this special sort of consciousness is that we must form a pure concept of an object – that is to say, we must make use of the Categories.

The Principle of the Original Synthetic Unity of Apperception is the central claim in Kant's Deduction of the Categories. If Kant is correct that that we must be conscious of the necessary unity of our synthesis of our representations if we are to make a judgment about an object then he is also entitled to conclude that the conformity of an object to the Categories or Pure Concepts of the Understanding is required for this object to count as an object of our faculty of understanding – and, therefore, as an object of empirical cognition. But in that case he can draw a further conclusion about the Categories or Pure Concepts of the Understanding. He can conclude that these concepts serve as representations of objects insofar as they count as

objects of experience. He can conclude that they serve as the blueprints for an object of experience.

My aim in this chapter is to address the crucial question: What is Kant's justification for the Principle of the Original Synthetic Unity of Apperception? Why does Kant think that this claim is true? My answer to this question can be stated briefly as follows: Kant thinks that if we reflect on what is special about the act of judgment, we find that making a judgment to the effect that, say, there are clouds over Oxford, or that England will win the forthcoming Ashes series, involves thinking that certain concepts or judgments ought to be connected in a particular way – or, in other words, that if someone connects these concepts in another way then he has done something wrong. In making a judgment, we think that connecting concepts in a particular way is something that is required. But Kant maintains that we can think of such a requirement only if we are conscious that all consciousness of connections between concepts has a sort of unity, which is to say, only if we are conscious that all consciousness of connections between concepts or judgments constitutes a sort of whole. And Kant thinks that we can become conscious of this unity of consciousness only if we are conscious that our synthesis of our representations has a unity that is necessary in the sense that I discussed in §3.4.

This is perhaps the most original part of the interpretation that I defend in this thesis, and it will require careful exposition. In §4.1, I focus on the claim that an act of judgment involves the thought that a set of representations ought to be connected in a particular way. I ascribe this claim to Kant on the basis of what he tells us in his famous discussion of the distinction between objectively valid and subjectively valid empirical judgments in the *Prolegomena*. In §4.2, I turn to the claim that the thought of

such a requirement is made possible by the consciousness of the unity of our consciousness itself. I emphasize that the unity of consciousness that is involved here is not restricted to the consciousness in a particular mind – say, your mind, or my mind, or the mind of the Queen of England. And I find evidence that Kant subscribes to this claim in both editions of the Deduction of the Categories. In §4.3, I compare Kant’s theory about what is involved in the act of judgment with the theories of other philosophers. And I argue that Kant’s theory constitutes a major advance over the theories of these other philosophers inasmuch as it involves the recognition of the distinctively normative aspect of an act of judgment. Finally, in §4.4, I consider another attempt to explain the connection between empirical cognition and the original synthetic unity of apperception – specifically, the theory recently defended by Kitcher (2011).

§4.1 Objective validity and necessary, universal validity

To understand why Kant accepts the Principle of the Original Synthetic Unity of Apperception – that is to say, the claim we can make use of our faculty of understanding only if those conditions that are satisfied that must be satisfied if we are to achieve consciousness of the necessary unity of synthesis – it is helpful to consider what he says in §18 of the *Prolegomena*. Here Kant divides ‘empirical judgments’, which he defines as judgments that ‘have their basis in the immediate perception of the senses’ (Ak IV:297) into two classes. He writes:

Empirical judgments, insofar as they have objective validity, are JUDGMENTS OF EXPERIENCE; those, however, that are only subjectively valid I call mere JUDGMENTS OF PERCEPTION. The latter do not require a pure concept of the understanding, but only the logical connection of perceptions in a thinking subject. But the former always demand... special concepts originally generated in the understanding, which are precisely what make the judgment of experience objectively valid. (Ak IV:298)

Before I go any further, I should point out that what Kant says about empirical judgments in this passage has seemed to many commentators to be flatly inconsistent with what he says about the act of judging in general in the *Critique*. Remember that in §19 of the second edition Deduction of the Categories, Kant rejects the definition of judgment as ‘the representation of a relation between two concepts’ on the grounds that there is something distinctive about the relation between representations that is represented in the act of judgment (B140). He then writes:

If... I investigate more closely the relation of given cognitions in every judgment, and distinguish that relation, as something belonging to the understanding, from the relation in accordance with laws of the reproductive imagination (which has only subjective validity), then I find that a judgment is nothing other than the way to bring given cognitions to the *objective* unity of apperception. (B141)

‘Only in this way’, he goes on to say,

does there arise from this relation *a judgment*, i.e. a relation that is *objectively valid*, and that is sufficiently distinguished from the relation of these same representations in which there would be only subjective validity... (B142)

Now, there are at least two discrepancies between these passages from the *Critique* and §18 of the *Prolegomena*. Here's the first problem: In the *Prolegomena* Kant ascribes objective validity to judgments themselves. For example, he defines judgments of experience as empirical judgments that are also objectively valid (Ak IV:298). But in §19 of the second edition Deduction of the Categories, Kant ascribes objective validity to the relation between concepts or judgments that is represented in an act of judging. He appears to think that a judgment involves the representation of a relation between two concepts or two or more judgments as objectively valid. He does not describe the judgments themselves as objectively valid.

The obvious way to iron out this particular crease is to suppose that a judgment counts as objectively valid in the sense that Kant has in mind in the *Prolegomena* just in case it consists in the representation of a relation between two concepts or two or more judgments as objectively valid. But this solution to the problem only serves to make the second discrepancy between §18 of the *Prolegomena* and §19 of the second edition Deduction of the Categories even more striking. In the *Prolegomena* Kant divides empirical judgments into those that are objectively valid and those that are only subjectively valid, which is to say, those in which what is represented is the objective validity of the relation between two concepts or two or more judgments, and those in which what is represented is the mere subjective validity of the relation between two

concepts or two or more judgments. In the Deduction of the Categories, however, Kant contrasts the act of judging in general, which is the representation of a relation between two concepts or two or more judgments as objectively valid, with the representation of a relation between the same concepts or judgments as merely subjectively valid. This time Kant assumes that a judgment as such counts as objectively valid. But if all judgments as such count as objectively valid, there can be no such thing as a judgment that is merely subjectively valid – and, therefore, no such thing as a judgment of perception (cf. Kemp Smith 1918, p.289, Beck 1978, p.50).

Another discrepancy between §18 of the *Prolegomena* and the *Critique* has to do with Kant's claim that judgments of perception 'do not require a pure concept of the understanding' (Ak IV:289). The trouble is that in the *Critique* Kant seems to think that all judgments whatsoever require the use of pure concepts of the understanding. For example, in §20 of the second edition Deduction of the Categories, having reminded his reader that '[the] action of the understanding... through which the manifold of given representations... is brought under an apperception in general, is the logical function of judgments', he goes on to say that since 'the categories are nothing other than these very functions for judging', 'the manifold in a given intuition also necessarily stands under the Categories' (B143). What matters here is Kant's claim that the Categories are the functions for judging. It is reasonable to conclude on this basis that someone must make use of the Categories if she is to engage in an act of judging. But in that case there can be no judgments that do not require the use of Categories, and, therefore, no judgments of perception (cf. Sassen 2008, p.271).

Some commentators have concluded on the basis of these discrepancies that what Kant outlines in the *Prolegomena* is a theory of empirical cognition that is flatly

inconsistent with the one that he defends in the *Critique* (cf. Beck 1978, p.49, Longuenesse 1997, p.180). Kemp Smith goes so far as to say that the distinction between judgments of experience and judgments of perception ‘cuts at the very root of Kant’s Critical teaching’ (1918, p.288). More recently, Sassen sides with ‘those who claim that in the absence of the categories [judgments of perception] cannot be judgments proper’, and adds that ‘all efforts to defend Kant against the charge of inconsistency are misguided’ (2008, p.281). If these commentators are right then it is dangerous to use what Kant says in the *Prolegomena* to elucidate his claims about the act of judgment in the Deduction of the Categories.

Allison proposes that we can avoid any tension between §18 of the *Prolegomena* and the Deduction of the Categories if we understand the distinction between judgments of perception and judgments of experience as a distinction between ‘two conceptions of judgment, rather than, as initially presented, two kinds of empirical judgment’ (2004, p.180). The idea here is that what Kant means by a judgment of perception is an empirical judgment ‘according to the common view (the standard empiricist account)’ (2004, p.180), whereas what Kant means by a judgment of experience is an empirical judgment as he himself understands it. The trouble with this solution to the problem is that it flies in the face of the text of the *Prolegomena*. Kant clearly defines judgments of experience and judgments of perception as different sorts of empirical judgment. There is no evidence whatsoever that he is talking about a distinction between two different theories or conceptions of empirical judgment.

Another commentator who thinks that the two texts can be reconciled is Longuenesse (1997). She explains the source of the difficulty in the following way. In §19 of the second edition Deduction of the Categories, Kant writes that only in a

certain way ‘does there arise... a judgment, i.e. a relation that is objectively valid, and that is sufficiently distinguished from the relation of these same representations in which there would be only subjective validity’ (B142). He then adds that ‘[in] accordance with the latter I could only say: if I carry a body, I feel an impression of weight, but not: it, the body, is heavy’ (B142). Longuenesse remarks that ‘[it] is of course tempting to see in this confrontation... an echo of the distinction made in the *Prolegomena* between judgments of perception and judgments of experience’ (1997, p.187), and to conclude that Kant ‘renounces the opposition between two kinds of judgment in favor of the opposition between mere empirical and association and judgment’ (1997, p.187). But Longuenesse insists that this temptation must be resisted. On her view, Kant does not commit himself in §19 of the second edition Deduction of the Categories to the claim that all judgments are objectively valid. His aim in this section of the *Critique* is to contrast the sort of judgment that would be ‘merely derived from empirical associations’ with the sort of judgment that arises from ‘a function of judging that we would consider as *original*’ (1997, p.187).

Whereas Allison’s solution to the problem flies in the face of the text of the *Prolegomena*, Longuenesse’s solution flies in the face of the text of the *Critique*. Kant’s aim in §19 of the second edition Deduction of the Categories is to work out what is distinctive about the relation between concepts or judgments that is represented in an act of judging – that is to say, ‘in every judgment’ (B141), ‘in a judgment in general’ (B140). And in the course of his discussion he contrasts this relation with ‘the relation in accordance with the laws of reproductive imagination (which has only subjective validity)’ (B141, cf. B142). He also clearly writes that only in a certain way ‘does there arise a *judgment*, i.e. a relation that is *objectively valid*’ (B142). Longuenesse does not

explain how we can interpret this except as a claim that implies that all judgments as such are objectively valid. There is no evidence in this section of the *Critique* that Kant allows for even the conceptual possibility of ‘judgments merely derived from empirical associations’.

Fortunately, there is another way to explain the discrepancies between the *Prolegomena* and the *Critique* on the subject of judging, objective validity, and the role of the Pure Concepts of the Understanding, that does violence to neither of these texts. I propose that Kant uses the word ‘judgment’ in more than one sense. In the sense that he has in mind in the *Critique*, a judgment is defined as the representation of an relation between two concepts or two or more judgments as objectively valid. In this narrower sense, there can be no such thing as a judgment of perception. But in the wider sense that Kant has in mind in the *Prolegomena*, an empirical judgment is defined as the representation of a relation between several perceptions, whether it is a representation of this relation as objectively valid or as merely subjectively valid (cf. Ak IV:301). If this is correct then there is no contradiction between Kant’s claims in the *Prolegomena* and the *Critique*. When Kant claims in the *Critique* that all acts of judgment involve the Categories, what he has in mind is the thought that all representations of relations between two concepts or two or more judgments as objectively valid involve the Categories. When in the *Prolegomena* he allows for acts of judgment that do not involve the Categories, what he has in mind is the thought that there can be representations of relations between several concepts or several judgments, that is to say, representations of relations between concepts or judgments as merely subjectively valid, that do not involve the Categories.

With these difficulties out of the way, we can consider the more substantive

claims that Kant makes in §18 of the *Prolegomena*. He writes:

All of our judgments are at first mere judgments of perception; they hold only for us, i.e. for our subject, and only afterwards do we give them a new relation, namely to an object, and intend that the judgment should also be valid at all time for us and for everyone else; for if a judgment agrees with an object, then all judgments of the same object must also agree with one another, and hence the objective validity of a judgment of experience signifies nothing other than its necessary universal validity. But also conversely, if we find cause to deem a judgment necessarily, universally valid... we must then also deem it objective, i.e. as expressing not merely a relation of a perception to a subject, but a property of an object; for there would be no reason why other judgments necessarily would have to agree with mine, if there were not the unity of the object - an object to which they all refer, with which they all agree, and, for that reason, also must harmonize among themselves. (Ak IV:298)

The claim that Kant is making here is neatly summarized at the start of §19:

Objective validity and necessary universal validity (for everyone) are therefore interchangeable concepts, and although we do not know the object in itself, nonetheless, if we regard a judgment as universally valid and hence necessary, objective validity is understood to be included. (Al IV:298)

If an empirical judgment is objectively valid, it is necessarily, universally valid. And if an empirical judgment is necessarily, universally valid, it is objectively valid. Or, to use the language of §19 of the second edition Deduction of the Categories: Someone represents a relation between two concepts or two or more judgments as objectively valid just in case she represents a relation between these concepts or judgments as necessarily, universally valid. Now, if it is true that only judgments of experience count as judgments in the sense that Kant has in mind in the *Critique*, we can draw the following conclusion: According to Kant, in order for someone to engage in an act of judging, she must represent a relation between two concepts or two or more judgments as necessarily, universally valid.

So what exactly does Kant mean by the necessary, universal validity of a judgment? Kant appears to characterize this sort of validity in at least two different ways in §18 of the *Prolegomena*. First, he expresses the claim that if a judgment is objectively valid then it is necessarily, universally valid as follows: If a judgment is objectively valid, we ‘intend that [it] should be valid at all times for us and for everyone else’ (Ak IV:298). So it seems that a relation between two concepts or two or more judgments counts as necessarily, universally valid insofar as we ‘intend that [it] should be valid at all times for us and for everyone else’. Second, he writes that if my judgment is objectively valid then ‘other judgments necessarily... have to agree with mine’. So it seems that a relation between two concepts or two or more judgments counts as necessarily, universally valid insofar as ‘other judgments necessarily... have to agree with mine’, that is to say, with a judgment in which these concepts or judgments are related in this way. Presumably, we should try to find a way of understanding Kant’s concept

of necessary, universal validity on which it makes sense to characterize it in both of these ways.

I suggest that what Kant has in mind when he refers to the necessary, universal validity of a judgment is the idea that when someone makes an objectively valid judgment – or, in other words, when someone represents a particular connection between two concepts or two or more judgments as objectively valid – part of what she is doing is thinking that this is the way that the concepts or judgments in question ought to be connected to one another. So for example, Kant thinks that part of what is involved in my judgment that the cup on the desk in front of me is round is the thought that the concept of the cup on the desk in front of me ought to be connected to the concept of roundness in a particular way. He thinks that this judgment involves the thought that a connection of concepts is in some sense required or demanded. And he thinks that the requirement to connect these concepts in this particular way is not limited to a particular person or a particular time. The thought that is involved in this judgment is that everyone who represents a connection between these concepts – no matter who they are or when they make their judgment – is required to connect the concepts in this particular way.

I believe that there is a very strong argument for this way of understanding Kant's concept of necessary, universal validity, based on the examples that Kant gives of judgments of perception in §19 of the *Prolegomena*. Before I can state this argument, I should mention a famous difficulty with these examples. Here is the key passage from §19:

Let us provide examples: That the room is warm, the sugar sweet, the wormwood repugnant, are merely subjectively valid judgments. I do not all require that I should find it so at every time, or that everyone else should find it just as I do; they express only a relation of two sensations to the same subject, namely myself, and this only in my present state of perception... these I call judgments of perception. (Ak IV:299)

Evidently, Kant thinks that the judgment that the room is warm is an example of a judgment of perception – that is to say, an example of an empirical judgment that is only subjectively valid. The same goes for the judgment that the sugar is sweet and the judgment that the wormwood is repugnant. The difficulty with these examples arises when we consider a claim that Kant makes in his discussion of judgment in §19 of the second edition Deduction of the Categories. He writes:

That is the aim of the copula ‘is’ in them [i.e. in judgments]: to distinguish the objective unity of given representations from the subjective... (B141-142)

He adds that

this word designates the relation of the representations to the original apperception and its *necessary unity*, even if the judgment is itself empirical, hence contingent, e.g. bodies are heavy. (B142)

Suppose that I form a representation of a connection between two concepts. And suppose that this representation can be expressed as the representation that A is B – for example, the representation that all men are mortal or that Socrates is mortal. What Kant tells us in §19 seems to imply that the ‘is’ in this judgment is a mark of objective validity. If I have a representation that can be expressed as the representation that A is B then this representation is a representation of a connection as objectively valid. However, the examples that Kant gives in the *Prolegomena* of representations of connections between perceptions as merely subjectively valid are all representations of the form: A is B. They are the judgments that the sugar is sweet, the wormwood is repugnant, and the room is warm (Ak IV:299). So here we appear to have another discrepancy between the text of the *Prolegomena* and the text of the *Critique*.

Let’s take a closer look at the first of Kant’s examples of a judgment of perception: the judgment that the room is warm. Kant’s claim about this judgment is that in making it, ‘I do not all require that I should find it so at every time, or that everyone else should find it just as I do’. I suggest that we can see what Kant has in mind if we reflect on the following case. Suppose that on a bitterly cold night, a soldier comes into a guard hut at the end of his shift on sentry duty to wake the soldier whose job it is to relieve him. Suppose that this other soldier has been asleep in a bunk bed. The first soldier remarks ‘it’s warm in here!’ as he comes into the hut. The second soldier exclaims ‘it’s freezing in here!’ as he crawls out of bed. The first soldier, in making his judgment, may be perfectly aware that the second soldier will not find the room warm. And he will be unlikely to conclude from what the second soldier has said that this other soldier has made a mistake – that is to say, he will be unlikely to conclude that the other soldier has made the wrong decision about the way things are.

The first soldier will not think that the second soldier has failed to make the judgment that he ought to have made. For in making his own judgment, the first soldier did not think to himself that everyone else ought to make the same judgment. He did not require everyone else to make the same judgment as him. If, by contrast, the first soldier had come into the guard hut and said ‘this room has only one window’, and the second soldier had crawled out of bed and remarked ‘this room has three windows’, the first soldier would conclude that the second soldier had made a mistake. In making the judgment that the room has only one window, the first soldier would have been thinking that everyone else ought to make the same judgment about the number of windows in the room – that if anyone made the judgment that the number of windows in the room was not one, they would judge incorrectly.

If this reading of Kant’s remark that in making a judgment of perception ‘I do not all require that I should find it so at every time, or that everyone else should find it just as I do’ is correct, there is good reason to accept the interpretation of Kant’s concept of necessary, universal validity that I have proposed in this section. When Kant refers to the necessary, universal validity of a judgment, what he has in mind is the normative component of this representation – that is to say, the thought that not only have I connected two concepts or two more judgments in a particular way, but that if anyone else fails to connect these concepts or judgments in this particular way, they have not connected them in the way that they ought to connect them. Now it is important that this requirement should not be misunderstood. Kant’s claim is not that in making the judgment that, say, the cup of tea in front of me is round, I think that every person at every time ought to be connecting the concept of the cup of tea in front of me with the concept of roundness as subject to predicate. The requirement that is

represented in this act of judgment is just that if anyone at any time connects the cup of tea in front of me with the concept of roundness, this person ought to connect these concepts as subject to predicate.

How should we deal with the problem of the apparent inconsistency between Kant's examples of subjective validity and objective validity in the *Critique* and the *Prolegomena*? I propose that what the discrepancy between the two texts shows us is that Kant has changed his mind over the question whether the distinction between subjective validity and objective validity is marked by the words we use when we express our empirical judgments. In the *Critique*, Kant clearly thinks that the word 'is' serves as a mark of objective validity. By contrast, in the *Prolegomena*, Kant thinks that a judgment that is expressed in the same words can count either as a judgment of perception or as a judgment of experience. Here is another example that he provides:

if I say: the air is elastic, then this judgment is to begin with only a judgment of perception; I relate two sensations to one another. If I want it to be called a judgment of experience, I then require this connection to be subject to a condition that makes it universally valid. I want therefore that I, at every time, and also everyone else, would necessarily have to connect the same perceptions under the same circumstances. (Ak IV:300)

Evidently, Kant thinks that what is expressed by the judgment that the air is elastic can be either an objectively valid judgment or a merely subjectively valid judgment. What determines which sort of judgment is being made is what is going on in our mind. If what I am thinking when I say that the air is elastic is just that I have connected the

concept of the air to the concept of elasticity in a particular way, my judgment is only subjectively valid. If, however, I also think that everyone else ought to connect the concept of this air to the concept of elasticity in this particular way, my judgment is objectively valid.

§4.2 Necessary, universal validity and the unity of apperception

In §4.1, I started to construct an answer to the question why Kant accepts the Principle of the Original Synthetic Unity of Apperception. I drew attention to the fact that in the *Prolegomena* Kant commits himself to the claim that if someone represents a connection between two concepts or two or more judgments as objectively valid then he also represents this connection as necessarily, universally valid. And I defended an interpretation of Kant's concept of necessary, universal validity. If my interpretation is correct then someone represents a connection between two concepts or two or more judgments as necessarily, universally valid, insofar as he thinks not only that he has connected these concepts or judgments in a particular way, but also that these concepts or judgments ought to be connected in this particular way. To put this point another way, Kant thinks that part of what is involved in representing the objective validity of a connection between two concepts or two or more judgments is thinking that this connection is required or demanded.

I propose that Kant is also committed to a claim about the connection between the necessary, universal validity of a connection between two concepts or two or more judgments and the original synthetic unity of apperception. Specifically, he is committed to the claim that I can think of a particular connection between two

concepts or two or more judgments as one that is required or demanded in the sense discussed in §4.1 – that is to say, I can think that these concepts or judgments ought to be connected in a particular way – only if I have used my faculty of original apperception to become conscious of the necessary unity of synthesis. My aim in this section is to defend the proposal that Kant is committed to this claim.

It will be helpful first of all to introduce an aspect of Kant’s discussion of the distinction between judgments of perception and judgments of experience that I neglected in §4.1. There I mentioned two ways in which Kant characterizes necessary, universal validity. I considered the idea that a connection between two concepts or two or more judgments counts as necessarily, universally valid insofar as we ‘intend that [it] should be valid at all times for us and for everyone else’, and also the idea that a connection between two concepts or two or more judgments counts as necessarily, universally valid insofar as ‘other judgments necessarily... have to agree with mine’, that is to say, with the judgment in which the concepts or judgments are connected in this particular way. However, Kant also writes in §20 of the *Prolegomena* that

judging can be of two types: first, when I merely compare the perceptions and connect them in a consciousness of my state, or second, when I connect them in a consciousness in general. The first judgment is merely a judgment of perception has thus far only subjective validity... (A1 IV:300)

And in §22, he writes that

Judgments are... either merely subjective, if representations are related to one consciousness in one subject alone and are united in it, or they are objective, if they are united in a consciousness in general, i.e. are united necessarily therein. (Ak IV:304)

Kant clearly thinks that a judgment counts as necessarily, universally valid insofar as what we do in making this judgment is connect two concepts or two or more judgments 'in a consciousness in general'. If we simply connect the same concepts or judgments in our own consciousness – that is to say, in a consciousness of our own state of mind – then our judgment does not count as necessarily, universally valid. We should read this as follows: When we make a judgment that is necessarily, universally valid, what we do is to think that this is the way that two concepts or judgments ought to be connected in every consciousness whatsoever, that is to say, 'in a consciousness in general', not just in our own consciousness at this moment.

Remember that two judgments can be said to disagree with one another in the sense that Kant has in mind in the *Prolegomena* insofar as the first judgment consists in the consciousness that two concepts or two or more judgments are connected in a particular way, and the second judgment consists in the consciousness that these same concepts or judgments are not connected in this way. This concept of the agreement or disagreement of one judgment with another can be easily extended to cover consciousness itself. One consciousness can be said to disagree with another consciousness insofar as the first consists in the consciousness that two concepts or two or more judgments are connected in a particular way, and the second consists in the consciousness that these same concepts or judgments are not connected in this way.

I propose that Kant is committed to the claim that there is a requirement that two consciousnesses agree with one another if and only if these consciousnesses have unity. Suppose that I am conscious at one moment that I have connected two concepts A and B in a particular way. And suppose that I am conscious at some other moment that I have connected two concepts C and D in a particular way. If these consciousnesses are entirely distinct – that is to say, if they lack unity – then there is no requirement that they should agree with one another. There is no requirement that if C and D are the same concepts as A and B then they are connected in the same way in the second consciousness as in the first consciousness. If, however, there is a unity of these consciousnesses then there is also a requirement that they should agree with one another. What's more, it is only if there is a unity of these two consciousnesses that there is a requirement that they should agree with one another. There is no reason why two consciousnesses must agree with one another if they do not have unity.

Now, if Kant is committed to this claim then it is plausible to ascribe to him a further claim about the connection between necessary, universal validity and the original synthetic unity of apperception. Kant thinks that a judgment counts as objectively valid only if it involves the thought not only that I have connected two concepts or two judgments in a particular way in my consciousness, but also that these concepts or judgments ought to be connected in the same way in every consciousness. Now, suppose that I am conscious that two concepts or two or more judgments have been connected in a particular way in my consciousness. And suppose that I am also conscious of the fact that all consciousness has unity. I suggest that according to Kant if I am conscious of this then I am in a position to think that there is a requirement that these concepts or judgments should be connected in the same way in every other

consciousness as they have been in my own consciousness. For I am in a position to think that since all consciousness has unity, there is a requirement that every consciousness agrees with every other consciousness – and, in particular, with my consciousness of this connection between two concepts or two or more judgments. What’s more, I suggest that according to Kant it is only if I am conscious of this that I am in a position to think that there is a requirement that two concepts or two or more judgments should be connected in the same way in every other consciousness as they have been in my own consciousness. Unless I am conscious of the unity of all consciousness I am in no position to move from the thought that I have connected two concepts or two or more judgments in a particular way to the thought that these concepts or judgments ought to be connected in this way. Finally, I suggest that according to Kant I am conscious of the unity of all consciousness if and only if I have achieved the sort of consciousness that Kant has in mind when he refers to the original synthetic unity of apperception – that is to say, the consciousness of the necessary unity of synthesis. Only if I am conscious that whenever I connect two concepts or two judgments together, part of what is going on is the very same act of synthesis, am I conscious of the unity of all consciousness itself (cf. §3.4). And so only if I have achieved the sort of consciousness that Kant has in mind when he refers to the original synthetic unity of apperception am I in a position to make a judgment that counts as objectively valid – that is to say, a judgment in the sense that Kant has in mind in the *Critique*.

So far in this chapter, what I have done is to make a proposal about the connection between the original synthetic unity of apperception and the thought that a connection between two concepts or two or more judgments is necessarily, universally

valid. I have ascribed to Kant the claim that someone can think that a connection between two concepts or two or more judgments is necessarily, universally valid – that is to say, he can think that these concepts or judgments ought to be connected in a particular way – only if he is conscious of the necessary unity of synthesis. And I have proposed that Kant accepts this claim for the following reason: He thinks that only if someone is conscious of the necessary unity of synthesis is he conscious of the unity of all consciousness itself. And he thinks that only if someone is conscious of the unity of all consciousness itself is he in a position to think that since he has connected two concepts or two or more judgments in a particular way in his own consciousness, there is a requirement that these concepts or judgments should be connected in the same way in every other consciousness.

What evidence is there that this is an accurate account of Kant's reasons for accepting the Principle of the Original Synthetic Unity of Apperception? There are two passages in the *Prolegomena* that are worthy of special attention. First, Kant writes in §20:

A completely different judgment... occurs before experience can arise from perceptions. The given intuition must be subsumed under a concept that determines the form of judging in general with respect to the intuition, connects the empirical consciousness of the latter in a consciousness in general, and thereby furnishes empirical judgments with universal validity; a concept of this kind is a pure *a priori* concept of the understanding... (Ak IV:300)

Here Kant tells us that the role of the Pure Concepts of the Understanding in empirical cognition is to render our judgments universally valid by connecting the 'empirical consciousness' of a given intuition 'in a consciousness in general'. This suggests that he thinks that the Categories serve as conditions of the possibility of empirical cognition because they are required to achieve a connection between consciousnesses – specifically, to achieve the thought that the consciousness of a particular connection between several perceptions is connected to every other consciousness.

A second and more important passage is the one that I quoted in §4.1, where Kant defends the claim that a judgment counts as objectively valid if and only if it counts as necessarily, universally valid. The crucial aspects of this passage for our purposes are the reasons that Kant gives for this claim:

All of our judgments are at first mere judgments of perception... only afterwards do we give them a new relation, namely to an object... if a judgment agrees with an object, then all judgments of the same object must also agree with one another, and hence the objective validity of a judgment of experience signifies nothing other than its necessary universal validity. But also conversely, if we find cause to deem a judgment necessarily, universally valid... we must then also deem it objective... for there would be no reason why other judgments necessarily would have to agree with mine, if there were not the unity of the object – an object to which they all refer, with which they all agree, and, for that reason, also must harmonize among themselves. (Ak IV:298)

Here Kant clearly commits himself to the following claim: There is a requirement that two judgments agree with one another if and only if these judgments agree with the very same object. On the one hand, 'if a judgment agrees with an object, then all judgments of the same object must also agree with one another', and on the other hand, 'there would be no reason why other judgments necessarily would have to agree with mine, if there were not the unity of the object - an object to which they all refer, with which they all agree'. To put this point a little differently, Kant's claim is that there is a requirement that one consciousness of a connection between two concepts A and B agrees with another consciousness of a connection between two concepts C and D if and only if each of these consciousnesses is related to the very same object.

Now, this is very close to the key claim that I ascribed to Kant in my proposal about his reasons for thinking that there is a connection between necessary, universal validity and the original synthetic unity of apperception. I proposed that Kant is committed to the claim that there is a requirement that two consciousnesses agree with one another if and only if these consciousnesses have unity - if, that is to say, there is a unity of consciousness. It may be objected that my proposal is flawed precisely because what Kant actually claims in the *Prolegomena* is that there is a connection between the requirement that two consciousnesses agree with one another and the unity of the object of consciousness, not the unity of consciousness itself. However, there are two points to bear in mind here. First, the concept of the original synthetic unity of apperception is conspicuous by its absence from the text of the *Prolegomena*. It seems as if Kant is deliberately avoiding talk of the original synthetic unity of apperception in order to provide a more accessible introduction to his theory of the understanding. So we should not assume that if Kant doesn't explicitly mention the unity of consciousness

then this is not what he has in mind. Second, it is not unreasonable to think that, for Kant, consciousness of the unity of the object can only consist in consciousness of the unity of consciousness itself. Kant clearly thinks that there is a requirement that our consciousnesses agree with one another if and only if these consciousnesses are related to the very same object (Ak IV:298). This raises the question: How can we become conscious that two consciousnesses are related to the very same object – that is to say, how can we become conscious that the object with which one judgment agrees is the very same object with which another object agrees? After all, our consciousness of the unity or identity of the object of these consciousnesses cannot consist in yet another empirical judgment. For the objective validity of this other empirical judgment would require a further consciousness of the unity of its object with the object of the first two empirical judgments. I suggest that Kant thinks that the consciousness of the unity of the object can only be a consciousness of the unity of synthesis. Only if I am aware that the consciousness of a connection between two concepts A and B and the consciousness of a connection between two concepts C and D involve the very same act of synthesis can I become aware that these consciousnesses have the same object, and, therefore, that there is a requirement that they agree with one another. But consciousness of the unity of synthesis is the same for Kant as consciousness of the unity of consciousness itself. So there is a close connection for Kant between the consciousness of the unity of consciousness and the consciousness of the unity of the object of consciousness.

There is also evidence to support my proposal about the connection between necessary, universal validity and the original synthetic unity of apperception in the Deduction of the Categories itself. In the section of the first edition Deduction of the

Categories called: Of the Synthesis of Recognition in the Concept, Kant raises the following question:

What does one mean... if one speaks of an object corresponding to and therefore also distinct from... cognition? (A104)

He then goes on to make a number of claims about what is involved in the thought of an object corresponding to cognition – that is to say, about what is involved in the use of our faculty of understanding.

For the moment, I shall set aside the first part of Kant's answer to his question about the object corresponding to our cognition. This runs as follows:

It is easy to see that this object must be thought of only as something in general = X, since outside of our cognition we have nothing that we could set over against this cognition as corresponding to it. (A104)

What matters more at this stage is the second part of Kant's answer. He writes:

We find, however, that our thought of the relation of all cognition to its object carries something of necessity with it, since namely the latter is regarded as that which is opposed to our cognitions being determined at pleasure or arbitrarily rather than being determined *a priori*, since insofar as they are to relate to an object our cognitions must also necessarily agree with each other in relation to it, i.e. they must have that unity that

constitutes the concept of an object. (A104-105)

This sentence must be treated very carefully. The first clause of the sentence is unclear in more than one way. First, Kant's claim could be that 'our thought of the relation of all cognition to its object' includes the thought of a necessity, and in that sense 'carries something of necessity with it'. But his claim could also be that the having of this thought makes something necessary, whether or not this necessity is something that is thought of. Second, it's not clear how Kant's talk of the thought of the relation of 'all cognition to *its* object' (my emphasis) should be understood. Kant could be talking about the thought of the relation of a cognition in general to its object. But he could also mean the thought of the relation that all of our cognitions, taken together, bear to their object.

Regarding the first ambiguity, there is reason to prefer the first reading. For it is natural to take the second clause of the sentence as making the claim that when we think of the relation of all cognition to its object, we regard this object as 'that which is opposed to our cognitions being determined at pleasure or arbitrarily'. In other words, when we think of this relation to an object, we regard the object as a source of a certain necessity, and, in this way, we think of a certain necessity. As for the second ambiguity, while the second reading is the more natural one *prima facie*, there are at least two reasons to prefer the first reading. One is that Kant has not argued for the claim that all of our cognitions are related to one and the same object. Another is that it's not clear why consideration of the thought of the relation of all our cognitions taken together to an object is relevant to Kant's discussion. For what Kant has set out to investigate is the thought of an object that is required for cognition. And this is the

thought of an object 'that corresponds to this intuition' (A92-93/B125), i.e. that corresponds to some particular intuition.

So Kant's claim in the first clause of the key sentence from the section Of the Synthesis of Recognition in the Concept is that if someone thinks of an object corresponding to a cognition – that is to say, if she makes use of her understanding – then she thinks of a certain necessity. What does this necessity consist in? As in §18 of the *Prolegomena*, Kant characterizes the necessity in two ways. First, what he has in mind is the fact that it is necessary for our cognitions to be determined in some way. The thought of the necessity is the thought that our cognitions are 'determined *a priori*', not 'at pleasure or arbitrarily'. Second, what Kant has in mind a necessity of 'agreement' among cognitions. When we think of the relation of a cognition to its object, we think of a certain necessity, viz. we think that it is necessary for other cognitions to agree with this cognition. Now, it is reasonable to assume that a cognition counts as determined in a particular way if the concepts or judgments that this cognition contains are connected in a particular way. And if this is correct then to say that it is necessary for our cognitions to be determined in some way is to say that it is necessary for the concepts or judgments contained in those cognitions to be connected in a particular way. And I take it that what Kant has in mind here is the thought that these concepts or judgments ought to be connected in this particular way. What's more, it is reasonable to assume that two cognitions disagree in the sense that Kant has in mind in this sentence insofar as the same concepts or judgments are connected in a particular way in one cognition but are not connected in the same way in the other cognition. And in that case, it is reasonable to conclude that Kant's claim in the first half of the sentence under discussion is the same as the one that he makes

in §18 of the *Prolegomena*. He thinks that if we reflect on what is involved in the act of thinking of an object, we will see that it involves the thought that two concepts or two or more judgments ought to be connected in a particular way whenever they feature in a cognition – that is to say, the thought that there is a requirement that all other cognitions agree with the cognition in which they are connected in this way.

So far, I have ignored the final clause of the sentence from the section: *On the Synthesis of Recognition in the Concept*. Here Kant writes that insofar as our cognitions are to relate to an object, ‘they must also necessarily agree with each other in relation to it, i.e. they must have that unity that constitutes the concept of an object’ (A104-105). What does Kant mean by the claim that our cognitions must ‘have that unity that constitutes the concept of an object’? One key question here concerns the interpretation of the ‘i.e.’. This can be taken to mean that what follows is simply a restatement of what goes before. On this reading, to say that our cognitions necessarily agree with each other is just to say that they have a certain sort of unity. But the ‘i.e.’ can also be taken to suggest that what goes before logically implies what follows. On this reading, Kant’s claim is that our cognitions necessarily agree with each other, which means that they must have a certain sort of unity, for only if they have this sort of unity is it necessary for them to agree with one another. Now, the context makes it clear that the second reading is the right one. For Kant has already discussed the unity that constitutes the concept of an object in this section of the first edition *Deduction of the Categories*. And he does not define it in terms of a relation of agreement among cognitions. He writes:

If, in counting, I forget that the units that now hover before my senses were

successively added to each other by me, then I would not... cognize the number, for this concept consists solely in the consciousness of this unity of synthesis.

He then adds:

The word 'concept' itself could already lead us to this remark. For it is this *one* consciousness that unifies the manifold that has been successively intuited, and then also reproduced, into one representation... one consciousness must always be found, even if it lacks conspicuous clarity, and without that concepts... would be entirely impossible. (A103)

There's a lot that needs to be said about these lines. What matters for our purposes is just this: Kant claims that it is '*one* consciousness' – that is to say, a unity of consciousness – that constitutes the concept of an object. So when Kant tells us later on in the same section that 'insofar as [our cognitions] are to relate to an object [they] must also necessarily agree with one another... i.e. they must have that unity that constitutes the concept of an object', what he means is that cognitions that are related to an object must agree with one another, which means that there must be a unity of the consciousness in these cognitions. He thinks that it is only if there is such a unity of consciousness that there is any requirement that they should agree with one another.

If this interpretation of the key sentence from the first edition Deduction of the Categories is correct then here too it seems that, according to Kant, the connection between the unity of consciousness – specifically, the original synthetic unity of

apperception – and the thought of an object – that is to say, the use of our understanding – turns on the fact that the thought of the unity of consciousness is an essential ingredient in the thought that there is a requirement that our cognitions should agree with one another. And this is what I proposed in §4.1.

Further evidence for my proposal can be extracted from §19 of the second edition Deduction of the Categories. This is the section where Kant discusses the definition of an act of judging. Here Kant writes that the aim of the copula in a judgment is ‘to distinguish the objective unity of given representations from the subjective’. He then adds:

this word designates the relation of the representations to the original apperception and its *necessary unity*, even if the judgment is itself empirical, hence contingent, e.g. bodies are heavy. By that, to be sure, I do not mean to say that these representations *necessarily* belong to one another in the empirical intuition, but rather that they belong to one another *in virtue of the necessary unity* of the apperception in the synthesis of intuitions. (B142)

The proposal that I have defended in this chapter allows us to make sense of the final sentence of this passage. If my proposal is correct then Kant thinks that what is required for us to think of a connection between two concepts or two or more judgments as one that is objectively valid is for us to think that there is a requirement that these concepts or judgments should be connected in a particular way in every consciousness. What’s more, he thinks that we can think that there is such a requirement only if we are consciousness of the unity of consciousness itself. For only

if consciousness has unity is there are a requirement that every consciousness agrees with every other consciousness. And it is reasonable to suppose that this is what Kant has in mind when he writes that when representations are connected in a judgment – that is to say, when they are represented as objectively connected – they ‘belong to one another in virtue of the necessary unity of the apperception in the synthesis of intuitions’. He means that it is because of the unity of consciousness that there is a requirement that the concepts or judgments are connected in a particular way.

Now there is an obvious objection that can be made against the interpretation that I have defended so far in this chapter. This objection has to do with the scope of the original synthetic unity of apperception. On the interpretation that I defended §3, the original synthetic unity of apperception has to do with the consciousness of the necessary unity of all of the acts of synthesis that I myself perform – that is to say, the fact that there is an act of synthesis that I perform whenever I perform an act of synthesis. However, on the interpretation that I have defended so far in this chapter, the necessary, universal validity of a connection between several concepts has to do with the consciousness that several concepts ought to be connected in a particular way in every consciousness whatsoever, no matter who does the connecting or when they do the connecting. So it seems that the thought of the necessary, universal validity of a connection between several concepts requires the consciousness of more than just the necessary unity of my own consciousness. It requires the consciousness of the unity of every consciousness whatsoever.

Here is one plausible way to deal with this objection. When Kant refers to other minds in the course of his discussion of necessary, universal validity in the *Prolegomena* – that is to say, when he tells us that a judgment is necessarily, universally

valid only if it is valid not only for us but also for everyone else (Ak IV:298) – he is referring to minds the faculties of which are constituted in the same way as our own. Kant thinks that if I am conscious that a particular connection of concepts is valid for me at all times then I am also in a position to become conscious that this connection is valid for anyone else who connects representations in the same way as I do – that is to say, anyone else whose consciousness has the same sort of unity as my own. And this is why he thinks that the original synthetic unity of apperception – that is to say, the consciousness of the necessary unity of our own acts of synthesis – makes it possible for us to become conscious of a connection between several concepts as necessarily, universally valid.

§4.3 Theories of judgment

So far in this chapter, I have defended an account of Kant's reasons for thinking that the Principle of the Original Synthetic Unity of Apperception is true. I have argued that, according to Kant, when someone makes a judgment – that is to say, when someone makes use of his faculty of understanding – he thinks not only that he has connected two concepts or two or more judgments in a particular way, but also that these concepts or judgments ought to be connected in this way whenever anyone is conscious of a connection between them. And I have argued that, according to Kant, someone can think that two concepts or two or more judgments ought to be connected in a particular way on the basis that he has connected them in this way in his own consciousness only if he is conscious of the necessary unity of consciousness itself, which is to say, only if he can relate this connection of concepts or judgments to the

original synthetic unity of apperception. However, I have largely ignored the question whether Kant's analysis of the act of judging – that is to say, his analysis of our faculty of understanding – is defensible. My aim in this section is to address this question. I argue that Kant's analysis of the act of judging or the faculty of understanding has at least one clear advantage over the accounts of other philosophers. It involves the recognition of the distinctively normative aspect of an act of judgment. If we want to assess the philosophical merits of the theory of the understanding that Kant defends in his *Critique of Pure Reason* then this is the claim upon which we should focus our attention.

Kant certainly isn't the only philosopher to have seriously considered the question what is involved in the act of judgment. Brentano, in one of the later chapters of his 1874 treatise *Psychology From an Empirical Standpoint*, asks what it is that distinguishes a case in which I judge that, say, there is a book on the desk in front of me, from a case in which I merely entertain the thought that there is a book on the desk in front of me (1973, p.201-202). He goes on to consider and reject a series of answers to this question that he finds in the works of other philosophers. And it will be helpful to examine some of these other answers.

Perhaps the most famous account of what is distinctive about the act of judgment is due to Hume. In the first book of his *Treatise of Human Nature*, Hume notes that we 'conceive many things, which we do not believe', and sets out to discover what is added to an idea – or, as he also calls it, a conception – when we affirm it or assent to it. Hume's first claim is this:

as 'tis certain there is a great difference betwixt the simple conception of

the existence of an object, and the belief of it, and as this difference lies not in the parts or the composition of the idea, which we conceive; it follows, that it must lie in the manner, in which we conceive it. (1973, p.94-95)

Hume's claim here rests on the assumption that if there is a difference between the belief that some object exists and the mere idea or conception of the existence of this object, there are only two possible explanations for this difference. On the one hand, it may be that when we form the belief that an object – Hume's example is God – exists, we take some new idea and add it to, or join with it, our idea of God's existence. On the other hand, it may be that when we form the belief that an object exists, we have the very same idea as before, but we have this idea in a new way. Hume rejects the first explanation. He holds that 'the belief of the existence [of the object] joins no new ideas to those, which compose the idea of the object'. And he concludes that what is distinctive about a case in which I not only think of the existence of God, but also believe that God exists, must be the particular way in which I think of the very same idea.

At this point, Hume draws on a key assumption:

When you would any way vary the idea of a particular object, you can only encrease or diminish its force and vivacity. If you make any other change on it, it represents a different object... (1973, p.96)

Hume concludes on this basis that 'as belief does nothing but vary the manner, in which we conceive any object, it can only bestow on our ideas an additional force and

vivacity' (1973, p.96). In brief, Hume claims that the difference between a case in which I merely entertain the thought that snow is white, or that God exists, and a case in which I form the belief, or judge, that snow is white, or that God exists is just this: In the second case, my thought of the whiteness of snow, or the existence of God, has more force and vivacity, is firmer and more vivid (1973, p.97), than in the first case. Now Hume's conclusion here may be misunderstood. To say that one idea has more force than another may be taken to mean that it has a more powerful effect on other ideas or our desires. To say that one idea is firmer than another may be taken to mean that it is more persistent. And there is a place in the *Treatise* where Hume says things that can be taken to suggest that this is what he has in mind. He writes:

An idea assented to feels different from a fictitious idea... And this different feeling I endeavour to explain by calling it a superior force, or vivacity, or solidity, or firmness, or steadiness. This variety of terms, which may seem so unphilosophical, is intended only to express that act of the mind, which renders realities more present to us than fictions, causes them to weigh more in thought, and gives them a superior influence on the passions and imagination. (1973, p.629)

However, it would be a mistake to conclude that Hume thinks that the nature of a belief, or a judgment, consists in its causal relations to passions or other ideas (cf. Stroud 1977, p.74). Hume thinks that the essential difference between a belief and a mere idea has to do only with what it is to like to have the idea. He compares this difference at one point to the difference between the perception of pale and vivid

shades of the same colour. But he also thinks that beliefs have more influence on our passions and other ideas than mere ideas. And he thinks that they have this greater influence precisely because of their greater force or vivacity. This felt vivacity of the ideas 'gives them more force and influence; makes them appear of greater importance; infixes them in the mind; and renders them the governing principles of all our actions' (1973, p.74).

One bad objection to this theory has been advanced by Bennett, who remarks that if Hume's analysis of belief is correct then 'there is no difference between believing that the Sahara is warm and entertaining the thought that it is extremely hot' (1966, p.294). This cannot be right. Take the mere idea that the Sahara is warm and the idea that the Sahara is extremely hot. On Hume's theory, if someone forms the belief that the Sahara is warm then she has the first of these ideas in a lively and forceful manner. If you like, this idea is conceived in a more intense fashion. But that is not to say that what the idea is an idea of is in any way intensified or made more forceful. This would mark a difference in the content of the idea, on Hume's theory, not the manner in which the same idea was conceived. Yet the fact that Bennett makes this objection may be excused on the following grounds. Once the claim that a belief is a firm or forceful idea is distinguished from the claim that a belief is an idea that tends to linger in the mind, or that tends strongly to affect our other ideas and desires, it is very hard to see what this firmness or liveness could consist in besides a firmness or liveliness in the very content of the idea. I can make sense of Hume's talk of the forcefulness of an idea if what this means is that the idea tends to influence other ideas, desires etc. And I can make sense of Hume's talk of the vividness or liveliness of an idea if what this means is the liveliness or vividness of what the idea is an idea of, say, a colour or a smell. But if

this is not what Hume means, it is hard to see how his talk of force and vivacity should be understood. And so it is hard to say whether his claim about the nature of belief even makes sense. How can force or vivacity, in any meaningful sense of the words, attach to the way in which an idea is conceived?

It may be objected at this point that what Hume provides in the *Treatise* is a theory of belief, not a theory of judgment, and that his theory should not, therefore, be compared with Kant's analysis of judgment in the *Deduction of the Categories*. But there is good reason to read Hume's claims about belief as claims about the act of judgment. After all, even Hume's claim that the belief in the existence of God involves the idea of the existence of God seems false if belief is taken in the usual sense of the word. In order for me to believe, at some moment in time, that God exists, I need not currently entertain the thought that God exists. To have a belief is not to have any sort of idea, or to have an idea in any sort of way. Conversely, it is plausible to suppose that I must entertain the thought that God exists in order to judge that God exists.

At the end of his discussion of Hume's theory of belief, Stroud remarks that Hume 'never seems to have entertained the idea that [the] connection between belief and the passions and the will might constitute the very difference he seeks between belief and mere conception' (1977, p.74). He goes on to suggest that the reason why '[no] adequate theory of belief has been given to this day' is because 'it has been investigated in virtually complete independence from the notions of passion, desire, will and action' (1977, p.74). Now, the historical part of this last claim is false. One of the theories that Brentano criticizes in his discussion of the difference between judgment and mere thought is what he calls 'the unfortunate idea' that this difference has to do with the effect of a thought on our desires and actions (1973, p.202). He

ascribes this theory of judgment to Bain:

Because we let something influence our actions and volition in a special way when we judge it to be true as opposed to merely having a representation of it, he [i.e. Bain] thought that the difference between holding something to be true and mere representation simply consists in this influence upon the will. The representation that exerts such an influence becomes, by virtue of the very fact that it exerts it, a belief. (1973, p.202)

What Brentano has in mind here is Bain's claim that 'the difference between mere conceiving or imagining... and belief, is acting, or being prepared to act, when the occasion arises' (1872, p.372).

It is certainly plausible to maintain that one of the differences between merely entertaining the thought that, say, there is a book on the desk in front of me, and judging that there is a book on the desk in front of me, has to do with the connection between such states of mind and our actions, emotions and desires. If, for example, I judge that someone has been saying slanderous things about a friend of mine behind her back, I will be inclined to get angry with this person, and, perhaps, to take revenge upon him. If, by contrast, I merely entertain the thought that someone has been saying slanderous things about my friend, I will be, at the very least, less inclined to get angry, and not at all inclined to take any action. The question is: Could this constitute the essential difference between a case in which I judge that someone has been saying

slanderous things about my friend and a case in which I merely entertain the thought that he has been saying such things?

Brentano's answer to this question is No. This is because he thinks that there must be an intrinsic difference between the two acts of the mind – between, say, entertaining the thought that there is a book on my desk and judging that there is a book on my desk (1973, p.202-203) – that accounts for the fact that these two acts of the mind have the different effects that they do. The difference between thinking and judging, according to Brentano, cannot consist solely in the different relations that thinking and judging bear to other acts of the mind. There must be some reason why these acts of the mind have the different effects that they do. There must be something about an act of judging that leads it to have different effects to an act of mere thinking. Brentano is surely right about this. We could also point out that we have no need to examine the effects of an act of our mind on our actions and feelings – to decide whether or not we have really made a judgment. I can judge at this moment that there is a cup on the table in front of me. I am perfectly aware that I have done more than merely entertain the thought that there is a cup on the table in front of me. But I am not aware of any of the consequences of the act of my mind.

A third theory of judgment that Brentano considers in his *Empirical Psychology* focuses not on the effects of the acts of our mind, but rather on the causes of those acts of the mind. Brentano ascribes to James Mill and Herbert Spencer the claim that

the [representation] of a union of two characteristics is accompanied by belief when an inseparable association is formed between these two characteristics in consciousness, i.e. when the habit of presenting the two

characteristics together has become so strong that the presentation of the one invariably and irresistibly calls the other into consciousness and becomes connected with it. (1973, p.203)

What Brentano has in mind here is the idea that the difference between entertaining the thought that there is a cup on the desk in front of me and making the judgment that there is a cup on the desk in front of me consists in the fact that the second act of the mind is the result of association – that is to say, the result of the tendency of our mind, when it has frequently encountered two objects A and B together, to think of B whenever it thinks of A and to think of A whenever it thinks of B (cf. A100). On this theory, if I am conscious of a connection between two concepts – say, the concept of the cup on the desk in front of me and the concept of roundness – and I want to know whether I have judged that the cup on the desk in front of me is round, what I need to do is to ask why it is that these concepts are connected in the way that they are in my consciousness. If I find that the cause of my representation is the fact that my mind has got into the habit of thinking of the concept of roundness when it thinks of the concept of the cup on the desk in front of me then I can conclude that my consciousness amounts to the judgment that the cup on the desk in front of me is round.

Brentano's objection to this theory is similar to his objection to Bain's theory.

He writes that

just as Bain confused a distinctive feature of the consequences with the intrinsic property of holding a thing to be true, the elder Mill and Spencer

proposed as a distinctive feature of this sort of thought process a characteristic which, at best, they could only have called the possible cause of its distinctive character. (1973, p.204)

Brentano thinks that there is an intrinsic difference between merely entertaining the thought that there is a cup on the desk in front of me and making the judgment that there is a cup on the desk in front of me. He thinks that if association is the cause of judgment then that is because it is the cause of our being in a state of mind that is differentiated in this way from merely entertaining a thought.

The theory of judgment that Kant defends in his *Deduction of the Categories* has one substantial advantage over each of the theories that I have mentioned so far in this section. It involves the recognition of a key fact about the act of judging that, say, there is a cup on the desk in front of me, as contrasted with merely entertaining the thought that there is a cup on the desk in front of me. The fact in question is that an act of judging has a distinctively normative aspect. That is to say, when I judge that the cup on the desk in front of me is round, part of what I am thinking is that the concept of the cup on the desk in front of me ought to be connected in a particular way to the concept of roundness. It is a consequence of what I am thinking that if someone connects these concepts, but does not connect them in this way, they have gone wrong. They have failed to connect the concept of the cup on the desk in front of me and the concept of roundness as they ought to connect them. To put this point another way, part of what I am thinking when I judge that the cup on the desk in front of me is round is that there is a rule governing the connection of these concepts.

Kant is also surely right to resist the idea that the thought that a set of concepts ought to be connected in a particular way can be analyzed in terms of the thought that these concepts always have been connected in this way in the past. This is a point that Kant makes at the end of §19 of the second edition Deduction of the Categories. He writes that only in virtue of the transcendental unity of apperception

does there arise... a judgment, i.e. a relation that is objectively valid, and that is sufficiently distinguished from the relation of these same representations in which there would only be subjective validity, e.g. in accordance with the law of association. In accordance with the latter I could only say 'if I carry a body, I feel a pressure of weight', but not 'it, the body, is heavy', which would be to say that these two representations are combined in the object, regardless of any difference in the condition of the subject, and are not merely found together in perception (however often as that might be repeated). (B142)

Kant's claim at the end of this passage is that it is not enough for a connection between two concepts or two or more judgments to count as objectively valid that the concepts or judgments in question are generally found together. What is represented in a judgment is not just that there is what Hume calls a 'constant conjunction' of concepts. As Kant puts it in the *Prolegomena*, when I make the judgment that the air is elastic,

'[the] expansion is... represented not as belonging merely to my perception of the air in my state of perception or in several of my states or in the states of others but as necessarily belonging to it' (Ak IV:301).

It is because Kant recognizes that there is this distinctively normative aspect to the act of judgment that his theory of judgment is superior to Hume's.

§4.4 Kitcher on empirical cognition and the unity of apperception

I have already mentioned that there are other commentators who think that Kant's Deduction of the Categories turns on the claim that there is a connection between the original synthetic unity of apperception and a state of empirical cognition. One commentator who satisfies this description is Kitcher (2011). My aim in the final section of this chapter is to contrast Kitcher's account of the reasons why Kant is committed to such a claim with my own.

The first thing to bear in mind at this point is Kitcher's interpretation of Kant's concept of empirical cognition. As I explained in §2.3, Kitcher thinks that the sort of cognition that Kant is investigating in his Deduction of the Categories is 'not merely empirical, that is linked to sensory representations', but 'also conceptual'. And she claims that 'for Kant [this] means that it is a matter of 'marks' or grounds of cognition that are recognized as such' (2011, p.121). Kitcher equates this to the claim that the cognition that Kant has in mind in the Deduction of the Categories is 'rational empirical cognition', which is to say, '[empirical] cognition where the cognizer can give

the reason – or ground – of the cognition’ (2011, p.121). In order for me to achieve rational empirical cognition it is necessary for me not only to see or touch an object, say, the cup on the desk in front of me, or to bring this object under a concept, say, the concept of a body, but also to be in a position to ‘give the reason’ for bringing the object under this concept. I must be able to point to the fact that the object falls under some other concept, say, the concept of impenetrability, where there is a connection between falling under this other concept and falling under the concept of a body. Put simply, I must not only judge that an object I can see or touch etc. falls under some concept F, but must also be able to justify the judgment that the object falls under the concept F by appealing to the judgment that the object falls under some other concept G, where falling under the concept G serves as a mark of falling under the concept F.

Kitcher thinks that Kant’s strategy in his *Deduction of the Categories* is to investigate the conditions that must be satisfied in order for us to achieve a state of rational empirical cognition (e.g. 2011, p.85). And it is our ability to become aware of the grounds of our cognition that she thinks is particularly important. What’s more, Kitcher thinks that it is in the course of Kant’s investigation of the conditions that must be satisfied in order for us to achieve a state of rational empirical cognition that he discovers the importance of what she calls the Principle of Apperception (2011, p.115). She claims that the *Deduction of the Categories* ‘regresses from the possibility of rational empirical cognition to the truth of the apperceptive principle’, and then argues ‘that the principle of apperception requires that the categorical principles hold across all possible objects of the senses’ (2011, p.121). It should be clear from this that Kitcher’s interpretation of the structure of Kant’s *Deduction of the Categories* is very similar to the one that I have defended in this thesis.

Now what does Kitcher understand by the Principle of Apperception? One thing that is immediately clear is that she thinks that this principle is a claim about the conditions that must be satisfied if someone is to achieve a state of rational empirical cognition. At the start of her discussion of the Principle of Apperception Kitcher mentions what she takes to be Leibniz's claim that we are conscious that all of our various representations are states of a self that is also a persisting substance (2011, p.122). And she argues that we should expect Kant to treat this claim in the same way that he treats the Law of Causality, which is to say, the principle that every event has a cause. Just as Kant substitutes for the principle that every event has a cause the principle that '[it] is necessary for cognition that events are understood as standing in relations of necessary succession' (2011, p.123), we should expect him to substitute for the claim that all of our representations are states of a self that is also a persisting substance the principle that 'the assumption that different representations all or necessarily belong to a common subject is required for [rational empirical] cognition' (2011, p.122). This last claim is precisely what Kitcher finds in the first edition Deduction of the Categories. She expresses the Principle of the Apperception as follows:

All the varied empirical consciousness must be combined in one single self-consciousness (for cognition to be possible). (2011, p.123)

'All representations must (if cognition is possible) belong to a single self-consciousness.' Alternatively: it is necessary for cognition that representations are connected to a single 'I'. (2011, p.124)

She also writes that on Kant's theory,

The metaphysical proposition that any representation must belong with all others to a continuing self is replaced by the claim that in human cognition, empirical consciousness has a necessary reference to original apperception, that is, to one (continuing) consciousness. (2011, p.123)

Embedded in this principle is a claim that Kitcher calls 'the I rule', which is the claim that 'all representations belong with others to a single self' - or in other words: 'for anything that is a representation, it must belong with others to some continuing self' (2011, p.123). Kitcher clearly thinks that the Principle of Apperception asserts a connection between rational empirical cognition and this I rule. Unfortunately, it's not clear at first exactly what the connection between rational empirical cognition and the I rule is supposed to be. Kitcher writes at one point that '[part] of the burden of the A deduction is to establish this highest principle, the thesis that the holding of the I-rule is necessary for cognition' (2011, p.124), which suggests that the Principle of Apperception should be understood as the claim that only if the I rule holds - that is to say, only if it is true that 'all representations belong with others to a single self' - can someone achieve a state of rational empirical cognition. Further support for this interpretation is Kitcher's identification of the Principle of Apperception with the claim that 'these representations must be combined together in one identical self for cognition to be possible' (2011, p.124). She also writes that in the Principle of Apperception the I rule is 'removed from metaphysics and relocated among the

assumptions that must hold for cognition to be possible' (2011, p.124). On the other hand, Kitcher tells us that the Principle of Apperception has the same structure as the claim that 'the thesis of determinism [is] an implicit assumption that cognizers must make if they are to acquire any empirical cognition' (2011, p.6). She defines the Principle of Apperception as the claim that 'the assumption that different representations all or necessarily belong to a common subject is required for [rational empirical] cognition' (2011, p.122). And she talks of Kant's 'general program' of 'recasting' the doctrines of metaphysics as 'presuppositions of cognition' (2011, p.122). This suggests that the Principle of Apperception should be understood as the claim that someone can achieve a state of rational empirical cognition only if she assumes that the I rule is true – that is to say, only if she presupposes that 'all representations belong with others to a single self'.

To settle the question which of these interpretations of Kant's Principle of Apperception Kitcher accepts it will be helpful to consider her account of the reasons why Kant is committed to this principle. Her discussion focuses at first on the example of the application of the concept of being four in number to a set of objects – say, four strokes on a piece of paper (2011, p.128). Kitcher ascribes to Kant the claim that in order for someone to achieve rational cognition of the fact that there are four of these strokes it is necessary for her not only to apply the concept of being four in number to what she can see in front of her, but also to apply the concept of being one in number to the first stroke, the concept of being two in number to the first and second strokes etc., and to be conscious that it is on the basis of her application of the concept of being one in number to the first stroke, and her application of the concept of being two in number to the second stroke etc., that she has applied the concept of being four in

number to the set of strokes on the piece of paper. The idea here is that it is a mark of a set of objects falling under the concept of being four in number that it contains an object that falls under the concept of being two in number etc. As Kitcher puts it,

the counter needs to be conscious of his representations of '1', '2', and so on, as 'marks', or 'partial representations' that are the ground or basis of the whole representation '4'. (2011, p.128)

The point could also be made in terms of Kant's famous example of the wild man and the civilized man looking at a house. The civilized man achieves rational empirical cognition of the object that he can see only if he not only applies the concept of a house to this object but is also conscious that he is doing this on the basis that the object he can see has windows and doors and a chimney, etc. One important point that Kitcher thinks is made clear by this sort of example is that there is no need for the concepts that serve as 'marks', or 'grounds of cognition', to be such that if an object falls under these concepts then it follows that it falls under the concept in the judgment (Kitcher 2011, p.127, p.129). The concept of having windows, doors and a chimney can serve as a basis upon which to apply the concept of a house to an object in spite of the fact that an object can have windows, doors and a chimney even if it is not a house.

So far, this is just a more detailed explanation of Kitcher's interpretation of Kant's concept of empirical cognition. The point is that Kitcher ascribes to Kant the claim that achieving rational empirical cognition requires a person to be 'aware of her partial representations (e.g. the representations of a certain shape and of self-propelled motion) and of her mental act as the basis on which she judges the object to be a 'dog''

(2011, p.130). The first step in her account of the basis of the Principle of Apperception is to ascribe to Kant the claim that '[this] awareness not only permits cognition of the object', '[but also] essentially involves recognizing a relation across the partial cognitions or mental states and the judgmental state', viz. 'the relation of necessarily belonging together' (2011, p.130). To see what Kitcher means by this we need to know what she understands by this relation of 'necessary connection', or 'necessarily belonging together'. Here is her attempt to clarify this relation:

Since the partial representations are the basis of the judgment, it could not be a judgment – an example of [rational empirical] cognition – in their absence. Unless the partial representations could be used in producing such a judgment, they could not participate in [rational empirical] cognition. (2011, p.131)

Kitcher thinks that someone who achieves rational cognition that the strokes she can see in front of her are four in number

can see these representations [i.e. the representation of the first stroke, the representation of the second stroke etc.] as exemplifying a relation of necessarily belonging together, since the judgment of '1' would be impossible without the representation of the '1', and the representation of '1' would be impossible as ground of cognition without the judgment. (2011, p.136)

This is not especially easy to understand. I take it that what Kitcher has in mind is this: Suppose that the civilized man applies the concept of a house to the object that he can see in front of him. And suppose that he does this on the basis of his application to the same object of the concept of having windows. On the one hand, the concept of having windows owes its status as a 'ground of cognition', or 'mark', to the fact that it has served as the basis for the application of another concept – specifically, the concept of a house. On the other hand, the concept of a house owes the fact that its application to the object in a judgment constitutes a case of cognition to the fact that it has been applied on the basis of another concept – specifically, the concept of having windows. So the concept of a house and the concept of having windows depend on one another for the fact that they are elements in a state of cognition. If this is right, we can conclude that when Kitcher refers to the idea that rational empirical cognition requires the recognition of 'a relation across the partial cognitions or mental states and the judgmental state' – specifically, 'the relation of necessarily belonging together' (2011, p.130) – what she has in mind is the idea that it requires the recognition that the marks and the judgment depend on one another for their status as elements in a state of rational empirical cognition.

The next step in Kitcher's interpretation is absolutely crucial. She ascribes to Kant the claim that '[in] recognizing the relation of necessarily belonging together between [...] representations, [we] recognize them as exemplifying the I rule' (2011, p.136). As she also puts it,

When representational states are recognized as necessarily connected together, they are recognized as instances of the I rule. (2011, p.142)

Remember that what Kitcher means by the I rule is the rule that 'all representations must belong to a single consciousness' (2011, p.135), or, in other words, the rule that 'different representations all or necessarily belong to a common subject' (2011, p.122). So the claim that she is ascribing to Kant is that when someone recognizes that a collection of her representations are necessarily connected, she recognizes that these representations satisfy the rule that all representations must belong to a single consciousness or self.

Once again this is not easy to understand. A clue that can help us to make sense of Kitcher's interpretation is the fact that she thinks that the I rule works in a similar way to the rules that are associated with concepts in general. Kitcher thinks that all concepts are associated with rules. An example of such a rule in the case of the concept of a body is the rule that if an object is a body then it is extended. And she ascribes to Kant the claim that rules of this sort are involved in every state of cognition. She thinks, for example, that if someone recognizes an object as a body on the basis that this object is extended, he avails himself of the rule that if an object is a body then it is extended. Now, just like the rule associated with the concept of a body, the I rule mentions two concepts. The first is the concept of a representation and the second is the concept of belonging to an identical consciousness or self, viz. the I. And I take it that what Kitcher means by recognizing states of our mind as 'instances of the I rule' is applying to these states of our mind the concept of belonging to the I on the basis of the application to these states of our mind of the concept of a representation. For what the I rule claims is that if something is a representation then it belongs to the I. Furthermore, I take it that Kitcher identifies the concept of belonging to the I with the

concept of necessarily belonging together with – or, in other words, being necessarily connected to – other representations.

If this is right then the claim that Kitcher ascribes to Kant can be spelled out as follows: Two things are going on whenever someone achieves a state of rational empirical cognition. First, this person applies one concept to an object on the basis of the application to this object of some other concept. For instance, she applies the concept of body to this object on the basis of the application to this object of the concept of extension. Second, and crucially, this person also applies the concept of ‘necessary connection’, or ‘necessarily belonging together’, to her concepts – in this case, to her concept of body and her concept of extension. And she applies this concept of necessary connection to her concepts on the basis of her application to these concepts of the concept of representation. Just as she takes the concept of extension as the ground of her cognition of the object as a body, so she takes the concept of representation as the ground of her cognition of the concepts as necessarily connected to one another. Just as she represents the object as satisfying the rule that if something is a body then it is extended, so she represents her concepts as satisfying the rule that if something is a representation then it is necessarily connected to other representations.

At this point it is clear that the second of the two interpretations of Kant’s Principle of Apperception is the one that Kitcher accepts. She reads the Principle of Apperception as the claim that someone can achieve a state of rational empirical cognition only if she presupposes the truth of the I rule – that is to say, only if she presupposes that if something is a representation then it is necessarily connected to other representations in a single consciousness. She does not read it as the claim that someone can achieve a state of rational empirical cognition only if the I rule holds –

that is to say, only if it is true that if something is a representation then it is necessarily connected to other representations in a single consciousness

As I have already noted, there are many apparent similarities between Kitcher's interpretation of the Principle of Apperception and its place in Kant's Deduction of the Categories and the one that I have defended in this thesis. Kitcher thinks that Kant's strategy is to investigate the conditions that must be satisfied if someone is to achieve a state of empirical cognition (cf. §2.1). And she thinks that the Deduction of the Categories turns on the claim that there is a connection between a state of empirical cognition and the unity of consciousness or apperception (cf. §3.1). What's more, she thinks that the key to Kant's argument for his claim about the connection between a state of empirical cognition and the unity of consciousness turns on the idea that a state of empirical cognition requires the consciousness of a sort of necessary connection between a number of representations (cf. §4.1, §4.2). However, there are also some crucial differences between Kitcher's interpretation and the one that I have defended in this thesis. One difference concerns the interpretation of Kant's concept of empirical cognition. In §2.2, I argued that there is insufficient textual evidence for Kitcher's claim that what Kant means by a state of empirical cognition is a state of rational cognition – that is to say, a state in which someone not only makes a judgment that such and such is the case but also counts as conscious of the grounds of her judgment. I argued that the sort of state that interests Kant in his Deduction of the Categories is just one in which someone makes a judgment that such and such is the case.

Another related difference concerns the nature of the necessary connection consciousness of which is required for a state of empirical cognition. Kitcher thinks that this is a relation between the judgment and the grounds of the judgment in a state

of empirical cognition – say, between the judgment that the object I can see in front of me is a house and the recognition that this object has one of the marks of a house, viz. windows. The necessity of the connection consists in the fact that the judgment and the grounds of the judgment depend on one another for their status as elements of a state of cognition. If, however, the interpretation that I have defended in this thesis is correct, the necessary connection that Kant has in mind in his Deduction of the Categories is a connection between the various concepts or judgments that are connected in a judgment that such and such is the case – say, between the concept of the cup on the desk in front of me and the concept of roundness in the judgment that the cup on the desk in front of me is round. The necessity of the connection consists in the fact that what is thought in a judgment is that these concepts or judgments ought to be connected in a particular way, which is to say, that connecting these concepts or judgments in a particular way is required or demanded.

I propose that once Kitcher's account of the connection between rational empirical cognition and the unity of consciousness is spelled out in detail, it becomes clear that it is open to a serious objection. It is hard to see how Kitcher can avoid ascribing to Kant the claim that all rational empirical cognition of objects involves cognition of the unity of the self. Suppose that someone achieves rational empirical cognition of an object as a body. If Kitcher is right then Kant thinks that this cognition involves judging that the object is a body on the basis that it is, say, extended. He thinks it involves making use of the rule – what Kitcher calls an 'object rule' (2011, p.130) – that if something is a body then it is extended. Furthermore, if Kitcher is right then Kant thinks that the rational empirical cognition of an object as a body involves recognizing the concept of extension and the judgment that this object as a body as

necessarily belonging together, that is to say, as necessarily belonging to one and the same consciousness or self – and, what’s more, recognizing this on the basis that this concept and this judgment are representations. In this way, it involves making use of the rule – what Kitcher calls the I rule – that if something is a representation then it necessarily belongs together with other representations. The question is: Why can we not conclude on this basis that the cognition of the object as a body involves the cognition of our representations as necessarily belonging together in one consciousness? After all, there is a judgment of the right sort – namely, the judgment that our concepts and judgments necessarily belong together – and also a ground of cognition – namely, the application to these concepts and judgments of the concept of representation.

However, if Kant thinks that the cognition of the object as a body involves the cognition of our representations as necessarily belonging together in one consciousness then his theory is open to at least two serious objections. One is that it is inconsistent with Kant’s claim in the Paralogisms of Pure Reason that we cannot achieve any sort of cognition of the necessary unity of the self (cf. §3.3). Kant writes here that

[the] identity of the consciousness of myself in different times is... only a formal condition of my thoughts and their connection, but it does not prove at all the numerical identity of my subject, in which – despite the logical identity of the I – a change can go on that does not allow it to keep its identity... (A363)

Evidently, Kant does not think that we are able to achieve rational cognition of the identity of our self. Another objection to Kant’s theory, when understood in the way

that I have just indicated, is that it involves a vicious regress. For in order for us to recognize by means of the I rule that our concepts and judgments necessarily belong together, it is necessary for us to recognize that our application of the concept of necessarily belonging together and our application of the concept of being a representation also necessarily belong together. And this requires another application of the I rule in turn. However, the recognition that two states of our mind stand under the I rule requires yet another application of the I rule. And so on ad infinitum. I conclude that since it is hard to see how Kitcher can avoid ascribing to Kant a theory that is open to these objections, her account of the connection between a state of empirical cognition and the original synthetic unity of apperception should be rejected.

Conclusion

In this thesis I have defended an interpretation of Kant's Deduction of the Categories that differs in important respects from the more familiar interpretations in the secondary literature. If my interpretation is correct then many philosophers have failed to realize how key claims and concepts in the Deduction of the Categories should be understood. They have made false assumptions about the aim and the structure of Kant's argument. And they have also failed to notice aspects of Kant's theory of the faculty of understanding that are worthy of serious philosophical consideration.

My aim in this final section is to draw attention to some of the key conclusions defended in this thesis. In §1, I focused on Kant's aim in his Deduction of the Categories. In particular, I focused on the question what it means to demonstrate and explain the objective validity of the Categories. I argued that a concept counts as objectively valid in the sense that Kant has in mind insofar as it serves as a representation of an object. And I argued that a concept serves as a representation of an object in the sense that Kant has in mind insofar as two conditions are satisfied: first, there is a relation of agreement or conformity between the concept and the object, and second, it is no accident that there is this relation of agreement or conformity between the concept and the object. I used the various theories of the Categories that Kant criticizes in his famous letter of 21st February 1772 to Marcus Herz to illustrate this concept of representation.

I believe that this interpretation of Kant's concept of objective validity constitutes a major advance in our understanding of the Deduction of the Categories. Commentators do not always make it especially clear how they understand Kant's

concept of objective validity. Longuenesse, for example, thinks that a concept counts as objectively valid insofar as it is related to an object, but she doesn't tell us explicitly what this relation to an object is supposed to consist in (1998, p.17). Those commentators who do make it clear what they understand by the objective validity of a concept, such as Van Cleve (1999), tend to interpret it in terms of mere instantiation. But this is a serious mistake. If we interpret Kant's concept of objective validity in this way then we cannot understand how the various theories of the Categories that Kant mentions in the *Critique* and in his letter to Herz - including, crucially, the theory to which he himself subscribes - count as explanations of the objective validity of the Categories. In particular, we cannot understand why a demonstration that the Categories are required for experience of objects amounts to a demonstration that the Categories are objectively valid. To understand this we need to see the connection between Kant's concept of objective validity and his concept of representation. And we need to understand what Kant thinks is required for a concept to serve as a representation of an object.

The conclusions that I defended in §2 should be more familiar, even if they are not uncontroversial. I argued that Kant thinks that the conformity of an object to the Categories is required for this object to count as an object of experience because he thinks that the conformity of an object to the Categories is required for this object to count as an object of thought - that is to say, as an object of our faculty of understanding. I concluded on this basis that the strategy of Kant's Deduction of the Categories is to investigate the conditions that must be satisfied if we are to make use of our faculty of understanding, or think an object, in order to show that these conditions include the conformity of the object to the Categories. I also argued that what Kant

means by a case in which someone thinks of an object – that is to say, a case in which someone makes use of his faculty of understanding – is a case in which he makes a judgment about an object, say, the judgment that there are clouds over Oxford, or that England will win the forthcoming Ashes series.

Many commentators do not accept these conclusions. Some think that Kant's strategy in his Deduction of the Categories is to show that there is a connection between the Categories and the mere intuition of an object (e.g. Allison 2004), while others think that what Kant wants to do is to show that the Categories are required in some sense for consciousness in general. However, there is an increasing tendency for commentators to accept that Kant's strategy in his Deduction of the Categories is to investigate the conditions that must be satisfied in order for someone to achieve a state of empirical cognition – that is to say, a state of experience – where this state is understood in a fairly thick sense as involving the use of the faculty of understanding (e.g. Kitcher 2011). What is original in my discussion of Kant's concept of empirical cognition and the strategy of his Deduction of the Categories is the set of arguments that I use to defend my claims. If we want to understand Kant's Deduction of the Categories, it is crucial for us to recognize that the passages cited in support of the claim that Kant thinks that the Categories are required for the mere intuition of an object or for consciousness in general do not show any such thing. It is also crucial for us to recognize that the question whether or not a state of empirical cognition is factive has no bearing on the question of Kant's strategy in his Deduction of the Categories. It is also important that we should have a clear sense of the problems associated with the claim that what Kant means by empirical cognition is empirical knowledge, and that what he means by empirical cognition is empirical judgment on the basis of reasons.

In §3, I defended an interpretation of Kant's Principle of the Original Synthetic Unity of Apperception. I argued that this principle is a claim about the conditions that must be satisfied if we are to think of an object – that is to say, if we are to make use of our faculty of understanding. And I argued that this principle consists in the claim that if we are to make use of our faculty of understanding then those conditions must be satisfied that must be satisfied if we are to achieve a very special sort of consciousness. I then argued that the special consciousness in question is the consciousness – by means of our faculty of original apperception – of an act of synthesis that is necessary in the sense that we perform it whenever we connect any of our representations together. Finally, I argued that the consciousness of an act of synthesis that is necessary in this sense is what Kant understands by a Pure Concept of the Understanding.

Of special importance in §3.1 was my analysis of the Principle of the Original Synthetic Unity of Apperception in §16 of the second edition Deduction of the Categories. I believe that if we want to understand the structure of Kant's argument then it is crucial for us to recognize that Kant's claim at the start of this section is a claim about the conditions that must be satisfied if we are to think of an object – that is to say, the conditions that must be satisfied if we are to make use of our faculty of understanding. This is a point that has been overlooked by numerous commentators. As for §3.2, I have already mentioned that my interpretation of Kant's concept of apperception is similar to the one recently defended by Dickerson (2004). However, since Dickerson misinterprets Kant's concept of representation (cf. §1.4), he fails to see the implications of his interpretation for the crucial distinction between empirical apperception and original apperception. And I believe that once we have recognized

these implications we are in a far better position to make sense of Kant's complicated theory of the various operations of the mind that are involved in empirical cognition.

The claims that I defended in §4 are the most original conclusions in this thesis. Very few commentators on Kant's Deduction of the Categories have recognized that this argument turns on a claim to the effect that we must achieve a special consciousness – in particular, the consciousness of the necessary unity of synthesis – if we are to achieve a state of empirical cognition. And so very few commentators have even raised the question why Kant might think that this claim is true. In §4, I argued that Kant's commitment to the Principle of the Original Synthetic Unity of Apperception is based on his acceptance of two other claims. The first is the claim that when a person makes a judgment about an object – say, the judgment that there are clouds over Oxford – part of what she thinks is that a set of concepts ought to be connected in a particular way, which is to say, that a particular connection of concepts is in some sense required. The second is the claim that the thought of this requirement is made possible by consciousness of the necessary unity of all consciousness of connections between concepts. My argument for this interpretation of Kant's justification for the Principle of the Original Synthetic Unity of Apperception was based on detailed consideration of passages of the *Prolegomena* and the *Critique* that have been seriously misunderstood by commentators. If this interpretation was correct then Kant's argument for the Principle of the Original Synthetic Unity of Apperception – and, therefore, the theory of the objective validity of the Categories that he defends in his Deduction of the Categories – is based on a crucial idea about the distinctively normative character of an act of judgment that has escaped the notice of previous commentators. To recognize the place of this idea in Kant's argument is to take a

crucial step towards understanding the value of the Deduction of the Categories as a work of philosophy.

Bibliography

This thesis uses the conventional 'A/B' system for references to the first and second editions of the *Critique of Pure Reason* (e.g., A19/B33). References to Kant's other writings are to *The Berlin Academy Edition of the Works of Immanuel Kant* and take the form 'Ak' followed by volume and page number (e.g., Ak X:366). All translations are taken from the *Cambridge Edition of the Works of Immanuel Kant*. All references to 'the *Critique*' are references to Kant's *Critique of Pure Reason*. All references to 'the *Prolegomena*' are references to his *Prolegomena to any Future Metaphysics that will be able to come forward as Science*.

- Allais, L. (2009) *Kant, Non-conceptual Content and the Representation of Space*,
Journal of the History of Philosophy 47:3 (pp.383-413).
- Allison, H.E. (2004) *Kant's Transcendental Idealism: An Interpretation and Defense: Revised and Enlarged Edition*, Yale University Press, New Haven CT.
- Ameriks, K. (1978) *Kant's Transcendental Deduction as a Regressive Argument*, Kant-Studien 69:2 (pp.273-287).
- Ameriks, K. (1982) *Recent Work on Kant's Theoretical Philosophy*, American Philosophical Quarterly 19:1 (pp.1-24).
- Aquila, R.E. (1983) *Representational Mind: A Study of Kant's Theory of Knowledge*, Indiana University Press, Bloomington IN.
- Bain, A. (1872) *Mental and Moral Science*, Longmans, Green, and Co., London.
- Beck, L.W. (1978) *Essays on Kant and Hume*, Yale University Press, New Haven CT.

- Bennett, J. (1966) *Kant's Analytic*, Cambridge University Press, Cambridge.
- Bennett, J. (1971) *Locke, Berkeley, Hume: Central Themes*, Oxford University Press, Oxford.
- Bird, G. (2006) *The Revolutionary Kant: A Commentary on the Critique of Pure Reason*, Open Court Publishing Company, Chicago and LaSalle, IL.
- Buroker, J.V. (2006) *Kant's Critique of Pure Reason: An Introduction*, Cambridge University Press, Cambridge.
- Descartes, R. (1985) *The Philosophical Writings of Descartes: Volume I; Translated by John Cottingham, Robert Stoothof and Dugald Murdoch*, Cambridge University Press, Cambridge.
- Descartes, R. (1986) *Meditations on First Philosophy, With Selections from the Objections and Replies; Translated by John Cottingham with an Introduction by Bernard Williams*, Cambridge University Press, Cambridge.
- Dickerson, A.B. (2004) *Kant on Representation and Objectivity*, Cambridge University Press, Cambridge.
- Förster, E. (ed.) (1989) *Kant's Transcendental Deductions: The Three Critiques and the Opus postumum*, Stanford University Press, Stanford CA.
- Guyer, P. (1987) *Kant and the Claims of Knowledge*, Cambridge University Press, Cambridge.
- Guyer, P. *The Transcendental Deduction of the Categories* in Guyer, P. (1992) *The Cambridge Companion to Kant*, Cambridge University Press, Cambridge.
- Helmholtz, H. *The Facts of Perception* in Ewald, W. (ed.) (2007) *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, Oxford University Press, Oxford.

- Henrich, D. (1969) *The Proof-Structure of Kant's Transcendental Deduction*, *The Review of Metaphysics* 22:4 (pp.640-659).
- Henrich, D. (1994) *The Unity of Reason: Essays on Kant's Philosophy*, Edited with an Introduction by Richard Velkley, Harvard University Press, Cambridge MA.
- Kant, I. (1992) *Lectures on Logic*; Translated and Edited by J. Michael Young, Cambridge University Press, Cambridge.
- Kant, I. (1997) *Lectures on Metaphysics*; Translated and Edited by Karl Ameriks and Steve Naragon, Cambridge University Press, Cambridge.
- Kant, I. (1998) *Critique of Pure Reason*; Translated and Edited by Paul Guyer and Allen W. Wood, Cambridge University Press, Cambridge.
- Kant, I. (1999) *Correspondence*; Translated and Edited by Arnulf Zweig, Cambridge University Press, Cambridge.
- Kant, I. (2002) *Theoretical Philosophy After 1781*; Edited by Henry Allison and Peter Heath; Translated by Gary Hatfield, Michael Friedman, Henry Allison and Peter Heath, Cambridge University Press, Cambridge.
- Kant, I. (2005) *Notes and Fragments*; Edited by Paul Guyer, Translated by Curtis Bowman, Paul Guyer and Frederick Rauscher, Cambridge University Press, Cambridge.
- Kant, I. (2007) *Anthropology, History and Education*; Edited by Günter Zöller and Robert B. Loudon, Translated by Mary Gregor, Paul Guyer, Robert B. Loudon, Holly Wilson, Allen W. Wood, Günter Zöller and Arnulf Zweig, Cambridge University Press, Cambridge.
- Kemp Smith, N. (1923) *A Commentary to 'Kant's Critique of Pure Reason'*, Macmillan Press Ltd., Basingstoke.

- Kitcher, P. (1990) *Kant's Transcendental Psychology*, Oxford University Press, Oxford.
- Kitcher, P. (2011) *Kant's Thinker*, Oxford University Press, Oxford.
- Lamarra, A., Palaia, R. and Pimpinella, P. (2001) *Le Prime Traduzioni della Monadologie di Leibniz, 1720-1721; Introduzione Storico-Critica, Sinossi dei testi, Concordanze Contrastive*, L.S. Olschki, Firenze.
- Leibniz, G.W. (1898) *The Monadology And Other Philosophical Writings; Translated With Introduction And Notes By Robert Latta*, Clarendon Press, Oxford.
- Leibniz, G.W. (1996) *New Essays On Human Understanding; Translated And Edited By Peter Remnant And Jonathan Bennett*, Cambridge University Press, Cambridge.
- Longuenesse, B. (1998) *Kant and the Capacity to Judge*, Princeton University Press, Princeton NJ.
- McAlister, L.L. (1973) *Franz Brentano: Psychology from an Empirical Standpoint*, Routledge & Kegan Paul Ltd., London.
- Nidditch, P.H. (ed.) (1975) *David Hume: Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, Oxford University Press, Oxford.
- Nidditch, P.H. (ed.) (1975) *John Locke: An Essay Concerning Human Understanding*, Oxford University Press, Oxford.
- Paton, H.J. (1936) *Kant's Metaphysic of Experience: A Commentary on the First Half of Kant's Kritik der Reinen Vernunft*, George Allen & Unwin Ltd., London.
- Putnam, H. (1981) *Reason, Truth and History*, Cambridge University Press, Cambridge.
- Sassen, B. (2008) *Varieties of Subjective Judgments: Judgments of Perception*, Kant-

- Studien 99:3 (pp.269-284).
- Schultz, J. (1784) *Erläuterungen über des herrn Professor Kant Critik der Reinen Vernunft*, Carl Gottlob Dengel, Königsberg.
 - Smit, H. (2000) *Kant on Marks and the Immediacy of Intuition*, *Philosophical Review* 109:2 (pp.235-266).
 - Strawson, P.F. (1966) *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*, Methuen & Co. Ltd., London.
 - Stroud, B. (1977) *Hume*, Routledge and Kegan Paul Ltd., London.
 - Tetens, J.N. (1777) *Philosophische Versuche über die menschliche Natur und ihre Entwicklung*, M.G. Weidmanns Erben und Reich, Leipzig.
 - Thiel, U. (2011) *The Early Modern Subject: Self-Consciousness and Personal Identity from Descartes to Hume*, Oxford University Press, Oxford.
 - Van Cleve, J. (1999) *Problems from Kant*, Oxford University Press, Oxford.
 - Walker, R.C.S. (1978) *Kant*, Routledge & Kegan Paul Ltd., London.
 - Walker, R.C.S. (ed.) (1982) *Kant on Pure Reason*, Oxford University Press, Oxford.
 - Weldon, T.D. (1945) *Introduction to Kant's Critique of Pure Reason*, Oxford University Press, Oxford.
 - Wolff, C. (1732) *Psychologia Empirica Methodo Scientifica Pertractata*, Francofurti et Lipsae.
 - Wolff, C. (1747) *Vernünfftige Gedancken von Gott, der Welt, und der Seele des Menschen; Auch allen Dingen überhaupt, den Liebhabern der Wahrheit mitgetheilet*, Halle.
 - Wolff, R.P. (1963) *Kant's Theory of Mental Activity: A Commentary on the Transcendental Analytic of the Critique of Pure Reason*, Harvard University Press,

Cambridge MA.