USE OF THESES

This copy is supplied for purposes of private study and research only. Passages from the thesis may not be copied or closely paraphrased without the written consent of the author.
TRANSCENDENTAL ARGUMENTS
AND THE
CONCEPT OF SELF-REFUTATION

Thesis submitted for the degree
of
Doctor of Philosophy
in the
Australian National University

by

D.A. WHIWEWEL

September, 1971
In general, all parts of the thesis describe the candidate's own original work. Whenever the candidate makes use of ideas propounded by other philosophers due acknowledgement is made in the text. The candidate is of course indebted to his supervisors for their many helpful suggestions and criticisms.

Signed:

D.A. Whewell.

(D.A. Whewell).
The overall aim of this thesis is to find out whether any version of the transcendental method could be made to yield conclusions which are both factual and indubitably certain. In other words, could any sense be given to the notion of an a priori-synthetic principle, and if so, could such a principle be arrived at by means of a transcendental deduction? An affirmative answer is given to both these questions.

Part I. 'The search for a transcendental proof'.

I try to show that although Kant recognizes at least two kinds of transcendental argument - transcendental deductions and transcendental expositions - only transcendental deductions are capable of giving rise to principles which are, in some sense, a priori-synthetic.

I defend the concept of an a priori-synthetic principle against the objections of radical pragmatists (like Quine) on the one hand, and traditional empiricists (like Russell) on the other.

I criticize Kant's reasons for thinking that transcendental deductions, as he conceives them, logically guarantee the truth of their (synthetic) conclusions, and I explain how the structure of his arguments would have to be changed in order to achieve this.
Main conclusion: The initial premise of a transcendental deduction must be self-guaranteeing in the sense that any attempt to deny it would be self-refuting.

Part II. 'The concept of self-refutation'.

I examine a wide range of self-refuting statements, and produce a comprehensive system of classification. This is designed to show why some forms of self-refutation can be used as the basis of a transcendental deduction, but not others.

I point out that some of the things which have been said about self-refuting statements would seem to suggest that transcendental deductions based on any form of self-refutation are impossible, but that these suggestions are unfounded.

Main conclusion: A transcendental deduction may be described as an argument of the transcendental form in which the initial premise is operationally self-guaranteeing.

Part III. 'The composition of a transcendental deduction'.

I consider what other requirements an argument must satisfy in order to count as a transcendental deduction.

I argue that it should be based on a Cartesian-style premise and should be presuppositional rather than implicational in form. This leads to a comparison of the transcendental and Cartesian methods.

I explain that the only way in which ultimate principles or axioms can be justified, on other than pragmatic grounds, is
by some form of transcendental deduction.

Main conclusion: Our revised notion of a transcendental
deduction has much in common with Kant's original conception,
although his claim that this method of argument is 'entirely
different from any hitherto conceived' is not altogether
justified.
# CONTENTS

<table>
<thead>
<tr>
<th>Introduction</th>
<th>PART I - In Search of a Transcendental Proof</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Section 1. The concept of a transcendental argument</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Section 2. Apodeictic certainty</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Section 3. Quine's attack on the doctrine of analyticity</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Section 4. The concept of an a priori-synthetic principle</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Section 5. The initial premise of a transcendental deduction</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Section 6. The principle of the possibility of experience</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Section 7. The structure of a transcendental deduction</td>
<td>68</td>
</tr>
</tbody>
</table>

| PART II - The Concept of Self-Refutation | Section 1. The introduction | 79   |
|                                          | Section 2. The various forms of self-reference | 79   |
|                                          | Section 3. Meaning and self-reference | 80   |
|                                          | Section 4. Empirical self-referring statements | 88   |
|                                          | Section 5. Passmore's classification | 100  |
|                                          | Section 6. Mackie's classification | 105  |
|                                          | Section 7. Bonney on weak prefixability | 107  |
|                                          | Section 8. Possible ways of classifying 'I have no beliefs' | 112  |
|                                          | Section 9. 'I know that' | 116  |
|                                          | Section 10. Self-cancelling utterances | 120  |
|                                          | Section 11. Operationally self-refuting statements | 121  |
|                                          | Section 12. Two forms of operational self-refutation | 123  |
PART III - The Composition of a Transcendental Deduction

Section 1. The Introduction
Section 2. 'I do not think' as a self-refuting statement
Section 3. The concept of self-evidence
Section 4. Some self-refuting arguments
Section 5. The cogito principle as a transcendental axiom
Section 6. Referentially self-guaranteeing statements
Section 7. The conclusion

Bibliography
INTRODUCTION

Vanity, according to A.J. Ayer, is 'the sine qua non of philosophers'. The reason he gives for this seemingly extravagant claim is that in philosophy there are no established criteria of truth and falsehood - as he presumes there are in the natural sciences - so that its practitioners are extremely reluctant to admit error.

Whether philosophers are as reluctant to admit error as Ayer claims is at least debatable, but the charge is by no means a new one. Immanuel Kant, for example, accused metaphysicians of being, in general, arrogant in their intellectual aspirations and dogmatic in their approach. He saw this as being due to their failure to agree on an objective procedure for settling philosophical disputes. In other words, he realized that in metaphysics there was, as yet, no commonly accepted criteria of truth and falsehood:

1Ved Mehta, Fly and the Fly-Bottle, p.75. 2e.g. Critique of Pure Reason, Axi and Axiiv.

*For Kant, a metaphysician is simply a philosopher who is not a natural philosopher (i.e. one who employs empirical methods).
So far are the students of metaphysics from exhibiting any kind of unanimity in their contentions, that metaphysics has rather to be regarded as a battle-ground quite peculiarly suited for those who desire to exercise themselves in mock-combats, and in which no participant has ever yet succeeded in gaining ever so much as an inch of territory, not at least in such manner as to secure him in its permanent possession.¹

If metaphysics is to regain the respect in which it was formerly held it must, according to Kant, be set 'upon the secure path of a science'² by the introduction of a new philosophical method. This method must achieve two things: firstly, it must guarantee the validity of its conclusions in such a way as to prevent them from being seriously questioned. And secondly, it must define the limits of metaphysical knowledge, either by providing the key to the solution of all metaphysical problems or by showing them to be insoluble. Failure to define these limits would leave some of the problems untouched and so enable the dogmatists to continue to engage in their 'mock-combats'.

In the transcendental method Kant believes that he has found the only way of achieving these aims. He is aware that such pretensions may seem 'arrogant and vainglorious',³ but argues

¹ Bxv. ² Bvii. ³ Axiv.
that, in practice, they are incomparably more moderate than the claims of all those writers who, on the lines of the usual programme, profess to prove the simple nature of the soul or the necessity of a first beginning of the world. For while such writers pledge themselves to extend human knowledge beyond all limits of possible experience, I humbly confess that this is entirely beyond my power.¹

In the Critique of Pure Reason the task of setting metaphysics upon the sure path of a science is identified with the task of defining the limits of philosophical argument. Such an approach is only feasible, however, if one already knows what a philosophical argument is. Yet any attempt to define this concept is bound to be contentious, for there is no general consensus of opinion amongst philosophers as to what form their arguments should take. Indeed, it is not even clear that there should be a distinctively philosophical mode of argument.

Kant, however, did not recognize this difficulty. 'Philosophy' he confidently declared 'is simply what reason knows by means of concepts';² that is to say, a philosophical argument is one which proceeds by concepts alone without the aid of sensory experience. To discover how much can be known

¹Axiv. ²A732/B760.
in this way is therefore, on this view, to reveal the scope of the whole metaphysical enterprise. 'the chief question' said Kant 'is always simply this:- what and how much can the understanding and reason know apart from all experience?'

In short, the Critique of Pure Reason can be regarded as an attempt to define the philosopher's task by determining the scope of a particular mode of enquiry. This approach might be justified on the grounds that it is by its methods rather than by its subject-matter that philosophy is distinguished from other arts or sciences. This is of course a familiar enough notion and one which is still widely, though by no means universally, accepted. It is implicit in the often repeated claim that philosophy 'is not a body of doctrine but an activity'. However, among the philosophers who believe this to be true there are differing views as to what kind of an activity it is.

For Kant, it was the activity of reasoning a priori from concepts. Hence his claim that 'philosophy can never be learned, save only in historical fashion; as regards what concerns reason, we can at most learn to philosophize'. Wittgenstein came to regard it as the art of destroying linguistically induced nonsense by bringing 'words back

\[1\] Axvii. \[2\] Wittgenstein, Tractatus Logico-Philosophicus, \[4.112. \[3\] A837/B865.
from their metaphysical to their everyday usage.¹ For John Wisdom it was, less specifically, the art of analyzing concepts: 'analytic philosophy has no special subject-matter. You can philosophize about Tuesday, the pound sterling, and lozenges and philosophy itself'.²

These methods are all broadly a priori in character, in the sense that they do not require access to a store of esoteric empirical information, but differ very much in scope. Wittgenstein, for example, seemed to think that the philosopher could do no more than expose the fallacies in metaphysical thought through an examination of the way words are actually used. Wisdom, however, maintained that such an examination could perform a wider and more positive role in that it could give us new insights into the nature of our classificatory system. Even metaphysical theories could be useful in this respect by drawing attention to distinctions which ordinary language tends to conceal. For instance, the theory that no empirical propositions can be certain, whatever its defects, does at least draw attention to the fact that empirical propositions are validated in a quite different way from, say, mathematical propositions, and that the point at which they are validated is much more difficult to define.

²Wisdom, Problems of Mind and Matter, p.2.
However, Wisdom agreed with Wittgenstein that it was no part of the philosopher's task to add to our information about the world. 'The analytic philosopher, unlike the scientist, is not one who learns new truths but one who gains new insight into old truths'. In other words, his task consists in making explicit knowledge which is already implied by the concepts we use. In fulfilling this task, we are therefore only reminding ourselves of facts which are already within our grasp, but whose logical significance we have perhaps overlooked.

This account is open to a number of serious objections, not least of which is that, in the end, our concepts are merely ways of organizing experience, and therefore cannot be adequately analyzed without reference to that experience. In other words, we must look not merely at the concepts themselves, but at those elements in our experience to which they are supposed to apply, and this, being a straightforwardly empirical procedure, may easily result in our noticing new facts. For instance, if we attempt to analyze the concept of memory without examining actual memory situations we may easily be influenced by ideas which are both empirically false and philosophically dangerous; e.g. the idea that the memory is a form of perception and that the only

---

1 Wisdom, Problems of Mind and Matter, p.2.
way in which we can recall an event is by conjuring up an image of it in our minds. In refuting such an idea we may well be adding to our information about the world.

Thus, what is termed the 'analysis of ordinary concepts' is in part no different from the analysis of ordinary (i.e. everyday) experience. The only justification for describing this as an a priori activity is that to be successful at it we need only reflect upon the kind of experiences with which we are already perfectly familiar; that is to say, we are not obliged to extend our experience of the world in the way that the natural scientist is.

Nevertheless, under Kant's criteria, this could certainly not be regarded as a legitimate philosophical technique. For it cannot be described as a priori in the tough sense of being wholly non-empirical, nor can it be said to lead to conclusions which are indubitably certain. This is not to say that all forms of conceptual analysis are unphilosophical in the same sense,* but only those which rely on, or give rise to, empirically based statements of fact. It remains to be seen...

*Kant does occasionally seem to suggest that only the analysis of a priori concepts can give rise to a priori (i.e. indubitable) knowledge, and that the analysis of empirical concepts can only give rise to empirical knowledge. 'For its (metaphysics) business is not merely to analyse concepts which we make for ourselves a priori of things ...'. B18.
whether or not they must all do so. Moreover, Kant may be right in claiming that not all statements of fact are empirically based. And if some of these statements can be proved independently of experience they could presumably be used as the basis of a more rigorous form of conceptual analysis.

Nevertheless, conceptual analysis in general is neither as rigorous as it was intended to be, nor does it appear to have finally solved any of the traditional problems of philosophy. This is evidenced by the fact that A.J. Ayer can still claim that philosophers are peculiarly prone to dogmatism by virtue of the fact that in philosophy (except where questions of formal logic are involved) there are no 'established criteria of truth and falsehood'.

Failure by successive generations of philosophers to remedy this state of affairs has caused an increasing number of philosophers to lose faith in the idea that philosophy could be transformed into an exact science if only the right methodology were found. They have also tended to lose faith in the idea that there is, or even should be, a distinctively philosophical mode of argument. Accordingly they are prepared to argue in any way that seems appropriate to the subject-matter, and see no reason for committing themselves in advance to a

---

\[1\] Ved Mehta, *Fly and the Fly-Bottle*, p.75.
specific type of methodology. And since the subject-matter itself is no longer determined by preconceived ideas of what may count as a philosophical argument they have tended to broaden rather than reduce the scope of their enquiries.

In the end, this relatively informal approach may prove more rewarding than those earlier attempts to transform philosophy into a highly specialized and disciplined study. Nevertheless, it has yet to be shown that those attempts were wrong in principle. In other words, no overwhelming reason has yet been found for saying that inconclusiveness is the inevitable fate of all distinctly philosophical (i.e. non-mathematical and non-empirical) modes of argument. On the other hand, seemingly powerful arguments have been adduced for saying that Kant's philosophical method, at least, is a non-starter, and that a priori reasoning on matters of fact cannot conceivably lead to conclusions which are absolutely indubitable. The purpose of this thesis is to argue against such a position; to show that a method of argument markedly similar to Kant's can, in some sense, be made to yield conclusions which are both indubitable and factual.
In Search of a Transcendental Proof

Section 1. The concept of a transcendental argument.

Kant makes at least four major claims on behalf of the philosophical method that he employs. These are: (1) that the discoveries made by this method are apodeictically certain without being analytic.\(^1\) (2) that it provides the key to the solution of all metaphysical problems.\(^2\) (3) that it provides the only access to metaphysical truth which is possible for perceptually-dependent beings like ourselves.\(^3\) (4) that this method is 'entirely different from any hitherto employed'.\(^4\)

It would be generally admitted, I think, by even the most loyal Kantians, that post-Kantian developments in physics, geometry, logic and philosophy, as well as more recent studies of the *Critique of Pure Reason* itself, have shown that none of these claims (with the possible exception of the last) is justified by any argument put forward in the *Critique*; that it is extremely doubtful whether claims '2' and '3' could be legitimately made on behalf of any kind of philosophical method whatever;\(^5\) that, for the most part, the conclusions which Kant arrives at by means of this method, far from being apodeictically

\(^1\) e.g. A xv. \(^2\) e.g. A xiii. \(^3\) e.g. A 7 8 7/ B 8 1 5. \(^4\) e.g. B 2 4. 
certain, are obviously false; and that the actual arguments, which are supposed to establish these conclusions, go wrong in not just one or two ways, but in a tediously large number of different ways. In other words, whatever the merits of the transcendental method, Kant's own particular application of it is clearly unsound.

However, I do not intend to embark on a detailed exposition of the particular theories and arguments advanced by Kant in the Critique in order to explain where and why they break down. My primary object is rather to determine whether any version of the transcendental method could be made to yield conclusions which were both factual and apodeictically certain. The question of how far such an argument would satisfy Kant's other requirements will also be commented upon. We will begin, however, by discussing the general nature of a transcendental argument.

The term 'transcendental' could be, and has been, applied to many different kinds of argument - most of them invalid. Kant himself appears to recognize at least two different types. In the Critique these are referred to as 'transcendental deductions' and 'transcendental expositions', although Kant is not consistent in his use of these terms. In the Prolegomena, on the other hand, they are referred to as the 'synthetic' and 'analytic' methods. However, the distinction
appears to be the same in each case. Examples of the transcendental deduction are mainly to be found in those sections of the Critique entitled 'the Transcendental Deduction' and 'the Analytic of Principles' whilst examples of the transcendental exposition are mainly to be found in the 'Aesthetic'¹ and in the first part of the Prologomena.²

The difference is very roughly as follows: according to Kant, a transcendental deduction is supposed to show that some concept or other is a necessary condition of the possibility of experience, and that this concept gives rise to a principle which is not only factually true but apodeictically certain: in other words, an a priori-synthetic principle. The main purpose of such an argument, he thinks, is to justify our use of certain a priori concepts and their associated principles which cannot be justified on purely empirical grounds. A transcendental exposition, on the other hand, is supposed to show that some concept or other is a necessary condition of the possibility of a particular kind of a priori-synthetic statement. Thus, at B40, he says:

I understand by a transcendental exposition the explanation of a concept, as a principle from which the possibility of other a priori-synthetic knowledge can be understood. For this purpose it is required (1) that such knowledge does really flow

¹ e.g. B40 & B48. ² e.g. Prologomena, sections 6 to 11, & pp.xv to xvi in Lucas' edition.
from the given concept. (2) that this knowledge is possible only on the assumption of a given mode of explaining the concept.

By means of this type of argument Kant tries to prove in the Aesthetic that space and time are a priori intuitions. For example, in the section entitled 'the transcendental exposition of the concept of space' he argues that the a priori-synthetic (sic) character of Euclidean geometry can only be explained on the assumption that space is an a priori intuition, and in the section entitled 'the transcendental exposition of the concept of time' he argues that the a priori-synthetic (sic) character of the mathematical theory of motion can only be explained on the assumption that time is an a priori intuition.

In fact, the distinction between transcendental deductions and transcendental expositions is not as clear as the above account might seem to suggest since, as it stands, Kant's definition of a transcendental exposition ('the explanation of a concept, as a principle from which the possibility of other a priori-synthetic knowledge can be understood') would seem to apply equally well to his transcendental deductions. For instance, Kant's transcendental deduction of the categories is supposed to show how an
a priori science of nature is possible,”¹ just as his transcendental exposition of the concept of space is supposed to show how an a priori science of physical space is possible.² Thus the transcendental deduction of the categories could easily be represented as the explanation of the concept of experience in general, as a principle from which the possibility of an a priori science of nature could be understood. In other words, it could, on Kant's definition, be regarded as a transcendental exposition.

Conversely, Kant's definition of a transcendental deduction at B177 ('the explanation of the manner in which concepts relate a priori to objects I entitle their transcendental deduction') seems to apply equally well to his transcendental expositions of space and time. For they are supposed to explain how the a priori character of space and time can be inferred from Euclidean geometry and the mathematical theory of motion. And this is taken to mean that the objects of our senses must conform to the constitution of our faculty of intuition.

This interpretation is strengthened by the fact that at B118 & B119 Kant argues that not only is a transcendental

¹B17, B164-165. ²B40-41.

*i.e. by showing that the fundamental concepts of Newtonian physics, e.g. causality, substance etc. are necessary conditions of the possibility of experience, and therefore a priori.
deduction of space and time necessary, but that it has already been performed (namely, by the metaphysical*¹ and transcendental expositions²): 'If, therefore, a deduction of such concepts (space and time) is indispensable, it must in any case be transcendental'. And later 'we have already, by means of a transcendental deduction, traced the concepts of space and time to their sources, and have explained and determined their a priori objective validity'.

However, the fact that transcendental deductions and transcendental expositions cannot be distinguished on the basis of Kant's definitions alone does not of course mean that they cannot be distinguished at all. For although they share a common form, their starting-points are quite different.

A transcendental argument as such may be described (but not necessarily defined) as an argument of the form: 'x is a necessary condition of the possibility of y. y is in fact possible. Therefore x is in fact the case'. I shall term this 'the transcendental form'. The basic difference, then,

*The purpose of a metaphysical exposition is to analyze an a priori concept. It is therefore a first step towards a transcendental deduction. ('By exposition ... I mean the clear, though by no means exhaustive, representation of that which belongs to a concept: the exposition is metaphysical when it contains that which exhibits the concept as given a priori'), B38.

¹B37 to B41, & B46 to B49. ²B41 to B46, & B49 to B54.
between transcendental deductions and transcendental expositions lies in the range of things for which the variable 'y' can stand. Thus, in the case of a deduction, it is always 'experience' or 'conceptualized experience', whereas in the case of an exposition it is a body of a priori-synthetic knowledge. In both cases, the conclusions are supposed to be not only factual but apodeictically certain; a claim we will be examining in some detail later on.

The following highly schematized argument would be an example of a transcendental deduction: The assumption that every event has a cause is a necessary condition of the possibility of experience. Experience is in fact possible (because it exists). Therefore it is certain that every event has a cause. The purpose of such an argument is to demonstrate the objective validity of an allegedly non-empirical concept (e.g. causality), and to show that its employment as a universal principle in Newtonian physics is perfectly justified, even though a fundamental principle of this kind cannot be justified on empirical grounds.

*The transcendental deduction of all a priori concepts has thus a principle according to which the whole enquiry must be directed, namely, that they must be recognized as a priori conditions of the possibility of experience, whether of the intuition which is to be met with in it or of the thought' (viz. conceptualized experience). A94/B126.
The following highly schematized argument would be an example of a transcendental exposition: That space is an a priori intuition is a necessary condition of the possibility of a geometry based on a priori-synthetic principles. A geometry whose axioms are a priori-synthetic is in fact possible (because this is true of Euclidean geometry). Therefore the fact that space is an a priori intuition is no less certain than those axioms; that is: to say, it is completely certain.

Such an argument has a dual purpose; namely (1) of explaining how a priori-synthetic judgements in Euclidean geometry are possible, and (2) of proving that space is a necessary element in the structure of our experience, or as Kant puts it, of explaining its 'a priori objective validity'.

If, however, we merely wish to distinguish the two types of argument, we may define a transcendental exposition quite simply as an attempt to explain how a particular type of a priori-synthetic knowledge is possible by tracing it back to its source in an a priori intuition. And we may define a transcendental deduction quite simply as an attempt to justify our use of certain non-empirical concepts and principles by showing them to be necessary conditions of the possibility of experience. The question now arises of whether modes of

\[8120.\]
argument so defined can ever be valid.

Clearly the fundamental weakness of any transcendental exposition must lie in its assumption that a body of a priori-synthetic judgements already exists, and in particular, that the axioms of Euclidean geometry (e.g. that the shortest distance between two points is a straight line)\(^1\) and the principles behind the mathematical theory of motion (e.g. that in the operation of one body on another, reaction is always equal and opposite to action) are a priori-synthetic.

As far as geometry is concerned, Kant believes that the axioms of Euclid can be known to be true of the world without recourse to experience, not through concepts alone, as is the case with other a priori-synthetic principles,\(^2\) but through the construction of concepts in a priori intuitions.\(^3\) His sole argument in support of this claim seems to be that we cannot represent to ourselves in intuition a geometrical figure which does not accord with these axioms.\(^4\) The import of this argument is that if we cannot visualize a thing it cannot exist (at least in so far as the phenomenal world is concerned). But since this could easily be due to the weakness of our powers of imagination and perception, it is a dangerous assumption to make.

---

\(^1\)B16. \(^2\)A731/B764. \(^3\)A731/B764. \(^4\)B16 & B299.
Moreover, post-Kantian developments in geometry and physics have shown that a non-Euclidean universe is perfectly conceivable, even if it is not visualizable. Indeed, it might be claimed that, of all possible systems of geometry, the Euclidean system is not necessarily the most useful for the description of the physical world. But if, after all, geometrical statements are not a priori-synthetic, they do not need to have their possibility explained by means of a transcendental exposition under Kant's definition of the term.

Similar objections can be made against Kant's only other attempt at a transcendental exposition in the *Aesthetic*, namely, that concerned with the concept of time at B48. In this argument he claims that the 'body of a priori-synthetic knowledge which is exhibited in the general doctrine of motion' can only be understood on the assumption that time is an a priori intuition.

The only support for this claim, apparently, is that 'the concept of motion, as alteration of place, is possible only through and in the representation of time'. But how this fact is supposed to make possible an a priori science of motion is far from clear. In any case, it is obvious from

---

2. B49.  
the development of quantum mechanics that the laws of motion are not a priori-synthetic. Rather, they are like the axioms of geometry; neither necessarily nor universally true of the physical world. For instance, one of the fundamental axioms of classical mechanics says that in the operation of one body on another, reaction is always equal and opposite to action. This axiom does not, however, appear to apply to the behaviour of atomic particles.

This point can be generalized in the following way: it is sometimes possible for an apparently viable scientific system, based on one set of axioms, to be either wholly or partly replaced by another equally viable system, based on a quite different set of axioms. And therefore Kant is not entitled to claim that some scientific laws are necessarily true of the world simply because they are the axioms of one particular system.

This means that if Kant is to explain how a body of a priori-synthetic knowledge is possible by using a transcendental exposition, he must somehow prove that such knowledge does in fact exist. In the transcendental expositions of space and time, however, he fails to do this.

On the other hand, he does try to prove by means of a transcendental deduction that the fundamental principles of Newtonian physics are a priori-synthetic. But having established
(let us suppose) the existence of an a priori science of nature, he is not then obliged to explain its possibility by means of a transcendental exposition.

This is because in order to understand how a priori-synthetic principles in the natural sciences are possible, we need only see the significance of what the transcendental deduction has already taught us; namely, that our experience is necessarily organized by certain a priori concepts, and that the principles according to which this organization takes place are a priori-synthetic. For, if they exist, such principles must, of necessity, be taken for granted in every type of empirical knowledge. If, on the other hand, the world did not conform to those principles any experience of it would be impossible, and therefore empirical science itself would be impossible. In short, the task of proving that certain scientific principles are a priori-synthetic involves showing how it is possible for them to be so. In other words, there is no room for a transcendental exposition (as opposed to a transcendental deduction) here either.

It would seem then that transcendental expositions, as conceived by Kant, are either fallacious or else indistinguishable from transcendental deductions. On the

*These alternatives are not of course mutually exclusive.  
1 e.g. B124-127, A96-97, B150, B166.  
2 e.g. B165, B218ff.
other hand, a more acceptable type of argument emerges once the more obvious defects of the expositional form are removed. These are: (1) the unproved assumption that a body of a priori-synthetic knowledge already exists. And (2) its links, via the concept of an a priori intuition, with the now almost wholly discredited doctrine of mind-made nature; viz. that the a priori intuitions of space and time are imposed by the intuited subject on the material which is presented to him in sensation.

If we omit these objectionable elements from our concept of a transcendental exposition, we are left with the bare idea of an argument which attempts to show that a particular concept or principle is necessary to the employment of some body of knowledge. Such an argument is at least feasible, and indeed comes fairly close to what R.G. Collingwood saw as the main purpose of metaphysics, namely to bring to light the absolute presuppositions of human thought at some date in history.¹

The description of a similar type of transcendental argument has been given recently by A. Phillips Griffiths in a symposium on the subject:² 'I characterize a transcendental argument as one to the conclusion that the

truth of some principle is necessary to the possibility of the successful employment of a specified sphere of discourse'.

On this view, transcendental arguments have a wider role than that suggested by Collingwood. For the expression 'sphere of discourse' is meant to include ever type of speech-act (e.g. evaluations, commands, mathematical calculations etc.) and not just the knowledge claims of particular sciences (e.g. phrenology or Newtonian physics).

Both Collingwood and Phillips Griffiths would agree, however, that the principles established by this type of argument are a priori only in relation to a given sphere of discourse, whose legitimacy is in no way established by the argument, and are not absolutely a priori. They cannot therefore be regarded as apodeictically certain, and may eventually be revised in the light of experience, as was the Newtonian principle of universal causality.

Moreover, in some cases the sphere of discourse may be rejected altogether. For instance, 'one might reject phrenology on the grounds that if anything were to count as conditions under which true statements about people's personalities could be made, based entirely on the bumps on their heads, a principle would be required which is in fact refuted by the more general evidence of physiology'.

---

Looked at in this way, a transcendental argument can, at the most*, only
show the necessity of accepting the principle on the part of anyone who claims seriously to employ the relevant sphere of discourse, or of abandoning such a claim. It is not then suggested that in a transcendental conclusion the truth of the principle is established.

This is, of course, far removed from Kant's original concept of a transcendental argument, although clearly a development of it. We are perhaps justified in using the term in this revised sense in order to remind ourselves of those links. We would not, however, be justified in describing the argument as a transcendental deduction, for the term 'deduction' has a legal or juristic connotation,¹ which is not applicable in this case.

Thus, according to Kant, a deduction is concerned with the vindication of a right (quid juris), rather than with establishing a fact (quid facti),² namely our right to make use of certain non-empirically derived concepts and their associated principles, which might otherwise be regarded as suspect. His transcendental deduction of the categories, for

*He later suggests that not even this can be achieved. A similar view is put forward by S. Korner in the Monist, July 1967. One of my chief aims in this thesis is to show how the difficulties envisaged here can be overcome.

¹A84/B116. ²Ibid.
instance, is supposed to establish our right to take for
 granted the validity of certain fundamental concepts in
Newtonian physics.

The kind of argument described by Phillips Griffiths,
on the other hand, is merely intended to establish the
hypothetical truth that if we are to employ a particular
sphere of discourse we are forced to take the validity of
a particular concept or principle for granted. But it does
not establish our right to do so. It is not therefore a
deduction.

If it is to be regarded as a transcendental argument
at all, it could more reasonably described as a
transcendental exposition, since, as we have seen, it has
several features in common with Kant's original notion; for
instance, in taking as its starting-point the existence of
a particular body of knowledge.

The declared aim of this thesis is to determine whether
or not any version of the transcendental method could be made
to yield conclusions which are not only factually true, but
apodeictically certain. It is now clear that the transcendental
exposition, whether widely or narrowly defined, does not come
into this category. It remains to be seen whether the
transcendental deduction fares any better.

If we accept some of the current definitions of a
transcendental deduction, the prospects do not appear too
bright. S.Korner, for instance, claims that a transcendental deduction ... can be defined quite generally as a logically sound demonstration of the reasons why a particular categorial schema is not only in fact, but also necessarily employed, in differentiating a region of experience.

He then proceeds to adduce fairly compelling reasons why an argument of this kind is impossible.

It would, for example, be necessary to show that every possible method of differentiating a region of experience belonged to a single schema. But we can only demonstrate the uniqueness of a particular schema on the basis of internal evidence, for we cannot compare it with other rival schemas, as, ex hypothesi, there are none, nor can we compare it with undifferentiated experience, since we would have to employ statements which presupposed some method of differentiation in order to effect the comparison. But it is surely impossible to demonstrate the uniqueness of a particular schema on internal evidence alone. As Korner says 'such an examination, at best, could only show how the schema functions in the differentiation of a region of experience, not that it is the only possible schema to which every differentiation of the region must belong'.

The chief difficulty with this argument lies in the fact that one cannot even refer to a region of experience.

without using a method of differentiation. Thus, to admit the possibility of a region of experience being differentiated by two totally dissimilar methods of differentiation is surely to imply that a region of experience can be distinguished independently of any particular method of differentiation. But this is clearly impossible. It could therefore be argued that if the method of differentiation is totally different (in the sense of exhibiting different schemas) then ipso facto the region of experience is different.

However, even if Korner is right, and he probably is, it only means that transcendental deductions, as he conceives them, are impossible, and not that they are impossible on any interpretation. In other words, although Kant's conception of a transcendental deduction is clearly covered by Korner's definition, other definitions are perhaps possible which would satisfy many of his criteria, but which would not be subject to the same objections.

For instance, under a new definition, it might be possible to construct a logically sound deduction which did not involve the requirement that a particular categorial schema was the only possible one for differentiating a region of experience. Kant himself goes some way towards admitting the possibility of this through his doctrine of transcendental
idealism; namely, that there is an unbridgeable gulf between things as they appear to us and things as they really are in themselves. For this allows for the possibility of modes of intuition and forms of categorization which have nothing at all in common with our own, but which, because of our physical limitations, we are prevented from ever using, or even knowing about: 'We cannot judge in regard to the intuitions of other thinking beings, whether they are bound by the same conditions as those which limit our intuitions and which for us are universally valid',¹ and again, 'we are acquainted with no kind of intuition but our own sensible kind and no kind of concepts but the categories, and neither of them is appropriate to a non-sensible object'.² This shows, I think, that even Kant did not regard his categorial schema as unique in any absolute sense. Nevertheless, this is not a very promising line of enquiry, since the doctrine of transcendental idealism is the least satisfactory part of the critical philosophy, and in any case, for our purposes, it does not go far enough.

According to Kornr, although it is not possible to demonstrate the uniqueness of any schema of differentiation, it is possible to show that a particular method of differentiation is a priori in relation to a given schema.³

¹B43/A27. ²B343/A287. ³The Monist, July 1967, p.327ff.
This, he believes, gives rise to a kind of metaphysical exposition.¹ He then argues that since a metaphysical exposition in this revised sense is both logically sound and philosophically useful, it should replace the somewhat dubious concept of a transcendental deduction. A metaphysical exposition, however, is not intended to give rise to conclusions which are both factual and indubitable, as Korner would be the first to admit, therefore it is not the kind of argument we are looking for.

Our best hope of finding such an argument still lies with some suitably revised version of a transcendental deduction. The problem of arriving at such a concept can be tackled in three distinct stages: firstly, we must list the minimal conditions which an argument must satisfy if it is to count as a transcendental deduction. Secondly, we must discover whether these conditions can in principle be satisfied by any argument. And thirdly, if there is such an argument, we must try to determine how close it actually comes to Kant's original conception.

The conditions which any argument must satisfy if it is to count as a transcendental deduction can be listed fairly briefly. They are (1) it must be an argument of the transcendental form; that is to say, it must be an argument

¹c.f. B38, B80.
of the form 'x is a necessary condition of the possibility of y. y is in fact possible. Therefore x is in fact the case'. Or, more briefly, 'x is the case because it is presupposed by y'. (2) The conclusion must in some sense be apodeictically certain or indubitable. (3) It must proceed in strictly a priori fashion without relying on any kind of empirical evidence. (4) In order to be termed a deduction, it must succeed in establishing our right to take certain facts, concepts, or principles for granted. In the course of this thesis I shall try to show how the many difficulties to which these conditions give rise can be overcome.

One obvious difficulty is that if we argue in a priori fashion through a series of logically connected premises, the conclusion can be no more certain than the premises on which it is based. Thus the requirement that the conclusion of a transcendental deduction be both factually true and apodeictically certain, or indubitable, can only be satisfied if there is some way of showing that the major and minor premises are themselves indubitable. But in the case of the minor premise this raises a very special problem, for it is itself a statement of fact.

Being an initial premise, any attempt to demonstrate its indubitability must carefully avoid presupposing other statements
of fact, for these would in their turn have to be proved. It must therefore be shown to be self-evident, although not in the usual empirical sense for this would be inconsistent with the requirement that the argument as a whole be strictly a priori. In other words, what is required at this stage is a type of presuppositionless argument; a condition which many philosophers would argue was impossible.

These then are the minimal conditions which any argument must satisfy if it is to be described as a transcendental deduction. If we find that an argument which satisfies these conditions is possible, we can then go on to determine how close it actually comes to Kant's original conception, and, in particular, whether it lends support to any of the major claims which Kant makes on behalf of the transcendental method; namely, that it provides a way of solving some of the traditional problems of philosophy; that this is the only way in which such problems could be solved; and that the method is distinctively philosophical, but different from all other forms of philosophical argument.

Our main task, however, is to determine whether there is any way of guaranteeing the truth of a statement of fact other than by appealing to experience, for this, ultimately, is what the transcendental method is all about.

1 See p. 10 of this thesis.
Section 2. Apodeictic certainty.

Whether or not a statement of the form 'it is apodeictically certain that $x$ is in fact the case' could, under the right conditions, be validly inferred from a statement of the form 'x is a necessary condition of the possibility of y, and y is in fact possible' depends not only on whether we are justified in taking the possibility of y for granted, but also on our interpretation of the expression 'apodeictically certain'.

Whatever interpretation we decide to adopt must at least be consistent with Kant's own account if we are to succeed in demonstrating the possibility of a genuine transcendental deduction. Let us therefore begin by taking a look at that account.

By far the most important fact about Kant's use of the term is that he uses it to refer to a standard of certainty which is, in some sense, absolute or ultimate. Thus, at Axxv in the first edition preface, for example, he says:

Any knowledge that professes to hold a priori lays claim to be regarded as absolutely necessary. This applies still more to any determination of all pure a priori knowledge, since such determination has to serve as the measure, and therefore as the supreme example of all apodeictic (philosophical) certainty.

In ordinary speech, to say of a statement that it is certain is to say, in effect, that there are excellent reasons for believing it and no good reasons for denying it. In this sense
a great many empirical truths are certain; e.g. that Paris is the capital of France, that Charles I of England was executed, that smoking is a major cause of lung cancer. It would be generally admitted, I think, that although by conventional standards these statements are all certain, they are not equally certain. For instance, it is more certain that Paris is the capital of France than that Charles I was executed, and more certain that Charles I was executed than that smoking is a major cause of lung cancer. As the evidence varies in each case so does the degree of certainty. Thus when the evidence is stronger and the reasons for doubting it harder to find, the statement can be said to satisfy a higher standard of certainty.

The idea that statements may enjoy varying degrees of certainty, although sanctioned by ordinary language, does not rule out, but rather lends support to, the notion of an ultimate standard of certainty; that is to say, a standard of certainty at which all rational doubt becomes impossible. At this point, the reasons for believing the statement are so good as to be infallible.

Amongst philosophers, the traditionally held view is that no empirically based reason can be infallible, since it is always possible to invent a story, however far-fetched which would explain away evidence of this kind; for example, as
a last resort, one could advance the theory that all
sense-perception is illusory. In other words, it is always
conceivable for a statement of this kind to be false.

On the other hand, according to this doctrine, it is
impossible to invent a story or describe a situation in which
analytically or logically true statements would be false. For
this reason, such statements were held to be absolutely
indubitable, and in principle always more certain than any
conceivable statement of fact.

It also follows from this view that no synthetic statement
(i.e. statement of fact) can ever be absolutely certain, since
to say of something that it is absolutely x (where x is a
predicate of some kind) is to imply that nothing could be more
x than that thing. To say of a mathematical calculation, for
example, that it is absolutely accurate is to say that no
calculation could possibly attain a greater degree of accuracy.
In other words, any statement of the form 'a is absolutely x,
but b is even more x' is strictly speaking self-contradictory.
Thus, if synthetic truths are in principle less certain than
analytic truths, it follows that they can never be absolutely
or apodeictically certain, in which case transcendental
deductions, as Kant conceived them, would be impossible. For
one would either have to deny that transcendental propositions*

*From now on, I will use the expression 'transcendental
proposition' in place of the more awkward 'conclusion of a
transcendental deduction'.
were apodeictically certain or that they were synthetic. It is perhaps worth noting that at least one commentator regards them as analytic. According to Jonathan Bennett 'the most interesting truths which Kant calls synthetic and a priori ... are unobvious analytic truths about the conditions under which certain concepts can have a significant use, affirmative or negative'.  

Kant, on the other hand, emphatically rejects the view that his philosophical conclusions are mere analytic truths, although he does want to say that they are certain in some ultimate sense. Indeed, his whole point in describing them as a priori-synthetic is to distinguish them from analytically true statements on the one hand, and ordinary empirically true statements on the other. Moreover, he clearly states that in order to be termed apodeictic, a proposition must be synthetic as well as absolutely certain. At A736 & B764 he says:

I divide all apodeictic propositions, whether demonstrable or immediately certain, into dogmata and mathemata. A synthetic proposition directly derived from concepts is a dogma; a synthetic proposition, when directly obtained through the construction of concept is a mathema. Analytic judgements ... do not extend our knowledge beyond the concept of the object, but only clarify the concept. They cannot therefore rightly be called dogmas.

---

It is of course possible that Kant is mistaken in thinking that transcendental propositions could be both non-analytic and absolutely certain, although I hope in due course to show that he is not. On the other hand, it must be admitted that his division of apodictic propositions into dogmata and mathemata is quite unacceptable since, as I remarked earlier,¹ the latter are inextricably bound up with the now much discredited doctrine of transcendental idealism; i.e. the doctrine that space and time do not belong to things as they really are in themselves, but are merely a priori intuitions. According to this theory, mathemata (identified by Kant as mathematical propositions) are apodictic because they 'describe' the a priori intuitions of space and time, rather than something a posteriori. The objections to this theory are well known,² and there is no need for me to set them down here. A far more important matter from our point of view is whether any statement could be correctly described as a dogma in the Kantian sense.

In order to be termed a dogma, a proposition must be: (1) apodeictically certain. (2) derived from concepts without the aid of intuition. (3) non-analytic. In other words, it must be a non-mathematical, a priori-synthetic principle. For Kant, all dogmas are transcendental propositions, and all transcendental propositions are dogmas. If, however, the

¹See p.22 of this thesis. ²See e.g. Strawson's Bounds of Sense.
traditional empiricist doctrine that no synthetic proposition can be as certain as any analytic proposition is correct, then dogmas, as Kant defines them, would not be possible, and therefore transcendental deductions would not be possible either. In other words, in order to defend Kant's concept of a transcendental deduction, we are obliged to attack the view that analytic propositions are in principle more certain than any conceivable statement of fact.

Section 3. Quine's attack on the doctrine of analyticity.

One fairly radical way of attacking the theory that analytic statements are intrinsically certain, whereas synthetic statements are in principle always open to doubt is to argue that there is in fact no real basis for the distinction between analytic and synthetic statements. Such an attack is to be found in Quine's article 'Two Dogmas of Empiricism'.

His arguments are designed to establish two main points: firstly, that any attempt to define the concept of analyticity in a philosophically illuminating way involves appealing (inevitably?) to terms which are equally in need of definition; that, for example, to explain analyticity in terms of cognitive synonymy would involve 'something like circularity'. The force of this point presumably lies in the presumption that we are not justified in drawing a distinction between analytic and

---

1 W.V.O. Quine, From a Logical Point of View, chapter II.
2 Ibid. p.30.
synthetic statements if we cannot give a precise account of the basis for this distinction. Secondly, in so far as the distinction has any validity at all, it is less sharply defined than the traditional doctrine would allow.

Quine argues on the one hand that no statement, not even an analytic one, is immune from revision in the light of experience, and on the other hand that no individual statement, not even an empirical one, is absolutely refutable by a direct appeal to experience. In other words, we are able to hang on to or reject whatever statements we like providing we are prepared to make the appropriate adjustments to the system of our beliefs as a whole, and providing we are prepared to put up with any inconvenience which that system might then have for us.

The force of this argument lies of course in the fact that according to the traditional theory, total immunity from revision is one of the chief distinguishing marks of synthetically true statements. Quine, however, maintains that this crucial aspect of the distinction cannot be upheld, and therefore the distinction, if it can still be said to exist, is a very murky one indeed.

Needless to say, this theory has met with considerable philosophical opposition. It is not, however, within the scope of this present work to determine whether or not Quine's
theory is correct. I propose, therefore, to tackle the problem of whether a genuine transcendental deduction, or something closely resembling it, could possibly exist, from both the Quinean and non-Quinean points of view. In fact, I hope to show that some kind of transcendental deduction is possible whichever of these theories is correct.

According to Quine, the distinction between necessarily true analytic statements and contingently true statements of fact is a spurious one. If he is right, it would surely be absurd to maintain that statements of fact can never attain that degree of certainty which analytic statements enjoy. Thus, one major objection to the possible existence of transcendental deductions would be removed. On the other hand, this would be achieved at the expense of denying that any statement at all was certain in the strict Kantian sense.

Indeed, it might be argued that from the Quinean point of view, even the notion that some statements are more nearly certain than other statements is strictly speaking inaccurate, and should be replaced by the notion of their being more firmly entrenched in the system than those other statements, and for that reason alone less susceptible to revision. We should then be forced either to abandon our quest for a genuine transcendental deduction altogether, or else to find a way of redefining the notion of a transcendental deduction.
in Quinean terms, without making use of the concept of ultimate certainty, at least in the form in which it is usually understood, but nevertheless sticking as closely as possible to Kant's original conception.

Quine has given the name 'theory of meaning'\(^1\) to that branch of semantics to which the concept of analyticity belongs. All the concepts belonging to this theory are for him equally suspect, since they are very closely related to one another. These include such concepts as 'meaning', 'synonymy' (or sameness of meaning), 'entailment' (or analyticity of the conditional), 'significance' (or possession of meaning). Clearly, then, if we are to succeed in our effort at making transcendental deductions respectable from the Quinean point of view, we cannot afford to use any of the above concepts in our exposition. In particular, of course, we must avoid making use of the analytic-synthetic distinction, even though it does play such a prominent part in Kant's own exposition.

Quine does not of course regard all semantic concepts as spurious, and indeed, regards that branch of semantics which he has termed 'the theory of reference'\(^2\) as perfectly legitimate. The main concepts in this theory are 'naming',

---

\(^1\)See his article 'notes on the Theory of Reference', Ch.VII, section 1, From a Logical Point of View, p.130.  
\(^2\)Ibid.
'truth', 'denotation' (or truth-of) and 'extension'. He also claims that 'in an explicitly quantificational form of language, the notion of ontological commitment belongs to the theory of reference'. Our task now is to see if we can explain how transcendental deductions might be possible in terms of concepts which, like those cited above, are respectable from the Quinean point of view.

As we have seen, the idea of a statement which satisfies the highest possible standard of certainty is central to the notion of a transcendental deduction, as opposed to a transcendental exposition. It looks, however, as though there is no room for such a concept within the framework of Quine's theory. According to this theory, although 're-evaluation of some statements entails re-evaluation of others, because of their logical interconnections' we have 'much latitude of choice as to what statements to re-evaluate in the light of a single contrary experience' and no statement is immune from such re-evaluation.

The only part of Quine's theory which is at all analogous to the traditional view that some statements attain a higher degree of certainty than other statements is the idea that some statements are more firmly entrenched in the system than other statements. For a firmly entrenched statement is

\[^{1}\text{ibid pp.130-131.}\ \ ^{2}\text{See ch.VII, 'Two Dogmas of Empiricism'; pp.42-43.}\]
one which is connected with many other elements in the
system, and is therefore less likely to be re-evaluated than
one with fewer connections. But clearly we have no right to
apply the concept of certainty to a statement simply because
we would be extremely unwilling to give it up, when the only
reason for this unwillingness is that it would involve us
in a great deal of inconvenience.

This leaves us with the problem of finding out whether
something akin to a transcendental deduction is still possible
if we repudiate the notion of absolute certainty on Quinean
grounds, or whether this would eliminate the distinction
between transcendental deductions and transcendental
expositions altogether.¹

The purpose of a transcendental exposition is to determine
what assumptions we are ultimately forced to make when we take
the validity of a particular system of knowledge for granted.
Kant employs this form of argument in the 'Aesthetic' where
he tries to show that space and time must be a priori intuitions
if mathematics is to be possible, and in the first part of the
Prolegomena where he tries to show, for example, that every
event must have a cause if physical science (which he identifies
with Newtonian physics) is to be possible.

According to Quine, all systems of knowledge (including
the logical and mathematical parts of those systems) are

¹See pp.21-25 of this thesis.
revisable in the light of experience, and no statement belonging to any system is immune from such revision. This seems to suggest that the most we can hope to succeed in doing by means of a transcendental argument is to uncover the fundamental presuppositions of a given system of knowledge at a particular stage of its development, although we can never rule out the possibility of the system's being changed in such a way as to bring about the introduction of a completely new set of assumptions. But obviously, to say that every statement we arrive at by means of a transcendental argument can only be affirmed in relation to a given system is to say that every successful or legitimate argument is a transcendental exposition.

As we have already seen,¹ Collingwood held a somewhat similar view, for he maintained that the only legitimate function of metaphysics was to ascertain what was absolutely presupposed in a given society, and how one set of presuppositions came to be replaced by another. Accordingly, he saw the Critique as being merely an investigation into the presuppositions of eighteenth century physics.

Not only is it impossible, on this view, to distinguish transcendental deductions from transcendental expositions, it is also difficult, if not impossible, to distinguish the

¹See pp.22-23 of this thesis.
transcendental method in general from certain non-philosophical modes of argument. For, thus understood, the conclusion of a transcendental argument simply states that in adopting a particular scientific system one is bound to make certain basic assumptions. And there is nothing distinctively philosophical about that.

Norman Kemp-Smith, arguing along rather different lines, comes to precisely the same conclusion. According to him, the transcendental method 'is really identical in general character with the hypothetical method of the natural sciences' and that 'considered as a method, there is nothing metaphysical or high-flying about it save the name'.

In order to preserve the distinction between transcendental deductions and transcendental expositions on the one hand, and between the transcendental and scientific methods on the other, it is necessary to refute the theory that any statement which is true according to one system of knowledge may be false according to another system, or modified version of the same system. The question now is whether this can be done without rejecting Quine's theory as a whole. I believe that an examination of some of his own basic assumptions will show that it can.

There is no such thing as a presuppositionless theory, as Quine would be the first to admit. Indeed, his own theory

\[\text{A Commentary to Kant's Critique of Pure Reason, p.xxxviii.}\]
takes for granted the legitimacy of certain semantic concepts (e.g. 'truth', 'denotation', 'extension' etc.) which he uses extensively in expounding his arguments, although he does question the legitimacy of certain other semantic concepts (e.g. 'analyticity', 'synonymy', 'entailment' etc.). But to say that the concepts of 'truth', 'denotation', and 'extension' are legitimate is to say, amongst other things, that it is possible to refer to something in order to talk about it, that some statements are true, and that a language of some kind exists. These are, therefore, basic assumptions in Quine's theory.

They are not, however, avoidable assumptions, for they are presupposed in all systems of knowledge, simply by virtue of their being systems of knowledge. For instance, take the statement 'a language of some kind exists'. Clearly, this is presupposed by every statement we make regardless of the body of beliefs from which it springs. In other words, in every system of knowledge or belief, the existence of a language of some kind is presupposed by every statement in that system.

We might say, therefore, that the statement is as firmly entrenched in the system as any statement can be. Indeed, it is so firmly entrenched that to re-evaluate it

1See From a Logical Point of View, p.130.
would not merely lead to a modification of the system or
even to its replacement by a different one, but rather to
the complete rejection of all systems.

At one point in his article 'On What There Is' Quine
says 'we commit ourselves to an ontology containing Pegasus
when we say Pegasus is'. Using similar terminology we might
now say that we commit ourselves to an ontology containing a
language whenever we say anything at all.

This conclusion is not, I believe, at variance with
any of the fundamental points in Quine's theory. We are not,
for instance, claiming that those statements which are
presupposed by all systems of belief are necessarily true,
nor that they are indubitable, in at least one sense of
indubitable.

A statement like 'a language of some kind exists' clearly
cannot be classified as a necessary truth, since we can easily
conceive of the non-existence of language. Nor, in at least
one sense, is the statement indubitable, since we are quite
free to reject the assumption that a language of some kind
exists, although if we do so sincerely we must regard all our
'assertions' as meaningless, and content ourselves with saying
nothing. In other words, a recalcitrant experience is quite
conceivable in such a case, although it could not then be
conceptualized by the person having the experience. Admittedly,

\[\text{From a Logical Point of View, p.8.}\]
therefore, the statement is immune from re-evaluation, but only because if it were not true there would be nothing to re-evaluate.

The distinction between those statements which can only be affirmed in relation to a particular system of knowledge (any scientific theory, for example) and those statements which can be affirmed in relation to any system of knowledge is a very important one from our point of view, because it enables us to distinguish transcendental deductions from transcendental expositions without appealing to the philosophically controversial notion of absolute certainty.

Thus, we may say that the purpose of a transcendental exposition is to determine what assumptions are fundamental within a particular system of beliefs, whereas the purpose of a transcendental deduction is to determine what assumptions are fundamental in any system; in other words, what must be taken for granted if we are to have any explicitly formulated beliefs at all.

If, as statement-makers, we are committed to the view that a language of some kind exists, we must also be committed to the existence of everything which is presupposed by this view. In his article 'On What There Is' Quine says that we commit ourselves to a particular ontological presupposition

\[\text{From a Logical Point of View, p.13.}\]
'if, and only if, the alleged presuppositum has to be reckoned among the entities over which our variables range in order to render one of our affirmations true'. The variable 'something' ranges over our whole ontology, and therefore if the statement 'there is something which is a language' presupposes the existence of persisting objects and recurring events, as Strawson claims, then it follows that we commit ourselves to an ontology containing persisting objects and recurring events every time we engage in discourse. An argument which showed we were in fact so committed would count as a transcendental deduction, according to our revised definition.

It is perhaps worth noting that Quine himself is quite prepared to say what must exist if language is to be possible. Thus, at one point in his article 'Identity, Ostension, and Hypostasis' he remarks that 'there is every reason to rejoice that general terms are with us, whatever the cause. Clearly language would be impossible without them, and thought would come to very little'. He is even more interested, however, in discussing the ontological commitments of particular kinds of discourse. For example, in his article 'Reification of Universals' he argues that the existence of universals is presupposed in classical mathematics.

1The Bounds of Sense, Part Two, Ch.III. 2From a Logical Point of View, p.77. 3From a Logical Point of View, Ch.VI.
An argument which attempted to uncover the presuppositions of classical mathematics, however, would count as a transcendental exposition rather than as a deduction.* For a deduction is concerned only with those presuppositions which cannot be re-evaluated in the light of any conceivable experience. But a principle is not immune from such re-evaluation simply because it is presupposed by a particular form of discourse, for it may be possible to deny it in terms of concepts which belong to a totally different form of discourse. For instance, the language of psycho-analysis is very different from the language of physiology, and yet it is conceivable that in the light of future experience the one may eventually completely supersede the other in the explanation of human behaviour.

*It should perhaps be mentioned, in passing, that there are various ways in which one might try to argue that a particular entity is presupposed by a particular form of discourse (or by language in general), but not every attempt would count as, or even look like, a transcendental argument; whether deduction or exposition. For example, one might argue that a particular entity is irreducibly presupposed by a particular form of discourse on the grounds that every previous attempt to eliminate all reference to that particular entity from the discourse has failed; e.g. the behaviourist's attempt to eliminate mental concepts from the language of human behaviour. In other words, one might try to establish the required conclusion indirectly by undermining or overthrowing arguments against it. A transcendental argument, on the other hand, is supposed to establish its conclusions directly.
On the other hand, a principle does not have to be presupposed by all forms of discourse, or even by all forms of statement-making, to be included in a transcendental deduction. It need only be the kind of principle which must be presupposed when anyone attempts to deny it. In other words, there must be no possibility of its being denied in terms of concepts belonging to another sphere of discourse. To take only a very trivial example: the statement 'it is physically possible for the concept of experience to be exemplified' is not, on the face of it, presupposed by the statement 'two plus two is equal to four'. Nevertheless, it clearly cannot be re-evaluated in the light of experience, nor is there any other way of refuting it, for it is itself an empirical statement.

My main aim in this section has been to re-interpret the concept of a transcendental deduction in terms that Quine himself would regard as legitimate, and at the same time to remove some of the more obvious objections suggested by his particular brand of pragmatism. Much, however, remains to be done before we can conclude that transcendental deductions are in fact possible. For instance, we must take a much closer look at that seemingly privileged class of statements which, as we have claimed, cannot be re-evaluated in the light of experience. We must also try to determine whether, despite
their apparent triviality, they give rise to any interesting conclusions.

Section 4. The concept of an a priori-synthetic principle.

According to the traditional view, the notion of ultimate certainty is supposed only to apply to analytic statements, whereas on Quine's view it cannot, apparently, be applied to any statement. Both theories, therefore, deny the possibility of there being any indubitable, synthetic truths, albeit for completely different reasons. This leaves us with the problem of explaining how transcendental deductions, as opposed to transcendental expositions, might still be possible. In the previous section, I tried to show how there might still be room for a form of argument closely resembling a transcendental deduction, even if Quine's theory is correct. I shall now try to explain how transcendental deductions are possible in terms of the traditional concepts.

The traditional empiricist view is that no proposition can achieve a higher standard of certainty than an analytic proposition, because the truth of an analytic proposition is quite literally undeniable. It is undeniable because, qua proposition, there is literally no way in which it could meaningfully be denied.

According to the traditional doctrine, analytic propositions are unique in this respect. For it is usually argued that
synthetic propositions can always be meaningfully denied, and so always fall short of the standard of certainty set by analytic statements.

The second part of this doctrine is clearly incompatible with the concept of a transcendental deduction. For a deduction is supposed to give rise to principles which are both synthetic and apodeictically certain; that is, as certain as anything can be.

In order to defend the concept of a transcendental deduction, we must see whether we can accept the first part of the doctrine without, at the same time, accepting the second part. I shall approach this problem by taking a closer look at what Kant meant by the term 'a priori-synthetic'.

When Kant says that transcendental propositions (i.e. non-mathematical, a priori-synthetic principles) can be derived from concepts without the aid of intuition, he is not of course claiming that they can be deduced from these concepts by means of the law of contradiction, for that would make them analytic. Indeed, he explicitly states that his transcendental proofs are based on 'principles of possible experience, and not on the principle of analysis (the law of contradiction).\(^2\)

It might be thought that at least part of the point being made here is that the transcendental mode of argument is presuppositional rather than implicational in form. For,

\(^1\) e.g. A736/B764. \(^2\) e.g. the footnote to A596/B624.
according to Kant's rather narrow concept of analyticity, a statement of the form 'x but not y' is self-contradictory if, and only if, x strictly entails y; i.e. where y is part of the meaning of x. A presupposition, however, is not part of the meaning of a statement, and therefore to assert the statement whilst denying its presupposition is not, on this view, self-contradictory. In other words, if x presupposes y without entailing it, then 'x but not y' is not self-contradictory, and 'if x then y' is not analytic.

For example, one could say that the statement 'Smith will attempt to swim the English Channel next month' presupposes but does not entail (in the strict sense) that Smith will be alive next month, and therefore the statement 'Smith will attempt to swim the English Channel next month, but he won't be alive next month' is self-defeating rather than self-contradictory. To say this, however, is merely to recognize the existence of a self-defeating principle (if x logically presupposes y, then a statement of the form 'x but not y' is self-defeating and so logically false). But clearly a proposition which is arrived at by means of the self-defeating principle alone is as much a logical truth (in the sense in which an analytic truth is a logical truth) as one arrived at by means of the principle of self-contradiction.
Since Kant defines analyticity in terms of the narrowly formulated version of the principle of self-contradiction, he would of course have to say that the statement 'if Smith attempts to swim the English Channel next month, then he will be alive next month' is non-analytic, although necessarily true. Similarly, if every event's having a cause is a necessary condition of the possibility of human experience, as Kant alleges, then the statement 'I have just witnessed a purely random event' would be self-defeating. Consequently, the statement 'if I have just witnessed an event, then every event has a cause' would be a non-analytic necessary truth.

The generalized point of these examples, of course, is that insofar as the transcendental method is presuppositional rather than implicational in form, it will give rise to propositions which are necessarily true, but (in the Kantian sense) non-analytic. Far more than this, however, is involved in saying that some propositions are a priori-synthetic, as I will now try to show.

If the alleged distinction between analytic and a priori-synthetic propositions could be entirely accounted for in terms of the difference between the entailment relationship and the presuppositional relationship, the distinction would be of no great philosophical importance. Indeed, it could be eliminated altogether by the simple expedient of extending
the concepts of analyticity, entailment, contradiction etc. so as to include presuppositional relationships; thus, in effect incorporating the self-defeating principle into the principle of self-contradiction. In other words, the need for a three-fold classification of propositions (viz. analytic, synthetic and a priori-synthetic) would only have arisen for Kant, because the logical terms which he employed had been too narrowly defined.

If this had in fact been the case, Kant's claim that his list of a priori-synthetic principles was complete would have been utterly absurd, for such a list would have had to contain an indefinitely large number of propositions including, for example, the unlikely proposition that 'if Smith attempts to swim the English Channel next month, then he will be alive next month'. Fortunately, however, we only have to examine the method of 'proving' these non-mathematical a priori-synthetic principles (i.e. the transcendental method) to realize that although these presuppositional inferences figure very prominently in these proofs, they do not in themselves give rise to a priori-synthetic principles.

As we have seen, a transcendental deduction is an argument of the form 'x is a necessary condition of the possibility of y. y is in fact possible. Therefore it is apodeictically

\[ \text{e.g. A148/B188.} \]
certain that x is in fact the case', or more briefly, 'the proposition y presupposes the proposition x, therefore x is
a priori-synthetic'. These formulations of the argument
clearly show that it is not the presuppositional statement
'if y then x', where x is logically presupposed by y, that is
supposed to be a priori-synthetic, but the categorical assertion
of x itself. Hence, the mere fact that a statement of
presuppositional inference cannot be classified as analytic
in the strict Kantian sense does not mean that it should be
regarded as a priori-synthetic. On the other hand, a priori-
synthetic principles are supposed to arise out of, or be
derived from, at least some such principles.

According to the argument of the Second Analogy, for
every example, the principle of universal causality is a priori-
synthetic, because its being true is a necessary condition of
the possibility of experience. In other words, the
presuppositional statement 'if there is (or could be) such a
ting as conscious experience, then every event has a cause'
is supposed to show that the sub-statement 'every event has
a cause' is a priori-synthetic.

It cannot of course be the case that every presuppositional
statement gives rise to a principle of this kind. For example,
what a priori-synthetic principle could possibly be derived
from the statement 'if Smith attempts to swim the English
Channel next month, then he will be alive next month'? That
Smith will be alive next month is obviously not such a principle.

This strongly suggests that a priori-synthetic principles are not, after all, the result of a special kind of inference. They can therefore only be based on a special type of premise. The nature of the premise will be examined in the next section.

Section 5. The initial premise of a transcendental deduction.

An important characteristic of any formal inference, whether presuppositional or implicational, is that the inferred statement may be more, but not less, certain than the statement from which it is inferred. For example, the evidence for the assertion 'all men are mortal' can never be as good as the evidence for the assertion 'the man Socrates is mortal', although the particular statement is of course entailed by the universal statement. Similarly, there is likely to be far more evidence for the claim that Smith will be alive next month than for the claim that Smith will attempt to swim the English Channel next month, although the former statement is presupposed by the latter.

In neither case, therefore, would one set out to prove that the inferred statement was true by first establishing the truth of the premise from which it was derived. In a transcendental deduction, on the other hand, one is supposed to be able to prove that a particular statement or principle
is absolutely certain, because it is presupposed by some other statement. It follows that the statement by which it is presupposed must itself be absolutely certain if the proof is to be a valid one.

The fact that the initial premise of a transcendental deduction must be certain, if the conclusion is to be certain, creates a serious problem for Kant, since if the conclusion is to be synthetic, as he alleges, it must be based on a synthetic and not on an analytic truth; and a synthetic truth, moreover, which is itself apodeictically certain. Kant, however, takes a very different view of the matter, and claims that his main aim in using the transcendental method is to 'develop knowledge out of its original seeds without seeking the support of any fact'; and furthermore insists that his philosophical conclusions are in fact 'deduced wholly in abstracto out of concepts'.

Although Kant does admit that for the purposes of exposition, it is sometimes useful to base philosophical arguments on acknowledge facts (this being the so-called 'analytic method' as employed in the Prologomena) he also clearly believes that in a truly philosophical argument no such appeal to facts should be made (the so-called 'synthetic method'). This also follows of course from his definition of philosophy as 'knowledge obtained by reason from concepts'.

1See p.29 in Prologomena, Lucas edition. 2Ibid p.34. 3B741.
In accordance with these views, Kant attempts to base his transcendental deduction at B131 on an analytic proposition, and therefore on a proposition which, in a very obvious way, can be derived solely from concepts. What this initial premise asserts is that 'it must be possible for the I think to accompany all my representations'. In other words, in order for these representations to belong to me, they must first of all be thinkable, otherwise they could not even be counted as potential representations. And secondly, they must be thinkable by me; that is, it must be possible to combine them in a single consciousness.

Kant calls this principle 'the transcendental unity of apperception', claiming it to be 'the highest principle in the whole sphere of human knowledge', although at the same time analytically true: 'the principle of the transcendental unity of apperception is itself, indeed, an identical, and therefore analytic, proposition'.

As we have seen, however, there is no way in which an apodeictically certain, synthetic proposition can formally be derived from an analytic proposition. Analytic propositions can give rise only to other analytic propositions. Moreover, it is obvious that any conclusions derived from such a principle will be empty unless its constituent concepts can be shown to

\[ 1^\text{B131}. \ 2^\text{B135}. \ 3^\text{B135}. \]
be instantiated in some way. Thus, if representations did not in fact exist the principle of the unity of apperception would be empty, and unworthy of serious philosophical consideration. That representations do in fact exist is clearly a synthetic, and not an analytic, truth, although it could perhaps be argued (in Cartesian fashion) that it would be impossible to doubt the existence of representations, as this would involve doubting whether one could think (for Kant, even thoughts are counted as representations).

We may safely conclude, therefore, that Kant's claim that he is able to derive important philosophical conclusions solely from concepts without appealing to any facts cannot possibly be correct. Granted, then, that transcendental deductions have to be based on synthetic premises, the question arises how such premises could be made to yield propositions which are apodeictically certain.

Section 6. The principle of the possibility of experience.

At one point, Kant does appear to admit that, in the end, transcendental deductions must be based on synthetic premises, and synthetic premises, moreover, which are merely contingently true. Thus at A737/B765 he says:

"Through concepts of understanding pure reason does, indeed, establish secure principles, not however directly from concepts alone, but always only indirectly through relation of these concepts to..."
something altogether contingent, namely, possible experience.

He then goes on to say that when such experience ... is presupposed, these principles are indeed apodeictically certain; but in themselves, directly, they can never be known a priori. Thus no one can acquire insight into the proposition that everything which happens has its cause, merely from the concepts involved.

What Kant appears to be saying here is that transcendental propositions (i.e. dogmata) can only be regarded as apodeictically certain, and therefore as a priori-synthetic, as long as the possibility of experience is taken for granted.

As it stands, however, this claim is open to the apparently serious objection that all presuppositional inferences could be classified as a priori-synthetic on exactly the same grounds. Thus, as we have seen, it could equally well be argued that the statement 'Smith will be alive next month' is a priori-synthetic, if the statement that Smith will attempt to swim the English Channel next month is taken for granted.

There are two possible ways of meeting this objection: one is to argue that the presuppositions of experience are in a quite different category from the presuppositions of anything else, because what is presupposed by experience cannot be known through experience, and must therefore be known a priori.*

*This, I believe, is really Kant's view.
The other, is to argue that we are in some way bound to take the possibility of experience for granted, whereas we are obviously not forced to assume that, for example, Smith will attempt to swim the English Channel next month. I will try to show that a far better case can be made out for the existence of a priori-synthetic principles on the basis of the second line of argument than on the basis of the first.

According to Kant, 'necessity and strict universality are ... sure criteria of a priori knowledge'.1 The reason for this is that 'experience teaches us that a thing is so, but not that it cannot be otherwise'2 and furthermore 'experience never confers on its judgement true or strict, but only assumed or comparative universality, through induction'.3 Kant also believes that anything which is 'indispensable for the possibility of experience'4 must satisfy the same criteria, and therefore be non-empirical. The fundamental presuppositions of Newtonian physics are cited as falling into this particular category; e.g. 'every event has its cause', 'the quantum of substance in nature always remains the same'.

It is true of course that if something is a necessary condition of the possibility of experience, then its existence can be determined without recourse to experience, and so,

---

1B4. 2B3. 3B3. 4B5.
in that sense, it can be said to be knowable a priori. We could, on the basis of this, make claims to the effect that 'within the field of our experience such and such must always be the case' or 'the world of which we are aware necessarily has such and such attributes' which at least appear to satisfy the criteria of universality and necessity.

It is important to realize, however, that such knowledge is not a priori in the true Kantian sense of being 'absolutely independent of all experience'. For, as Kant admits, to be a priori in this sense, a principle must not be 'derived from any proposition except one which also has the validity of a necessary judgement'. But, clearly, we cannot claim to know that something is actually the case, simply because it is indispensable for the possibility of experience unless we can also claim to know that experience itself is actual. However, the actuality of experience is in no sense a necessary truth.

In other words, in the last resort, knowing that within the field of our experience, such and such must always be the case, or that the world of which we are aware, necessarily has certain attributes, involves knowing that experience itself is actual (and therefore possible) as a result of the fact that we ourselves (or I myself) have conscious experiences.

Perhaps the main reason why Kant feels justified in regarding the fundamental principles of Newtonian physics as
a priori-synthetic rather than analytic is that they are not only necessarily true (according to him), but true of the phenomenal world (i.e. the world of our experience), whereas analytic truths are non-descriptive and therefore not about the world.

The fallacy here of course is that even if these principles do state the necessary conditions of the possibility of experience, it does not follow that they are true of the world in the sense which is needed in order for them to be classified as synthetic. Indeed, it is fairly obvious that at most they would be true of the world in precisely the same sense as the proposition 'red is a colour' is, on one interpretation, an analytic truth about redness. In other words, on this view, the Newtonian principles would be analytically true of any world of which we could be aware; i.e. a phenomenal world; that is to say, they could be analytically derived from the concept of such a world.

I have so far considered, and rejected, three possible reasons for classifying transcendental propositions as a priori-synthetic; namely (1) that transcendental deductions are presuppositional rather than implicational in form. (2) that what is logically presupposed by experience can be known independently of experience. (3) that such propositions are both necessarily true and 'true of the world'. There is,
however, another possibility at which I have already hinted; that is, that the propositions upon which transcendental deductions are based belong to a class of contingently true propositions which are necessarily presupposed in any attempt to deny them. This class consists mainly, but not entirely, of those propositions which we are logically forced to take for granted in all our assertions.

It seems not unreasonable to apply the term a priori-synthetic to such propositions, since they are like analytic propositions in that there is no way in which they can be meaningfully denied, but like ordinary empirical propositions, in that they describe states of affairs which might conceivably have been different. Kant would of course have objected to this, on the grounds that non-contingency or necessity is, and must be, a defining characteristic of a priori-synthetic truth. Later on, however, I will try to meet this objection by arguing that there is a sense in which a proposition can be both contingent and, at the same time, no less certain than if it had been analytic. But first we will try to determine whether any of Kant's transcendental deductions could be based on a proposition belonging to this privileged class.

As we have already seen, transcendental deductions can only give rise to a priori-synthetic principles if they are

1See pp.45-50 & p.60 of this thesis.
based on premises which are themselves both synthetic and apodeictically certain. Thus, we cannot show that something is actually the case, simply because it is a necessary condition of the possibility of experience, unless we can in some way show that we are logically entitled to take the actuality of experience for granted.

Although it is only contingently true that experience is actual, Kant could have defended its status as a source of a priori-synthetic principles, by appealing to his so-called principle of significance. This is the principle that 'all concepts, and with them all principles, even such as are possible a priori, relate to empirical intuitions, that is, to the data for a possible experience. Apart from this relation they have no objective validity'. Kant offers no independent proof of this principle, although the Critique as a whole could be interpreted as an attempt to answer certain obvious objections to it; e.g. the objection that not all our concepts are derived from experience.

One very important consequence of this principle is that anything which is a necessary condition of the possibility of experience will, ipso facto, also be a necessary condition of the intelligibility of any concept, and therefore of language as a whole. Thus, if every event's having a cause is a

\(^7\text{B195.}\)
necessary condition of experience, it must also be a
necessary condition of the possibility of language. This
means that one cannot deny the principle of universal
causality without taking it for granted, for in the mere use
of language its truth is already presupposed. In other words,
if Kant's principle of significance is correct then the falsity
of the assertion 'there is no such thing as sensory experience'
can be determined on logical grounds alone, and hence the
truth of any statement which is presupposed by the concept of
experience will be quite literally undeniable.

Although there is no way in which the truth of the
statement 'sensory experience of some kind actually exists'
can be meaningfully denied (that is, providing one accepts
the principle of significance) it is nevertheless not a
necessary truth, since we can quite well imagine what it would
be like for there to be no sensory beings in the world at all.
For this reason, to say that there is no such thing as
sensory experience is neither self-contradictory nor self-
defeating, although the claim that the statement, qua
utterance, was in fact true would be self-defeating.* The
same can of course be said of such statements as 'no sentence
ever conveys anything' and 'no language of any kind could ever

*The distinction between self-defeating propositions and self-
refuting assertions, which is an important one here will be
fully explained in Part II.
exist'.

Although statements like these are empirical, and describe perfectly conceivable situations, there is logically no way in which a person could correctly claim them to be true. I will be examining this type of statement in some detail in Part II under the title of 'self-refuting statements'. All that need concern us for the moment, however, is the mere fact of their existence, from which we can infer the existence of other empirical statements, whose truth we are logically forced to take for granted whenever we assert anything at all; e.g. 'some sentences are meaningful' and 'some kind of a language is possible'. Such statements as these surely deserve to be termed apodeictic since, like analytic propositions, there is no way in which they can be meaningfully denied.

Section 7. The structure of a transcendental deduction.

We are now, I think, in a position to provide at least a tentative answer to the question: how could a statement of the form 'it is apodeictically certain that x is in fact the case' be validly inferred from a statement of the form 'x is a necessary condition of y'. For, if x is necessarily presupposed or entailed by y, and y belongs to that class of empirically true statements which we cannot deny without assuming to be true, then it follows that x will also be empirically, yet undeniably true.
The structure of a genuine transcendental deduction, therefore, will look something like this: to deny $x$ (e.g. every event has a cause) is to presuppose the truth of $y$ (e.g. experience or language is actually possible). But the truth of $y$ presupposes the truth of $x$, therefore $x$ must in fact be true. Or alternatively, to assert $x$ (e.g. some events do not have causes) is to presuppose the falsity of $y$ (experience or language is actually possible). But to assert the falsity of $y$ is to presuppose the falsity of $x$, therefore $x$ must in fact be false.

It seems not unreasonable to apply the term a priori-synthetic to philosophical conclusions arrived at in this way, since the proof is a priori and the conclusions synthetic.

On the other hand, a priori-synthetic principles in this sense are contingently and not necessarily true. This cuts across Kant's classification to some extent. For it is a fundamental tenet of the Critique that we cannot know anything to be necessarily the case as the result of experience, and that we cannot know anything to be contingently the case independently of experience, that is, a priori. Thus, at B4 he says 'necessity and strict universality are ... sure criteria of a priori knowledge'. In other words, for Kant, necessity or non-contingency is, and must be, a defining characteristic of a priori-synthetic truth.
However, I believe that Kant is wrong both in his belief that a priori-synthetic principles so defined actually exist, and in his belief that no contingently true propositions could reasonably be described as a priori-synthetic. Thus I have tried to show how an a priori-synthetic principle might be defined as an empirically true proposition which, at the same time, is literally undeniable.

It should be mentioned at this point that the possibility of using this type of principle as the basis of a transcendental deduction has already been recognized by at least one student of the transcendental method. Thus, Barry Stroud, in an article in the Journal of Philosophy for May 1968, has given the following description of such an argument:

Suppose we have a proof that the truth of a particular proposition S is a necessary condition of there being any meaningful language, or of anything's making sense to anyone. For brevity, I will say that the truth of S is a necessary condition of there being some language. If we had such a proof we would know that S cannot be denied truly, because it cannot be denied truly that there is some language. The existence of a language is a necessary condition of anyone's ever asserting or denying anything at all, and so if anyone denies in particular the proposition that there is some language it follows that it is true. Similarly, it is impossible to assert truly that there is no language. This suggests that there is a genuine class of propositions each member
of which must be true in order for there to be any language, and which consequently cannot be denied truly by anyone, and whose negations cannot be asserted truly by anyone. 1

Stroud then goes on to argue that a transcendental deduction, constructed along these lines, is not by itself capable of establishing any of those things for which the transcendental method was originally designed; that is to say, it cannot demonstrate the objective validity of those concepts and principles which the epistemological sceptic calls into question.

This means, for instance, that it cannot (without invoking some version of the verification principle) be used to demonstrate the existence of the external world, the existence of other minds, the continued existence of unperceived objects, the reliability on the whole of sense-perception, the reliability on the whole of memory, or the truth of the principle of induction.

The general claim that transcendental deductions alone cannot refute scepticism does not of course mean that they are incapable of refuting or proving anything. This is important from our point of view, since we are mainly concerned with showing that transcendental deductions of a kind are possible, and with investigating in depth some of their more interesting logical characteristics. We can best achieve 1pp.252-253.
these aims by concentrating our attention on relatively straightforward and unsophisticated versions of the argument. We need not, therefore, concern ourselves with the much larger question of whether or not they provide the final solution to some of the most difficult and long-standing problems in philosophy.

On the other hand, Stroud does raise a couple of points which are well worth considering. Firstly, that "talk about 'language in general' or 'the possibility of anything's making sense' is so vague that there seems no convincing way of deciding what it covers and what it excludes." And that for that reason, the conclusions of any genuine transcendental deduction are likely to be either suspect or utterly trivial.

This is certainly a point that will have to be considered, but since it raises a number of vital questions concerning the nature of self-refuting statements in general (for instance, whether more than one type of self-refuting statement may be invoked in a transcendental deduction) it cannot be adequately discussed without a great deal more background information on the subject. This will be provided in Part II.

*In due course, I will try to show that it is possible to base transcendental deductions on rather more precise principles than these.

The second point which Stroud raises can be dealt with now. It is that, for any proposition \((S)\) which states a necessary condition of language in general, or of the possibility of anything's making sense,

the skeptic can always very plausibly insist that it is enough to make language possible if we believe that \(S\) is true, or if it looks for all the world as if it is, but that \(S\) needn't actually be true. Our having this belief would enable us to give sense to what we say, but some additional justification would still have to be given for our claim to know that \(S\) is true.\(^1\)

This additional justification can only be acquired, apparently, by appealing to some version of the verification principle. For it would be necessary to show that one of the conditions of anything's making sense was that we were in a position to know that such and such was the case, and not that we merely believed that it was. 'Hence the meaning of a statement would have to be determined by what we can know. But to prove this would be to prove some version of the verification principle'.\(^2\)

Stroud does not argue that any such appeal to the verification principle is illegitimate, but only that to demonstrate the self-defeating character of scepticism by means of a transcendental argument 'would amount to nothing more and nothing less than an application of some version of the verification principle, and if this is what a transcendental

---

\(^1\)ibid p.255.  \(^2\)ibid p.255.
argument is, then there is nothing special or unique, and
certainly nothing new, about this way of attacking scepticism'.

The suggestion that, in attempting to establish objective
truths, transcendental deductions invariably end up by
appealing to some version of the verification principle (in which
case, they would be superfluous) can be refuted quite simply.
For although Stroud's claim that we can discover only what we
must believe in order for language to be possible, not what we
are in a position to know, is true of at least some transcendental
deductions, it is certainly not true of all, as the following
example will show.

A necessary condition of the meaningful use of any language
is the existence, at one time or another, of a thinking being.
For language, as opposed to a meaningless jumble of 'symbols'
presupposes the existence, at some stage, of a being capable
of understanding what is said, or what he himself says.

The point, of course, is that we are not merely required
to believe that such a being exists, but that he actually does exist. In other words, from the self-guaranteeing statement
'language in general is possible' we can infer the existence
of a thinking being as an objective fact, albeit a contingent
one. And no appeal to the verification principle is needed to
effect the transition from what we believe to what we know.

1ibid 255-256.
The knowledge claim is justified, even from the sceptic's point of view, because there is no way in which it can be meaningfully denied. It is a presupposition of discourse, and therefore to doubt it would be to doubt all statements, including necessarily true ones, even though it is not itself a necessary truth.

On closer examination, it can be seen that even those transcendental deductions which appear to lend support to Stroud's theory need not necessarily invoke the verification principle in order to justify certain knowledge claims, although they must do so if they are to justify certain other, more ambitious, knowledge claims.

For instance, it might be argued that the statement 'language in general is possible' is not only empirically true, but self-guaranteeing. But that a necessary condition of making any kind of empirical assertion is the ability to employ empirical concepts (e.g. 'language'). And that a necessary condition of employing empirical concepts is the ability to bring items of experience under concepts, and this in turn presupposes a certain degree of regularity and uniformity in the world. Hence, to deny that the world exhibits a degree of regularity and uniformity is to deny the truth of the self-guaranteeing statement 'language in general is possible'.
Stroud would argue, rightly I believe, that the most we are entitled to infer from this argument is that the world must appear to us as ordered and regular, and not that it must really be so. And that, in fact, this further claim cannot be established without invoking the verification principle. But not even Kant would quarrel with this, for he readily admits that his conclusions are only true of the world as it appears to us (the phenomenal world) and not of the world as it really is in itself (the noumenal world) which is unknowable.¹

Moreover, it should not be thought that a transcendental deduction of this kind can only justify a tentative knowledge claim of the form 'it seems as though x is the case, but it might not really be so' as Stroud seems to imply. We need only look at the argument under consideration to see this. For it entitles us to say not merely that it seems as though the world exhibits a certain degree of uniformity and orderliness, but that the world is, in fact, perceived as exhibiting a certain degree of uniformity and orderliness.*

*N.B. We are not (at this stage anyway) entitled to make the further claim that the world is an entity existing independently of the percipient. Nor that the world is necessarily perceived as exhibiting a degree of uniformity and orderliness. However, the point is we cannot coherently assert that it is not perceived in this way.

¹e.g. B303.
A more general objection to Stroud's claim that transcendental deductions are faced with a gap between meaning and truth which they can only cross by invoking some version of the verification principle is that this claim is based on the questionable assumption that the only principle which could function as an axiom in a transcendental deduction is one to the effect that a meaningful use of language is possible. However, there are other possibilities.

For instance, it might be argued that the 'I think', on which Descartes based his philosophical system, would do equally well. It might even be claimed that this is what Kant's own deductions are based on. Or again, it might be argued that the proposition 'some assertions are true' could function as a transcendental axiom, since there is no way in which it could be meaningfully denied.* These possibilities will all be discussed in Part II.

As it has so far been described, the transcendental method is parallel in at least one respect to the geometrical method; that is to say, it attempts to demonstrate the truth of certain theorems by logical inference from a set of axioms. Unlike the axioms of geometry, however, the axioms of transcendental philosophy are not at all arbitrary or replaceable, since there is no way in which they can be...

*There are special difficulties associated with this one. See section 12 Part II.
meaningfully denied. This whole approach is of course strongly reminiscent of Descartes' attempt to derive an entire metaphysical system from a single indubitable empirical premise. Indeed, as I shall try to show in due course, the transcendental axioms themselves have a strongly Cartesian flavour.
PART II

The Concept of Self-Refutation

Section 1. The introduction.

In Part I, I argued that if the expression 'apodeictically certain' is to be ascribed to any statement whatever, then it should at least be ascribed to those statements which we are forced to take for granted if we are to say anything at all. I pointed out that one of the main characteristics of such a statement is that any attempt to deny it is necessarily self-refuting. However, there are many other kinds of self-refuting statement besides this one. In this part of the enquiry, I propose to examine as wide and varied a range of them as possible, not merely because the subject is intrinsically interesting, but with the ultimate aim of showing how it is possible for propositions derived from one type to be used as axioms in a transcendental deduction, whereas propositions derived from various other types are not suitable for this purpose.

No analysis of the concept of self-refutation can be regarded as adequate if it does not take into account the many different kinds of self-refuting statement and the complex relationships that exist between them. This can best be shown by means of a general and comprehensive classification.

Only two philosophers have so far attempted to classify the various types of self-refuting statement: Professor Passmore in chapter 4 of *Philosophical Reasoning*, and J.L.
Mackie in the *Philosophical Quarterly* for July 1964. Both of these will be discussed in the following sections, and on the basis of this I hope to come up with an even more comprehensive, detailed, and exhaustive classification.

It should perhaps be pointed out, in passing, that almost any statement which is put forward as an example of a self-refuting statement could, on a different interpretation, be regarded as a non-self-refuting statement; indeed, it might even be more natural to interpret it in that way. From our point of view, however, this possibility can be safely ignored, since we are only interested in finding out what kinds of self-refuting statement there could be. Thus, whenever I describe a statement as self-refuting or self-guaranteeing, I wish it to be interpreted in such a way (even if that way is eccentric or unusual) that it is self-refuting or self-guaranteeing.

Section 2. The various forms of self-reference.

In order to be self-refuting a statement must be, in some sense, self-referring, and the way in which it is self-refuting will be at least partly determined by the kind of self-reference involved.

There are basically two classes of self-referring statement, although, as we shall see, it is necessary to sub-divide them in various ways. In the one class, the statement refers to
itself as a whole (i.e. as a single statement), whereas in the other class, the statement merely refers to a sub-statement within itself. I shall describe those statements belonging to the former class as wholly self-referring, and those belonging to the latter class as partly self-referring.

The following statements may be interpreted as wholly self-referring: 'this is a strange proposition', 'this sentence is meaningless', 'everything I say is true', 'I am now talking very loudly'. The first example refers to itself as a proposition, the second refers to itself as a sentence, the third refers to itself as an assertion, and the fourth refers to itself as a speech-act (in this case, by describing the manner in which it is made).

Wholly self-referring statements may also differ from one another in being either explicitly and exclusively self-referring, or implicitly and incidentally self-referring. Thus the statement 'this is a very strange proposition' explicitly refers to itself and to no other proposition, whereas the statement 'everything I say is true' refers to a whole class of statements of which it is itself a member. I shall term wholly self-referring statements of the former type 'exclusively self-referring' and those of the latter type 'incidentally self-referring'.

The following statements are partly self-referring: 'it can be proved that the world is round', 'I sometimes write
that I cannot write', 'I think that I know nothing'. These are all second-order statements of which the first-order statement forms a constituent part (unlike, for example, the second-order statement 'what you have just said is false'). Thus, 'it can be proved that the world is round' is a second-order statement about the proposition 'the world is round' and 'I sometimes write that I cannot write' is a second-order statement about the proposition 'I cannot write', and 'I think that I know nothing' is a second-order statement about the proposition 'I know nothing'.

It might be thought that some or all of these statements could be re-expressed as first-order statements without any real change of meaning. If this was so, they would be partly self-referring only because they were expressed in a particular way. For example, although one would be more likely to say 'it can be proved that the world is round' rather than 'the world's roundness can be proved' this is obviously just a linguistic accident. The latter statement, however, is grammatically similar to the statement 'the cushion's softness can be felt, which is clearly not a second-order statement and in no sense self-referring. One might be tempted to conclude from this that 'the world's roundness can be proved' is also non-self-referring.

This would be a mistake, I think, since, logically speaking, only propositions can be proved, and therefore to say of anything...
that it can be proved is necessarily to make a second-order statement. Thus, although 'the world's roundness can be proved' appears to be a statement about the world's roundness, it is in fact a second-order statement about the proposition 'the world is round'. And since it is a second-order statement about the sub-statement 'the world is round' it is also partly self-referring.

Any of the partly self-referring statements cited above could be true, and therefore none of them is directly self-refuting. We must therefore ask ourselves what it is that distinguishes those partly self-referring statements which are self-refuting from those which are not. In other words, to what sub-class of self-referring statements do those self-refuting statements belong?

One does not need to look far to realize that some partly self-referring statements are self-referring in a double sense. In these cases, not only does the second-order statement refer to a sub-statement within itself, but the sub-statement in turn refers back to the second-order statement. For example, 'it can be proved that nothing can be proved', 'I sometimes write that I cannot write', 'I am telling the truth when I say that everything I say is true'.

In the first example cited above, the sub-statement 'nothing can be proved' entails that any second-order statement...
of the form 'it can be proved that ...' is false. And in the second example, the sub-statement 'I cannot write' entails that any second-order statement of the form 'I sometimes write that ...' is false (providing the reference of 'I' is the same in both cases and 'I cannot write' is taken to mean 'I never write'). And in the third example, the sub-statement 'everything I say is true' entails that any second-order statement of the form 'I am telling the truth when I say that ...' is true (again, providing the reference of 'I' is the same in both cases).

Partly self-referring statements which are self-referring in this double sense I shall term 'reflexively self-referring'. Those which are self-referring in the more marginal sense I shall term 'partially self-referring'. Clearly a partially self-referring statement can be false (e.g. 'it can be proved that the world is square') but never self-refuting. Its only importance from our point of view is the light it throws on the concept of self-reflexiveness.

There is perhaps no a priori method of proving that for a statement to be self-refuting, it must be either wholly self-referring or reflexively self-referring. But if, after a detailed investigation, it can be shown that a wide range of statements normally termed self-refuting can be comfortably accommodated in one or other of these categories, and if we fail to discover any counter-examples, then this would constitute
some kind of a proof. In other words, if the various types of self-refuting statement can be successfully classified on the basis of such an assumption, then the assumption is surely justified.

A statement may also refer to itself via another statement. Here are three examples: (1) Smith says 'my next statement can be proved'. Smith's next statement is 'nothing can be proved'. (2) On one side of a black-board is written 'the statement on the other side of this black-board is true'. On the other side of the black-board is written 'the statement on the other side of this black-board is false'. (3) Smith says 'Jones' next statement will be a strange one'. Jones' next statement is 'what Smith has just said is true'. This type of self-reference may be termed indirect self-reference.

From the point of view of our investigation, the problems which arise in the case of indirect self-reference are no different in principle from those which arise in the case of direct self-reference. On the other hand, it is doubtful whether any indirectly self-referring statement could be logically self-refuting, since it is surely a contingent matter that some statement or other is identifyingly referred to by some other statement. For example, it is surely a contingent fact that when Smith says 'my next statement can be proved' the statement which he goes on to make is that nothing can be
proved. Someone might have forced him at gun-point to say something different, for example. Or again, it is surely only a contingent fact that the statement on the other side of the black-board says what it does say.

Even if indirectly self-referring statements are interpreted as logically self-referring, they present us with no new problems, since they can be shown to be logically equivalent to directly self-referring statements, and therefore classifiable in precisely the same way; viz. as either wholly self-referring or reflexively self-referring. Thus the mutually self-referring statements in the first example jointly entail the reflexively self-referring statement 'the statement that nothing can be proved can be proved'. The two statements in the second example jointly entail the wholly self-referring statement 'this statement is false'. The two statements in the third example jointly entail the reflexively self-referring statement 'Jones' statement that Smith's statement that Jones' statement is a strange one is true is a strange one'. The logical form of this last highly awkward statement becomes clearer if we let 'd' stand for the expression 'the Jones-statement that ... is strange' and 'p' for the Jones-statement 'the statement that the Jones-statement is strange is true'. We may then say that d entails that p is true and p entails that dp is true.
The following diagram shows how the various types of self-referring statement that have been described in this section are related to one another:

```
SELF-REFERRING STATEMENTS

WHOLLY SELF-REFERRING       PARTLY SELF-REFERRING

INCIDENTALLY SELF-REFERRING1 EXCLUSIVELY SELF-REFERRING2

PARTIALLY SELF-REFERRING3 REFLEXIVELY SELF-REFERRING4

1 e.g. 'all the statements on this page are legible'.
2 e.g. 'this statement is false'.
3 e.g. 'it can be proved that some statements are false'.
4 e.g. 'it can be proved that nothing can be proved'.
```
Section 3. Meaning and Self-Reference.

If, as we have seen, transcendental deductions must be based on a premise which is self-guaranteeing because any attempt to deny it would be self-refuting, and if we accept the view that self-refuting statements are always in some sense self-referring, then we are faced with what is potentially a serious objection to the possibility of there being any genuine transcendental deductions at all.

The objection arises out of the theory that all self-referring statements are meaningless. For if this were so, it would imply that self-refuting statements and their contradictories were also meaningless, in which case no genuine transcendental deduction would be possible.

The theory that all self-referring statements are meaningless was originally advanced by Bertrand Russell as a direct consequence of his theory of types. Since then, it has been strongly advocated on other grounds by such philosophers as Jorgen Jorgensen and Alf Ross.

The chief advantage of the theory is, of course, that it disposes very neatly of the so-called self-referential

paradoxes, of which the liar paradox is perhaps the best known example (viz. if the statement 'this sentence expresses a false proposition' is interpreted self-referringly, it follows that it can only be true by being false, and can only be false by being true)."

The theory's chief disadvantage, from our point of view, is that it appears to rule out as meaningless even those self-referring statements which do not generate paradoxes. For this reason, it is generally held to be a highly controversial theory, and since its formulation, has been strongly attacked by such philosophers as H.L.A. Hart, Karl Popper, and Niels Christensen.

I do not myself propose to attack the theory directly, since I can afford to concede a great deal of what its proponents claim without in any way damaging my own thesis. For it is not necessary for me to prove that the theory is completely false, but only that it does not apply to literally all self-referring statements. Thus, all I have to do is to show that at least some self-referring statements are meaningful, namely those which enter into transcendental deductions.

---

1 'Self-Referring Laws' in Festkrift till Karl Olivecrona, 1964.
2 'Self-Referring Statements', Mind 1954.
*For further examples see W.V.O. Quine, From a Logical point of View, pp.132-134.
Both Popper\(^1\) and Hart\(^2\) give several examples of statements which are in a sense self-referring, but at the same time perfectly meaningful. One of Popper's examples is the statement 'I am now whispering so softly that dear old Socrates cannot make out what I am saying' spoken by someone in the presence of Socrates. One of Hart's examples is the statement 'this sentence is written in English'. The truth or falsity of both these statements can be determined in a straightforwardly empirical manner, and therefore they are clearly meaningful.

However, as Alf Ross has pointed out,\(^3\) if one accepts a slightly modified version of the self-reference theory, one is not bound to regard these last two statements as meaningless. For, as he claims, it seems a perfectly reasonable hypothesis that 'the vice of self-reference occurs when one tries to express in a sentence a meaning that refers to the meaning of the same sentence' (e.g. 'this sentence expresses a false proposition'), and that no such defect occurs 'in a sentence expressing a meaning that refers either to the sentence itself as a grammatical construction or to the speech-act as a sound sequence'. According to this account, the first example

can be regarded as meaningful because it refers to itself simply as a sound-sequence, and the second example can be regarded as meaningful because it refers to the sentence which bears the meaning and not to the meaning itself.

A much more radical modification of the self-reference theory has been suggested by H.L.A. Hart. In his opinion, the theory is true only of totally self-referring statements, and not of statements which are only partly self-referring.

It should be noted, however, that by 'totally self-referring' Hart means what I mean by 'exclusively self-referring' and by 'partly self-referring' he means what I mean by 'incidentally self-referring'. Thus, he regards an exclusively self-referring statement like 'this statement is true' as meaningless, but an incidentally self-referring statement like 'all the statements on this page (including this one) are true' as meaningful.

On an intuitive level, the theory seems to be quite plausible, because nothing at all seems to be involved in saying that the first statement is true, whereas one is tempted to say that one knows perfectly well what it would be like for the second statement to be true, namely, that all the other statements on the page would have to be true. For, if all the other

---

statements on the page were true, the meta-statement would also be true whether it was interpreted self-referringly or not. In other words, if we let a, b, c, etc. represent all the other statements, and if we let S1 represent the meta-statement that they are all true, then we may say that if a, b, c, etc. are all true then S1 must be true, and if S1 is true then it follows that all the statements on the page including the meta-statement are true, and this we may call S2 (S2 being the meta-statement interpreted self-referringly).

It is not difficult, however, to find reasons for doubting the validity of the above argument. One fairly cogent reason is that in interpreting the statement 'all the statements on this page are true' self-referringly, one is forcing it to function both as a meta-statement (in so far as it refers to all the other statements) and as a meta-meta-statement (in so far as it refers to itself as a meta-statement), and there is certainly something logically odd about this.

One might also reasonably argue that the mere fact that S1 and S2 are true under precisely the same conditions means that they cannot properly be distinguished from one another, and that therefore the move from S1 to S2 is empty. The general point here is that if the truth conditions of a statement are in no way altered by interpreting it self-referringly, then the statement is vacuous in so far as it is interpreted
self-referringly. This would seem to support Alf Ross's claim that 'if $S$ refers to $a$, $b$, $c$, and $S$ ... $S$ is meaningful in so far as it refers to $a$, $b$, $c$, and meaningless in so far as it refers to itself.'

Although I am prepared to accept the above conclusion for some incidentally self-referring statements, I hope to show that it does not apply to all such statements, not even to all those statements in which 'one tries to express in a sentence a meaning that refers to the meaning of the same sentence'. One such example is the statement 'no statement (including this one) can be proved'. This can be demonstrated, I believe, by means of the following arguments.

No philosopher has ever tried to argue that (what I have termed) reflexively self-referring statements are meaningless. Indeed, it would obviously be absurd to do so, since no special difficulties arise in determining the truth or falsity of reflexively self-referring statements which do not also arise in the case of non-self-referring statements. Thus, one is not in the least tempted to say that such statements as 'I sometimes write that I cannot write', 'it cannot be proved that nothing can be proved', and 'I know that I know nothing' are meaningless; for the first could easily be true, the second obviously is true, and the third is necessarily false.
Furthermore, none of the arguments advanced by the proponents of the self-reference theory (e.g. Russell's theory of types) has any tendency to show that reflexively self-refering statements are meaningless. In short, it is quite obvious that the theory was only ever intended to apply to (what I have termed) wholly self-refering statements.

If we let $S$ represent the statement 'no statement can be proved' and if we let $a, b, c$, etc. represent all other statements, then we may say that when $S$ is interpreted in a non-self-refering way, it merely entails that $a, b, c$, etc. cannot be proved. But when it is interpreted as an incidentally self-refering statement, it entails that the further proposition 'no statement can be proved' is itself unproveable. In other words, in this case it would entail the reflexively self-refering statement 'it cannot be proved that nothing can be proved'. But this, as we saw in the previous paragraph, is both meaningful and true. Hence $S$ must surely be meaningful when interpreted as a self-refering statement.

Neither of the objections I raised earlier against treating the statement 'all the statements on this page are true' as self-refering really apply to the proposition 'no statement can be proved'. For unlike the former statement its truth conditions do change when it is interpreted self-
referringly. Thus, it could be true in so far as it referred to all other statements, but false in so far as it referred to itself. For the unproveability of the statement 'all other statements are unproveable' cannot logically be inferred from the truth of that statement, in the way that the truth of the statement 'all the other statements on this page are true' can be inferred from the truth of that particular statement. In other words, the sentence 'no statement can be proved' clearly expresses a different proposition when interpreted self-referringly,* and in this respect at least differs significantly from a statement like 'all the other statements on this page are true'.

The above conclusion also takes the sting out of the objection that one should not force a statement to function simultaneously on two different logical levels (e.g. as a meta-statement and as a meta-meta-statement). Of course, the statement 'no statement can be proved' does function both as a meta-statement and as a meta-meta-statement, but in a perfectly harmless way. For the class of 'all statements' must include both first-order and second-order statements, and therefore any statement about such a class will function as a meta-statement in so far as it says something about first-order statements, and as a meta-meta-statement in so far as it

*This is only to say 'no statement can be proved' is different from 'no statement (except this one) can be proved'.

says something about second-order statements. However, there is no reason why one should not lump statements and meta-statements into one class in order to make a statement about them. There is nothing logically odd, for example, about the statement 'different levels of statement are discussed in this chapter'. And since the statement 'all statements are unproveable' is clearly intended to refer to all statements, regardless of their logical status, why should its own logical status prevent it from being included in that class?

Moreover, the sense in which a statement like this can be said to function on two levels at once is quite different from the sense in which a self-referring statement like 'all the statements on this page are true' can be said to function on two levels at once. For whereas the one is simply talk about different levels of statement, the other involves different levels of talk about statements. And it is this last feature which gives rise to logical difficulties, as I shall now try to explain.

As we have seen, the statement 'all the statements on this page (a, b, c, etc.) are true' and 'it is true that all the statements on this page are true' are logically equivalent in spite of their being on different logical levels. In other words, if one is true the other must be true, and if one is false the other must be false. Thus, in a sense, they are
merely alternative ways of saying the same thing (i.e. of asserting a, b, c, etc.). Therefore, to treat the statement 'all the statements on this page are true' as self-referring is, in effect, to say the same thing twice, on different levels, in the course of a single statement, which is not only logically odd but pointless.

The same objection cannot, however, be made against the statement 'no statement can be proved' when interpreted self-referringly, because saying that a, b, c, etc. cannot be proved is in no sense logically equivalent to the assertion 'it cannot be proved that a, b, c, etc. cannot be proved'. Whereas the former is a statement about a, b, c, etc. the latter clearly is not. Thus in the latter case, one cannot be accused of confusing two different levels of talk about statements, as opposed to merely talking about different levels of statement.

Another incidentally self-referring, but perfectly meaningful assertion, is the statement 'no language of any kind exists', and from our point of view a far more important one. The statement is self-referring, of course, because it directly entails the proposition 'no assertion can be made about anything'; i.e. for any x it is not the case that x is an assertion. In this way, the statement says something about a whole class of x's of which it is itself a member; namely,
that the class has no members. In other words, it is incidentally self-referring.

On the other hand, the statement is clearly meaningful because one can perfectly well conceive of a situation in which no language existed. Indeed, there was presumably a time when such a situation did exist. Nor is there any confusion of logical levels here, as there is in the case of some incidentally self-referring statements. Moreover, not only is the statement self-referring, it is also self-guaranteeing, for there is no way in which it can be meaningfully denied, since it states a necessary condition of discourse in general.

Although statements like 'no language of any kind exists' and 'assertions are never made' are self-referring, their contradictories are not self-referring, or at least, not in any obvious sense. The reason for this is that in denying a necessary condition of discourse one is committed to saying that the class of all statements to which one's denial belongs is empty, and that therefore one's denial is not a statement of any kind. In affirming a condition of discourse, on the other hand, one is merely committed to saying that the class to which one's statement, qua statement, belongs is not empty, but this does not in itself imply that the 'statement' is genuine; i.e. a member of the class. In other words, the
former statement is incidentally self-referring because it says something about all would-be statements and therefore about itself, whereas the latter statement does not refer to the whole class, but only to some unspecified members of it.

It is perhaps worth pointing out that even if one accepts Ross's dictum that a self-referring statement is meaningless when 'one tries to express in a sentence a meaning which refers to the meaning of the same sentence' one is not forced to conclude that statements like 'no language of any kind exists', 'no assertion can be made about anything', and 'no sentence conveys anything' are meaningless. For it could be argued that these statements cannot possibly refer to the meaning of other statements, since they deny the existence of all statements, and, indeed, of discourse in general. It could be said, therefore, that in none of these cases is one really referring to the actual meaning of the statement, but rather to the sequence of sounds, or marks on the page, or whatever it is that is purporting to be a statement. For example, what the putative sentence 'no assertions can be made about anything' says about itself is that it cannot itself be characterized as an assertion.

On the other hand, when a statement like 'no statement can be proved' is interpreted self-referringly, it clearly does refer (amongst other things) to its own meaning, for it
entails that what the sentence 'no statement can be proved' says is itself unproveable.


In future, I will describe self-referring statements which are logically true as self-guaranteeing, and self-referring statements which are logically false as self-refuting.

There is, of course, an indefinitely large class of self-referring statements of which the truth or falsity can only be determined empirically, and these are neither self-refuting nor self-guaranteeing. For instance, the wholly self-referring statement 'all the words which are used in this statement occur in the Oxford English Dictionary' comes into this class; if true it is empirically true, and if false it is empirically false. For although it happens to be true, one can perfectly well imagine a situation in which one of the words had been omitted as the result of some extraordinary oversight.

Another point about this example is that the same proposition could equally well be expressed in terms of a non-self-referring expression; e.g. 'the following words can be found in the Oxford English Dictionary ...'. In fact, this is likely to be true of almost all the empirically self-referring statements one thinks of, although it is possible to think of a few exceptions; e.g. 'most people would be puzzled by this statement'. This could not be re-expressed in a non-self-
referring way because the statement is puzzling, let us suppose, just because it is self-referring. On the other hand, its truth or falsity is undoubtedly an empirical matter.

It is also worth noting that these examples, especially the former, provide further evidence against the theory that all self-referring statements are meaningless. For although it is just possible to argue that the statement 'most people would be puzzled by this statement' is meaningless, on the grounds that the referring expression 'this statement' cannot logically refer to the very statement in which it occurs, it could not possibly be argued that the statement 'all the words which are used in this statement occur in the Oxford English Dictionary' is meaningless. For, as we have seen, it can be re-expressed in terms of a non-self-referring statement without loss of meaning. And if the non-self-referring statement is true, then the self-referring statement must also be true, and therefore not meaningless.

This is further proof, if any were needed, that the theory that all types of self-referring statement are meaningless is in need of modification, and also that we cannot classify a statement as meaningless solely on the grounds that it contains a referring expression like 'this statement' which is supposed to refer to the very statement in which it occurs.
Incidently self-referring statements may also be divided into those which are empirically self-referring and those which are logically self-referring. An example of an empirically self-referring statement would be 'all type-written statements are legible'. This is empirically self-referring because the statement itself just happens to have been typed, although it might easily have been written in some other way.

An example of a logically self-referring statement would be 'all the statements on this page are type-written'. It is logically self-referring because 'this page' is meant to refer to the page on which the statement is written. On the other hand, its truth or falsity can only be determined empirically and therefore it is neither self-guaranteeing nor self-refuting.

However, the truth or falsity of some incidentally self-referring statements can be determined logically; e.g. 'all generalizations are meaningless' and 'I sometimes say things which are false'. The former is self-refuting because a necessary condition of its truth is that it be meaningless, and if it is meaningless it cannot be true. The latter is self-guaranteeing because if it is true it is true, and if it is false it is true.

There is another type of incidentally self-referring statement which is more difficult to classify; namely, one
which is empirically self-referring, but logically self-refuting if its empirical self-reference is taken for granted; e.g. 'all type-written sentences are meaningless'. If the empirical self-reference of this last statement is taken for granted (i.e. that it is itself type-written) it can be regarded as logically self-refuting. In order for it to be true it must be meaningless, but meaningless utterances can be neither true nor false, therefore it can only be true by being untrue. In other words, if the second-order statement 'the statement that all type-written sentences are meaningless, is typed' is true, the first-order statement 'all type-written sentences are meaningless' must necessarily be false. Statements like this which are empirically false from one point of view, but self-refuting from another, may be termed pragmatically self-refuting.

Another example of a pragmatically self-refuting statement is the spoken statement 'I cannot speak'. It is empirically self-referring in much the same sense as the statement 'all type-written sentences are meaningless' is self-referring. In both cases the statement is about a whole class of statements or utterances of which it is itself a member. Thus, in more formal language, the statement 'I cannot talk' says that the class of my talkings is empty. But it just so happens that 'I cannot talk' is itself a member of the class
of my talkings, although it could just as easily have been a member of the class of my writings. Therefore, as a matter of empirical fact, the class is not empty, and hence the statement as a whole is empirically false. On the other hand, if its empirical self-reference is taken for granted, it can be regarded as logically self-refuting.

It cannot, of course, be argued that such statements are meaningless, because on the empirical level they describe a perfectly conceivable state of affairs. Thus one has no difficulty in imagining a situation in which all type-written sentences happened to be meaningless, or in which a given individual (myself) was unable to talk.

The truth or falsity of some reflexively self-referring statements may also be determined empirically; e.g. 'you know that all my assertions are true', 'I sometimes write that I cannot write', 'I sometimes believe that I have no beliefs'. Although none of these examples is directly self-refuting or self-guaranteeing, the logical inconsistency which exists between the first-order and second-order statements in the last two examples makes them indirectly self-refuting. This will be explained later.

Other reflexively self-referring statements are straightforwardly self-refuting; e.g. 'it can be proved that nothing can be proved' and 'I know that I know nothing',
whereas others are self-guaranteeing (in the sense of being both self-referring and logically true); e.g. 'it cannot be proved that nothing can be proved' and 'I do not know that I know nothing'.

Section 5. Passmore's Classification.

In chapter four of Philosophical Reasoning, Professor Passmore maintains that self-refuting statements can be classified into three main types: pragmatic, ad hominem, and absolute.

A statement is pragmatically self-refuting if the assertion which is being made conflicts with the way in which it is made. For example, if I say (aloud) that I cannot speak, or write that I cannot write, or shout that I am not shouting, or claim that I cannot pronounce a particular way whilst actually pronouncing it, then I am guilty of a pragmatic self-refutation. In each case, the way in which the speaker makes the assertion is 'the best possible counter-example to what he asserts to be the case'.

But, since it is an empirical fact that I make the assertion in the way I do, it cannot be claimed that I am contradicting myself.

Furthermore, 'it is always possible for the person accused of a pragmatic self-refutation to deny that the alleged

---

1 See Philosophical Reasoning, Duckworth 1961 (paperback), Ch.4, p.62. 2 ibid.
counter-example is a counter-example'. For instance, a person accused of saying that he cannot speak may claim that he was actually miming the words, whilst playing a track on a hidden tape-recorder. Such an attempt to avoid the charge may be far-fetched, or even ludicrous, but it is not logically absurd.

This account of pragmatic self-refutation is entirely consistent with my own account in section four. My main point was that a pragmatically self-refuting statement entails its own falsity when its self-reference is taken for granted, although the fact that it is self-referring must be demonstrated empirically. And this still stands.

An ad hominem self-refutation is one in which a person’s admission that he has made a particular assertion in a particular way is logically equivalent to an admission that the assertion is in fact false; e.g. 'I sometimes say that I cannot speak', 'I sometimes think that I cannot think'. These statements are not directly self-refuting since it might well be true that I sometimes say that I cannot speak or that I sometimes think that I cannot think, although I cannot then coherently maintain that I am in fact unable to speak, or that I am in fact unable to think. I shall have more to say about this indirect form of self-refutation later on.

---

1See *Philosophical Reasoning*, p.62.
An absolutely self-refuting proposition is a proposition whose falsity is necessarily presupposed in any assertion of its truth. In other words, it is a proposition 'p' such that to assert p is tantamount to asserting both p and not-p. Two fairly obvious examples of this would be: 'no sentences are intelligible' and 'there are no truths'. Thus, one cannot assert that no sentence is intelligible without taking for granted the intelligibility of that very assertion. And similarly, 'to assert that there are no truths is to assert both that there are and that there are not truths because, precisely, to assert is to assert to be true'.¹ I will be examining this type of self-refutation in some detail later on.

Section 6. Mackie's classification.

J.L. Mackie's classification of self-refuting statements, as set down in his article on self-refutation in the Philosophical Quarterly for July 1964, differs in at least two important respects from Passmore's. Firstly, he recognizes two distinct types of absolutely self-refuting statement, and secondly he claims to have discovered a form of self-refuting statement, not included in Passmore's list, which he terms 'operational self-refutation'. His class of pragmatically self-refuting statements is the same as Passmore's, however. In this article he attempts to give a strictly formal analysis

¹See Philosophical Reasoning, p.68.
of these four allegedly different modes of self-refutation.

Although, in this type of enquiry, formalization is
sometimes useful as a methodological instrument, its usefulness
as a method of exposition and clarification is rather limited.
In expounding Mackie's views, therefore, I shall use a less
formal, but I hope more intelligible, method of presentation
than the one which he uses. Accordingly, I shall employ only
two symbols 'p' and 'd', where p stands for a proposition of
some kind, and d for a proposition-forming operator upon
propositions; e.g. 'I think that', 'Smith believes that', 'it
is true that' and so on.

According to Mackie, one type of absolute self-refutation
occurs when the operator 'd' is 'truth-entailing' and the
proposition p is logically inconsistent with any proposition
which is formed by that operator. Truth-entailing operators
are those for which 'if dp is true, p itself must be true also'.
These include 'it is true that', 'it can be proved that', 'I
know that', 'it is certain that'. Thus, to say that it can be
proved that nothing can be proved is to imply that the
statement 'nothing can be proved' is in fact true, but this
statement in turn entails that any statement of the form 'it
can be proved that p' is in fact false, and hence that the
statement 'it can be proved that nothing can be proved' is in

\[^{1}\] J.L. Mackie, 'Self-Refutation - a Formal Analysis',
Philosophical Quarterly, July 1964, p.194.
fact false. In this way, the proposition 'it can be proved that nothing can be proved' entails its own falsity. In other words, if dp entails that p is true, and p entails that dp is false, then dp is absolutely self-refuting.

Such propositions as 'it is true that nothing is true', 'I know that I know nothing (literally)', 'it is certain that nothing is certain' are all self-refuting in the above sense.

According to Mackie, the other type of self-refutation occurs when the operator 'd' is 'strictly prefixable' and the proposition 'p' is logically inconsistent with any proposition formed by that operator. Strictly prefixable operators are those for which 'if p itself is true, dp must also be true'.

The only two examples which Mackie can think of are: 'it is true that' and 'it is possible that' in at least one sense of possible. The former is strictly prefixable because to assert anything just is to assert it to be true. The latter is prefixable because to say that something is true is, at the same time, to imply that it is possible. For this reason, the statements 'it is not the case that something is possible' and 'there are no truths' are absolutely self-refuting. They are self-refuting because to assert them is tantamount to asserting both dp and not-dp. For example, if nothing is possible then it follows both that nothing is possible and that it is not possible that nothing is possible.
Mackie maintains\textsuperscript{1} that an absolute self-refutation can only arise when the operator is either truth-entailing or strictly prefixable. However, he has overlooked another class of operators which can also give rise to self-refutations of this kind. These may be termed 'falsity-entailing' and include such propositions as 'it is logically impossible that' and 'it is false that'. Thus, the following statements are absolutely self-refuting: 'it is logically impossible for any proposition or state of affairs to be logically impossible' and 'it is false that at least one proposition is false'.

In both these examples, \( dp \) entails the falsity of \( p \), and \( p \) entails the falsity of any proposition formed by that particular operator for which \( d \) stands. In other words, if \( dp \) is true then \( p \) is false, and if \( p \) is false then \( dp \) is false. Hence \( dp \) requires its own falsity and is therefore absolutely self-refuting.

In all Mackie's examples of absolutely self-refuting statements, the operator \( d \) also occurs as part of the proposition \( p \). This is not of course an essential feature of such statements, since one operator very often entails another, thereby making possible a substitution which does not alter the self-refuting character of the statement. For example, to say that something is known is to imply that it is certain.

\textsuperscript{1}See Mackie's article in the \textit{Philosophical Quarterly}, July 1964, p.194, from 'This pragmatic self-refutation gives rise ...'. 
and to say that something can be proved is to imply that it is true. Thus 'it is known that nothing is certain' and 'it can be proved that nothing is true' are both absolutely self-refuting.

According to Mackie, it is important not to confuse these two types of absolute self-refutation with what he terms 'operational self-refutation'. A statement is operationally self-refuting if the operator $d$ is weakly prefixable and the proposition $p$ is logically inconsistent with any statement formed by that operator. Mackie fails to come up with a precise definition of 'weakly prefixable', but cites as examples the expressions 'I know that', 'I believe that' and 'I think that'. 'I know' is weakly prefixable because if someone puts forward any factual claim or makes any assertion, say that $p$, then he is implicitly committing himself also to making the claim that he knows that $p$.

Thus if I assert that I know nothing, I am implicitly committing myself to the claim that I know that I know nothing, and hence to a denial of what I originally asserted. From this it would follow that 'I know nothing' cannot be coherently asserted. Yet it may well be true, in that someone else may be able to say truthfully of me, 'he knows nothing'.

Similarly, 'I believe' and 'I think' are weakly prefixable because anyone who asserts that $p$ is implicitly committing

---

himself also to asserting that he believes or thinks that $p$
(in the sense in which 'I believe' or 'I think' is entailed by,
and does not exclude 'I know').

The question now arises whether Mackie's analysis of
operational self-refutation should be accepted. For if one
wished to attack it, one could do so in a number of ways.
For example, one could dispute his reasons for claiming that
the examples he gives us are operationally self-refuting, whilst
admitting that other reasons could be found for accepting the
claim. Or one might argue that his examples are not self-
refuting at all, or else that they are self-refuting in some
other sense, whilst still accepting the view that operationally
self-refuting statements actually exist. Or again, one might
want to argue that operationally self-refuting statements,
as Mackie conceived them, could not exist.

Section 7. Bonney on Weak Prefixability.

In the Philosophical Quarterly for October 1966, W.L.
Bonney has attacked Mackie's arguments in support of the
theory that 'I believe nothing' is operationally self-refuting
in the sense that, although it could well be true, it cannot
be coherently asserted. This theory is based on the claim
that 'anyone who asserts that $p$ is implicitly committing him-
self also to asserting that he believes that $p$' and that
therefore 'anyone who asserts that he believes nothing
implicitly commits himself to asserting that he believes that he believes nothing, and hence to denying his original assertion'.

Mackie argues that the truth of this claim can be demonstrated by reference to two principles: firstly, that "I assert that p' in the ordinary sense of assert entails 'I believe that p'", and secondly, 'if x asserts that p, then x is committed to asserting anything entailed by 'x asserts that p'".

As Bonney has pointed out, however, the second principle is open to the objection that, on any ordinary interpretation, 'I assert that' means the same as 'I make the factual claim that' and since lying consists in making a factual claim which one does not actually believe, it follows that 'I assert' cannot be held to entail 'I believe'. Thus, it could be argued that 'I believe nothing' is not self-refuting if in fact I do not really believe that I believe nothing, for it is quite possible that what I intend to be a lie is in fact true.

Bonney further argues that even if Mackie is using 'assert' to refer only to truthful claims, it still does not follow that 'I believe nothing' is self-refuting. For, in order to establish the more limited thesis that all truthful claims to believe nothing are self-refuting, it is necessary
to show that anyone who says that p (i.e. makes the factual claim that p) and also believes that p is committed to asserting that he believes that p. But clearly 'the fact that the speaker believes that p is no ground whatever for maintaining that he is committed to asserting that he believes that p. One might just as well argue that a person who says that p and does not believe it is committed to asserting that he does not believe it'.

Similar objections can be raised against any argument purporting to show that 'I know nothing' and 'I am not thinking' are operationally self-refuting on the grounds that if I assert that p then I am committed to asserting anything entailed by 'I assert that p' and that 'I assert that p' entails both 'I know that p' and 'I think that p'.

In the first place, one can lyingly assert something to be the case and therefore not know or think it to be true. In the second place, my knowing or thinking that p in no way commits me to asserting that I know or think that p whenever I say that p. In other words, in asserting something to be the case I do not thereby commit myself to saying that I assert (in the above narrowly defined sense) it to be the case.

The above arguments show, I think, that Mackie's reasons for saying that certain statements are operationally self-

---

refuting are invalid. It does not of course follow from this
that his conclusions are wrong. Bonney indeed claims that
Mackie is correct in saying that 'I believe nothing' is
operationally self-refuting, although he disagrees with the
arguments put forward in support of this view. He does not,
however, put forward any arguments of his own, although I
believe this to be a fairly simple task.

One might, for example, prove that 'I believe' is
prefixable for all assertions in the following manner: to
say that 'I believe that p' is, amongst other things, to
suggest that p is in fact true. I cannot therefore
consistently assert that p and at the same time say that I
do not believe it, since this would be to claim that p was
true whilst suggesting that it was false. Nor can I assert
that p and at the same time say that I neither believe nor
disbelieve it, since this would be to claim that p was true
whilst refusing to suggest that it was true (or false). Such
assertions are incoherent because it is I (the speaker
referring to himself) who makes them, but if someone else
said of me that I disbelieve, or neither believe nor
disbelieve, a proposition which is in fact true, he could
well be right. However, I cannot make such a claim myself,
because in making it I am forced to withdraw it.

*I believe that' does not of course exclude 'I know that'
in this context.
In other words, I cannot assert something and, at the same time, admit that I don't really believe it. For my assertion would then cease to be an assertion, as opposed to the mere utterance of a proposition. It follows from this that in making any assertion I implicitly commit myself to the only other alternative; namely, to saying that I do in fact believe it. In this sense, 'I believe that' is prefixable for all assertions.

If the above argument is correct, it follows that when a person claims to have no beliefs he in effect commits himself to saying that he believes that he has no beliefs, and therefore to a proposition whose truth entails the falsity of that assertion.

Section 8. Possible ways of classifying 'I have no beliefs'.

The statement 'I believe that I have no beliefs' is an example of what Passmore has termed an 'ad hominem self-refutation'; that is to say, the speaker's claim that he believes that he has no beliefs may be perfectly correct, since it is quite possible to hold inconsistent beliefs, but nevertheless, in a logical sense, it amounts to an admission that the belief is in fact false.

In this type of self-refutation, the second-order statement is logically inconsistent with the first-order

*N.B. 'Implicitly committed' means here 'logically prevented from denying the truth of'.*
statement, but this provides us with no means of knowing which statement is false and which true. For example, it is possible for me to make an assertion which I neither believe nor disbelieve, and therefore it is possible, although extremely unlikely, that when I claim to believe that I have no beliefs I do in fact have no beliefs; that is to say, my claim to believe that I have no beliefs may be false, in which case, it may still be true that I have no beliefs.

Thus, even if it can be shown that 'I believe that' is prefixable for all assertions, it does not follow that 'I believe nothing' cannot be coherently or truly asserted, but only that it cannot be truthfully asserted. Thus, we may always say to someone who claims to have no beliefs 'What you say may be true, but if it is then you at least cannot believe it'. Another point worth making is that the contradictory statement 'I have some beliefs' is not self-guaranteeing in the way that a statement like 'I sometimes make assertions' is. But more of this later on.

Although the statement 'I sometimes believe that I have no beliefs' is a form of ad hominem self-refutation, the fact that the operator 'I believe that' is prefixable for all assertions means that it differs in at least one important respect from such statements as 'I sometimes write that I cannot write' and 'I sometimes say that I cannot talk', which
are also forms of ad hominem self-refutation. For whereas in these last two examples it is possible to assert the first-order statement 'p' by denying the second-order statement 'dp', this option is not open in the case of a statement like 'I believe that I have no beliefs' in which to assert p is to commit oneself also to asserting dp. Once again, it should be noted that 'I have no beliefs' is not directly self-refuting since it may be true as long as the speaker does not himself believe it to be true. For this reason, I shall not describe it as operationally self-refuting as Mackie does.

In order to distinguish the two modes of indirect self-refutation I shall use the expression 'standard ad hominem self-refutation' in referring to the former and 'absolute ad hominem self-refutation in referring to the latter. It should be remembered, however, that the statement 'I have no beliefs' only comes into this latter category if 'I believe that' is prefixable for all assertions. The same of course would be true of a statement like 'I am not thinking' which could be shown to be indirectly self-refuting on similar grounds.

Even if the prefixability argument is not accepted, it is still possible to show that 'I believe that' and 'I am not thinking' are self-refuting, although in a much weaker sense. Thus one could argue that 'I believe nothing' entails the proposition 'I cannot sincerely say of any proposition that it is true (or false), or even that it seems to me to be true
(or false). On the speaker's own admission, therefore, the proposition is not being seriously asserted. Consequently, it does not deserve to be taken seriously. Of course, the statement is not self-refuting in the strict sense, since even a non-serious assertion may be true. However, it does have a tendency to undermine itself, and therefore to weaken our faith in it. For this reason, I prefer to characterize such statements as self-undermining rather than self-refuting. Other examples of this would be 'nothing is certain' and 'generalizations are invariably misleading and should be avoided'.

Apart from the prefixability argument there is another possible way of showing that 'I believe nothing' is self-refuting in the strict sense. This is by arguing that in order to make any assertion, I am forced to take for granted the truth of certain propositions (i.e. presuppositions of discourse); for example, that not all words are completely meaningless, and that at least some statements are true. But since this 'taking for granted the truth of certain propositions' is a kind of belief, it could be argued that 'I believe nothing' is logically incoherent. In other words, in order to make any assertion at all, I am forced to assume that not only do I have certain beliefs, but that at least some of those beliefs are true. However, this type of argument brings us very close to the whole notion of a transcendental deduction, and therefore I
prefer to postpone any further discussion of it at this stage.

Section 9. 'I know that ...'.

Mackie cites as another possible example of an operationally self-refuting statement, the statement 'I know nothing' on the grounds that it could be argued that 'I know that' is weakly prefixable for all assertions. This time, however, he does not attempt to adduce any arguments of his own in support of this contention. Furthermore, it seems to imply what is quite obviously false, namely that all assertions are knowledge claims.

If all assertions were knowledge claims, we would never be justified in making any assertion at all unless (a) we were certain of the truth of what we were saying, and (b) the evidence for it was overwhelming. But since we frequently do make assertions in defiance of both these conditions, it is quite clear that we do not wish to imply that we know whatever we assert to be the case. Thus the main purpose of such common expressions as 'I think', 'I believe', 'I suppose', 'I guess' is to enable us to make claims in a hesitant and undogmatic way.

It is true of course that when such expressions are omitted, the proposition would normally be interpreted as a knowledge claim, and that for this reason it would sound odd to make a categorical assertion and then to add that one did
not really know if it was in fact true; for example, to say 'Smith will leave town tomorrow, although I don't really know if he will'. But it certainly does not follow from this that 'I know that' is prefixable for all assertions in the sense in which Mackie says it is. In other words, we are not committed to saying that we know whatever we assert to be the case (in the sense of having the right to be certain of its truth), although we are perhaps committed to saying that we at least believe whatever we assert to be the case. But if 'I know nothing' does not belong to the class of operationally self-refuting statements, it does perhaps belong to the class of self-undermining statements.

It could of course be argued that 'I know that I know nothing' is self-refuting, but, once again, it is not operationally self-refuting, as I shall now try to show.

Section 10. Self-cancelling utterances.

It has been argued by Austin\(^1\) that 'I know that' is much closer to being a performative utterance, than a descriptive expression. There is clearly an element of truth in this, since 'I know that' is often used in such a way that there is virtually no difference between saying 'I know that x' and saying 'I vouch for the truth of x'; 'I vouch for' being an obvious example of a performative utterance.

If 'I know that' was to be interpreted as a performative expression then 'I know that I know nothing' would not be a proposition at all, and therefore could not possibly be self-refuting. It would mean something like 'I vouch for the truth of the proposition that I am never in a position to vouch for the truth of any proposition', and so would closely resemble such declarations as 'I promise to break all my promises' and 'I guarantee that all my guarantees are worthless'. In these cases, an undertaking is given in such a way that it is at the same time either broken or withdrawn. Such declarations may be termed 'self-cancelling utterances'.

Even if it is possible to interpret 'I know that' as a performative utterance in the majority of cases in which it occurs (although I do not believe it is) it does not of course follow that it can never be interpreted in any other way. Thus there are cases where a statement of the form 'I know that $x$' is clearly not intended to guarantee the truth of anything, but rather to provide information about the speaker. For example, a schoolboy who tells his teacher that he knows that Henry VIII was a powerful king who had many wives, on being asked whether he knows anything about the Tudors, is surely not vouching for the truth of those facts, but is merely providing his teacher with information about his state of knowledge on a particular subject.
Thus, if the statement 'I know that I know nothing' is interpreted as a statement about the speaker, as in the example above, it is self-refuting, but not operationally self-refuting. It is not operationally self-refuting in Mackie's sense, because it cannot be coherently asserted by anyone, let alone by the person making the assertion; that is to say, 'he knows that he knows nothing' is self-refuting in precisely the same way that 'I know that I know nothing' is.

In fact, both statements are self-contradictory and therefore necessarily false. They can be described as self-refuting only because they are reflexively self-referring. In future, I shall describe such statements as absolutely self-refuting. *For Mackie, the term has a somewhat different connotation. For he uses it to refer to any self-refuting statement which is necessarily (i.e. unavoidably) self-refuting; that is to say, one which cannot be coherently asserted by anyone, but I shall argue that some statements which come into this category should be regarded as operationally self-refuting.

Section 11. Operationally self-refuting statements.

We may conclude from the foregoing discussion that none of Mackie's examples are operationally self-refuting in the sense in which he thinks they are. According to him, a proposition is operationally self-refuting if it cannot be coherently asserted by the person to whom it refers, but
could be coherently asserted by someone else. But, as we have seen, the statement 'I have no beliefs' only commits the speaker to a kind of ad hominem self-refutation, in which either the first-order statement or the second-order statement may be true, but not both. Similarly, the statement 'I never think', when interpreted as a self-refuting statement, commits the speaker to an ad hominem self-refutation of the same kind.

The statement 'I know nothing', which is the only other example that Mackie gives us, is not operationally self-refuting because 'I know that' is not prefixable for all assertions. At the most, it is self-undermining. Nor can the statement 'I know that I know nothing' be interpreted as operationally self-refuting, for the simple reason that it is necessarily false and therefore cannot be coherently asserted by anyone; that is to say, it is a type of absolutely self-refuting statement.

Even though all Mackie's examples of operational self-refutation are really self-refuting in some other sense, it does not of course follow that operationally self-refuting statements, as he conceives them, could not possibly exist, nor even that they do not in fact exist.

It is only necessary to give examples of such statements to show that they do, in fact, exist. One such example would be 'I never make assertions', since the proposition 'x truly asserts that he never makes assertions' is necessarily false. On the
other hand, there is nothing logically incoherent in someone else's asserting of x that he never makes assertions. In other words, 'I never make assertions' could not possibly be a true assertion, although 'he never makes assertions' said of the same person, could well be true. For whether or not a person ever actually says anything is obviously a contingent matter. The statement 'I never make assertions' is therefore operationally and not absolutely self-refuting. Another example would be 'I never tell the truth'. Again, it cannot be coherently asserted by the person to whom it refers, but it could be coherently asserted of that person by someone else.

It is, I think, fairly clear from this discussion that it is a mistake to attempt to give an account of operational self-refutation solely in terms of the concept of weak prefixability, as Mackie does. For, as we have seen, even if 'I believe that' is weakly prefixable for all assertions, in the sense that if we deny that we believe that what we have asserted is true, we are in effect withdrawing our assertion, it does not follow that a statement like 'I have no beliefs' cannot be coherently asserted, but only that it cannot be sincerely asserted.

Fortunately, it is quite possible to give an account of operational self-refutation without making use of the somewhat dubious notion of weak prefixability. For instance, it is not
necessary to show that 'I assert that' is weakly prefixable for all assertions in order to show that the statement 'I never make assertions' is operationally self-refuting. Nor is it necessary to show that some expressions like 'I am telling the truth when I say that' is weakly prefixable for all assertions in order to show that 'I never tell the truth' is operationally self-refuting.

In the first case, one can easily demonstrate the incoherence of the assertion 'I never make assertions' by pointing out that it necessarily refers to the speaker himself. This means of course that in order for the assertion to be true (and considered only as a proposition, it could well be true) the meta-statement 'x truly asserts that he never makes assertions' must also be true. The meta-statement, however, is absolutely self-refuting, and therefore cannot possibly be true.

Similarly, the person who says 'I never tell the truth' is necessarily talking about himself, and therefore to say the assertion is true (and considered only as a proposition, it could well be true) is to claim that his assertion that he never tells the truth is true. Obviously any such claim would be absolutely self-refuting.

An operationally self-refuting statement is, therefore, one in which the assertion that a perfectly conceivable state
of affairs exists is sufficient to show that it does not exist. Mackie would argue that whereas such a statement cannot be coherently asserted by the person to whom it refers, it could be coherently asserted by someone else.

This is certainly true of statements like 'I never make assertions' and 'I never tell the truth'. On the other hand, it is not true of statements like 'nobody ever makes assertions' and 'nobody ever tells the truth'. For although they describe perfectly conceivable states of affairs, they cannot be coherently asserted by anyone. Mackie seems to regard such statements as absolutely self-refuting, but this, I believe, is a mistake. In fact, it would be far better to treat them as operationally self-refuting, as I shall now try to explain.

Section 12. Two forms of operational self-refutation.

According to Mackie, it is possible to distinguish at least two kinds of absolutely self-refuting statement: one occurs when the operator 'd' is truth-entailing and the proposition 'p' is logically inconsistent with any operator which is formed by that operator; e.g. 'it can be proved that nothing can be proved' and 'it is certain that nothing is certain'. The other occurs when the operator 'd' is strictly prefixable and the proposition 'p' is logically inconsistent with any proposition which is formed by that operator; e.g. 'there are no (absolute) truths', and 'nothing is possible'.
According to Mackie, absolutely self-refuting statements of both kinds are necessarily false, whilst their contradictionaries are necessarily true. He is therefore obliged to show that not only are reflexively self-referring statements like 'it cannot be proved that nothing can be proved' and 'it is certain that nothing is certain' necessarily true, but more controversially, that incidentally self-referring statements like 'there are some truths' and 'something is possible' are also necessarily true.

It is at this point that Mackie's classification diverges most sharply from Passmore's. For Passmore's class of absolutely self-refuting statements include statements whose contradictionary is in no sense necessarily true, but which nevertheless cannot be coherently denied. For example, the contradicitories of such statements as 'no (putative) sentence is intelligible' and 'no language of any kind exists' come into this category. Thus, one can perfectly well conceive of a situation in which all putative sentences were unintelligible, or in which no language of any kind existed, but nevertheless one cannot coherently assert that such a situation does exist, because in doing so one is forced to assume that it does not exist.

It could be said that the statement 'no (putative) sentence is intelligible' is self-refuting because the fact that some
sentences are intelligible is a necessary condition of statement-making in general. And the same can be said of the existence of a language of some kind. In fact, Passmore maintains that:

Only if a philosophical argument can show ... that a sentence can propose nothing - because what it asserts is inconsistent with the presuppositions of all proposing - is it pointing ... to an absolute self-refutation.  

Hackie, on the other hand, argues that:

In general, what could be called a condition of discourse sets limits only to what can be coherently asserted, it is concerned with the way in which saying one thing implicitly commits us to being prepared to say something else as well, and it therefore gives rise only to what we have called operational self-refutation.  

Clearly, Passmore's suggestion that all absolutely self-refuting statements are inconsistent with the conditions of discourse is unacceptable, at least in so far as reflexively self-referring statements are concerned. For statements like 'it can be proved that nothing can be proved' and 'I know that I know nothing' are in no sense inconsistent with the conditions

of discourse, although they are absolutely self-refuting.* The possibility remains, however, that the theory is true of all incidentally self-referring statements in this category. But if this is the case, then either Mackie's second class of absolutely self-refuting statements does not exist, or else it must include those statements which Passmore regarded as absolutely self-refuting.

Mackie admits that it is difficult to find many propositions which belong to his second class of absolutely self-refuting statements, and indeed can himself only think of two examples: namely, the incidentally self-referring statements 'there are no truths' and 'nothing is possible'. Passmore also regards the statement 'there are no truths' as absolutely self-refuting, and would perhaps agree that 'nothing is possible' could be classified in this way as well. However, they would disagree fundamentally in their reasons for saying this.

Thus, whereas Mackie would account for the absolutely self-refuting character of these statements in terms of the strict prefixability of the operators 'it is true that' and 'it is possible that', Passmore would account for it in terms of the necessary conditions of discourse. This means that

*It could of course be argued that since such statements are straightforwardly self-contradictory, there is not much point in classifying them as self-refuting. But, as I have already explained, I do not share this view. See p.123 of this thesis.
on Mackie's view, the propositions expressed by these statements are necessarily false, whereas on Passmore's view they are merely contingently false.

Mackie's case for saying that 'there are no truths' is absolutely self-refuting in the sense that it is necessarily false, whilst its contradictory 'there are some truths' is necessarily true, can be put in the following way: From the fact that 'it is true that' is strictly prefixable for all assertions, it follows that 'there are no truths' is logically equivalent to the reflexively self-referring statement 'it is true that there are no truths'. It therefore follows that 'there are no truths' is necessarily false, in which case 'there are some truths' must be necessarily true.

Or again, one could argue that there must be either some truths or no truths. If there are some truths, then it is true that there are some truths. If there are no truths, then it is true that there are no truths, and hence that there is at least one truth. Therefore 'there are some truths' is necessarily true.

The chief objection to these arguments is that there can be no truth without a language of some kind; that the concept of truth is inextricably bound up with the way in which we conceptualize our experience. In other words, we can talk of true statements, but not of independently existing
truths. No example of an objective, unconceptualized truth could ever be given, for it could only be identified by means of concepts, in which case it would come to us already conceptualized. And this at least seems to show that no clear meaning can be attached to the concept of such an entity.

It might therefore be argued that the existence of a language of some kind is a necessary condition of the truth of the proposition 'some truths exist'. But if this is the case, then it is only contingently true that some truths exist, since it is undoubtedly only contingently true that a language exists.

Thus, the reason why the proposition 'there are no truths' cannot be coherently asserted is not, as Mackie claims, because it is necessarily false, but because its falsity is presupposed in any assertion of its truth. Nevertheless, it makes perfect sense to say 'there might have been no truths' because precisely that situation would have obtained if there had been no language.

Of course, if there had been no truths, it would not have been true that there were no truths. But it does not follow from this that it would have been false that there were no truths. Indeed, it would have been neither true nor false, because within a situation of that kind there would be no concept to apply. Thus, although it does not make sense to say 'it might have been true that there were no truths' it does make
sense to say 'it is true that there might have been no truths'.

The apparent paradox is resolved once we realize that when we conceive of a situation in which there are no truths, we are conceiving of a situation which, if it existed, would be inconceivable; i.e. unconceptualizable. Obviously, we can do this because we do in fact have a conceptual system (i.e. a language). For if we have a language we can quite easily conceive of its non-existence.

It has now been shown, I think, that a good case can be made out for classifying 'there are no truths' with statements like 'no language of any kind exists' and 'all (putative) sentences are unintelligible'; that is to say, with statements which Mackie would classify as operationally self-refuting.

Moreover, the statement 'nothing is possible' is in precisely the same situation with respect to this argument. This is because the only reason for saying that the operator 'it is possible that' is strictly prefixable for all assertions is that whenever we make any assertion we are claiming that some proposition or other is true, and therefore we must also be claiming that it is both logically and empirically possible for that proposition to be true. In other words, 'it is true that' entails 'it is possible that'. Thus if it was true that no language existed then it would be logically and empirically possible for it not to exist. But just as the concept of
truth would not be applicable within such a situation, the
cognitive concept of possibility would not be applicable either. Hence
if 'there are some truths' is contingently true for the reasons
cited, then 'something is possible' must be contingently true
for the same reasons.

Thus, it very much looks as though 'there are no truths'
and 'nothing is possible' are not absolutely self-refuting in
the sense in which Mackie thinks they are. Indeed, one might
be tempted to go further and claim that no statements are
absolutely self-refuting in this second sense, for Mackie is
unable to think of any other examples. However, we can
scarcely hope to come to a definite conclusion on this matter,
since, as we have seen, it raises one of the most difficult
and intractable problems in the history of philosophy, namely
the relationship between truth, language, and reality.

If we do take the view that no statement is absolutely
self-refuting in Mackie's second sense, whilst agreeing with
him that only necessarily false propositions should be included
in this category, then we are bound to classify such
incidentally self-referring statements as 'there are no truths'
and 'nothing is possible' as operationally self-refuting, and
to classify only such reflexively self-referring statements
as 'it can be proved that nothing can be proved' and 'I know
that I know nothing', and 'it is true that nothing is true' as
absolutely self-refuting.
According to Mackie, a statement is operationally self-refuting if the operator 'd' is weakly prefixable and the proposition 'p' is logically inconsistent with any statement formed by that operator. The only examples he gives are 'I have no beliefs', 'I have no thoughts' (where 'I think that' is synonymous with 'I believe that') and 'I know nothing'.

He claims that in none of these cases can the proposition be coherently asserted by the person to whom it refers, although it could be coherently asserted of that person by someone else. But, for reasons already explained, the most that Mackie is entitled to claim is that these statements cannot be truthfully or sincerely asserted, and not that they cannot be truly or coherently asserted. Thus, either p or dp may be true, but not both. In other words, they are just another form of ad hominem self-refutation. 2

On the other hand, certain other statements are operationally self-refuting in Mackie's original sense; e.g. 'I never make assertions' and 'I never tell the truth'. For although, as propositions, they cannot be coherently asserted by the person to whom they refer, they can be coherently asserted of that person by someone else.

It is, as we have seen, quite possible to give an account of operational self-refutation without making use of the

1See p.116-123. 2See p.116.
somewhat dubious notion of weak prefixability. Indeed, the term cannot even be applied to some operationally self-refuting statements. For instance, statements like 'no language of any kind exists', 'no (putative) sentences are intelligible', 'nobody ever makes assertions', 'at least one statement-maker exists' state necessary conditions of the possibility of statement-making in general, and are therefore, on Mackie's own admission, operationally (as opposed to absolutely) self-refuting. And yet, no weakly prefixable operators, or indeed operators of any kind, seem to be involved in such statements.*

An operationally self-refuting statement may be defined quite simply as a proposition which cannot be coherently asserted by a particular person or class of persons, because one of the necessary conditions of its truth is that it remain unasserted by that person or class of persons. To be strictly accurate, it is only self-refuting when it is actually being asserted in that way. In other words, it is the assertion rather than the proposition which is self-refuting.

*It is perhaps worth noting that the operator involved in the operationally self-refuting statement 'no true assertion has ever been made' is strictly prefixable.

We are now in a position to distinguish at least two kinds of operationally self-refuting statement: namely, that where a proposition cannot be coherently asserted by a limited class of persons; e.g. 'I never make assertions' and 'I never tell the truth'. And that where it cannot be coherently asserted by an unlimited class of persons; e.g. 'nobody ever makes assertions', 'nobody ever tells the truth', 'no language of any kind exists'. I shall mark this distinction by describing the former as referentially self-refuting and the latter as universally self-refuting.

Whereas the proposition expressed by a referentially self-refuting statement can be coherently asserted by some people, but not by a particular person or group of persons, the proposition expressed by a universally self-refuting statement cannot be coherently asserted by anyone, even though it describes a perfectly conceivable state of affairs. For instance, the proposition expressed by the referentially self-refuting statement 'I never make assertions' could be coherently asserted of me by someone else, whereas the proposition expressed by the universally self-refuting statement 'nobody ever makes assertions' could not be coherently asserted by anybody.

The significance of this distinction will be made clear in Part III, in which the two strands of our investigation (transcendental deductions and the concept of self-refutation) will be brought together.
The following diagram shows how the various types of self-refuting statement that have been described in the preceding sections are related to one another:

1. e.g. 'all type-written statements are illegible'.
2. e.g. 'I cannot talk' (spoken).
3. e.g. 'it can be proved that nothing can be proved'.
4. e.g. 'I have no beliefs'.
5. e.g. 'I sometimes write that I cannot write'.
6. e.g. 'no language of any kind exists'.
7. e.g. 'I never make assertions'.
PART III

The Composition of a Transcendental Deduction

Section 1. The introduction.

In Part I, we examined in some detail the general requirements that any argument must satisfy if it is to count as a genuine transcendental deduction. We saw, for instance, that in a transcendental deduction one is supposed to be able to prove that a particular statement of fact is absolutely certain because it is presupposed by some other statement or principle. But this would only be possible if the statement by which it was presupposed was itself both factual (i.e. non-analytic) and absolutely certain. However, this is a requirement which, according to the standard doctrine of empiricism cannot be met. For, according to the empiricists, the truth of a statement of fact can only be demonstrated empirically, and yet no amount of empirical evidence would ever entitle us to say that it was absolutely certain (i.e. as certain as any analytic statement).

If we accept the whole of this doctrine, then clearly transcendental deductions are impossible. But do we have to accept the whole of it? Is there perhaps not some way of guaranteeing the truth of a statement of fact which does not involve the use of empirical data? In other words, are there
not some synthetic statements which could be shown to be true on the basis of internal (i.e. logical) evidence alone? For such statements would clearly be self-guaranteeing and therefore in no need of external or empirical proof.

I went on to argue that some statements are indeed self-guaranteeing in the sense that any attempt to deny them would be self-refuting. Statements of the necessary conditions of discourse come into this category; e.g. 'a language of some kind exists', 'some sentences are intelligible'. Such statements are at least as certain as any analytic statement, for there is no way in which they can be coherently denied. They could therefore presumably function as axioms in a transcendental deduction.

On the other hand, it must be admitted that the authenticity of such statements is by no means universally accepted. For instance, some of the things which have been said about self-refuting statements would seem to suggest that transcendental axioms in the above sense are impossible. Mackie, for example, has said that:

We might be tempted to believe that there is a special form of philosophical argument which enables us to establish positive conclusions by showing that certain contrary statements would be self-refuting. This would go against empiricism for if any view would literally refute itself its denial would be a necessary truth.
He adds that his own analysis shows that 'this challenge to empiricism evaporates on closer inspection'.

Or again, it might be argued that a statement can only be self-refuting or self-guaranteeing if it is in some sense self-referring, but that all self-referring statements are either meaningless or vacuous.

If either of these claims were true, transcendental deductions would be impossible. In order to deal with these objections a detailed analysis of the concepts of self-reference and self-refutation is obviously called for. And this means examining as wide and varied a range of self-referring and self-refuting statements as possible. This exercise was duly performed in Part II.

It was then discovered that although some self-referring statements might be meaningless or vacuous (e.g. all exclusively self-referring statements) others were neither meaningless nor vacuous, and these included statements which were (apparently) self-guaranteeing, and therefore suitable as transcendental axioms.

It was also discovered that Mackie's claim that if any view refuted itself, its denial would be a necessary truth, was only true of absolutely self-refuting statements; e.g. reflexively self-referring statements like 'it can be proved

---

that nothing can be proved', and 'it is certain that nothing is certain'. However, this fact creates no special problems, either for us or for the empiricists, since absolutely self-refuting statements are straightforwardly self-contradictory, whilst their contradictories are straightforwardly analytic. And no analytic statement can function as a transcendental axiom.

On the other hand, it was found that the denial of certain other self-refuting statements, although logically incoherent, did not give rise to analytic or necessary truths; e.g. 'no language of any kind exists' and 'no (putative) sentences are intelligible'. Such statements have been termed 'operationally self-refuting'.

The views which these statements express are not in themselves self-refuting, rather it is the assertion of them which is self-refuting. Nevertheless, it follows that they cannot be truly asserted and that their contradictories cannot be truly denied. But if a statement like 'a language of some kind exists' cannot be truly denied, then it is undeniably true, and therefore at least as certain as any analytic statement. It could therefore function as a transcendental axiom.

It would seem from our discussion so far that the only type of statement which could function as the initial premise in a transcendental deduction would be one whose contradictory was operationally self-refuting. If this was the case, then
a transcendental deduction could be re-defined as an argument of the transcendental form in which the minor premise (y is in fact possible) is operationally self-guaranteeing.

This view is re-inforced by the fact that none of the other types of self-refutation examined in Part II give rise to self-guaranteeing statements of any kind. At first sight the contradictory of an absolute ad hominem self-refutation may appear to be self-guaranteeing, but our analysis has shown that it is not.¹ For instance, 'I have no beliefs' is self-refuting in this sense. But although a person who says that he has no beliefs is committed to saying that he believes that he has no beliefs, and therefore to contradicting himself, his assertion that he has no beliefs could still be true as long as he does not intend it to be true.

There is, however, at least one other type of self-guaranteeing statement, which we have not so far considered, whose contradictory is in no sense self-refuting. And it is worth finding out whether this could function as a transcendental axiom.

A good example would be the wholly self-referring statement 'I sometimes say things which are false'. This statement is self-guaranteeing because in order for it to be false, it would have to be true, whereas if it was true it

¹See Part II, section 8.
would still be true. In other words, the assertion 'I sometimes say something false' is something I say, and so if it is false then it is true that I sometimes say things which are false, but if it is true that I sometimes say things which are false then it is true that I sometimes say things which are false. Hence the statement is self-guaranteeing.

On the other hand, its contradictory 'I never say things which are false' is in no sense self-refuting. It is therefore quite different from a statement like 'I sometimes make assertions' which is also self-guaranteeing, but whose contradictory 'I never make assertions' is operationally self-refuting.

The distinction here is roughly between the kind of statement which is self-guaranteeing because its falsity is a sufficient condition of its own truth, and the kind of statement which is self-guaranteeing because its truth is a necessary condition of the type of discourse to which it belongs. One could if one wished mark this distinction by describing the first type of statement as self-validating and the second as self-guaranteeing.

A strong objection to using the first type of statement as the initial premise in any type of philosophical argument is that it is very doubtful whether it makes sense to interpret it as a self-referring statement at all. For it is no different
in principle from statements like 'all the statements on this page are false' and 'this statement is true', which have been regarded by some philosophers (especially by those who accept Russell's theory of types) as meaningless in so far as they are interpreted self-referencing.

Even if this extreme view is rejected, it must be admitted that the statement 'I sometimes say things which are false' must function both as a statement and as a meta-statement if it is to succeed in guaranteeing its own truth (a task somewhat analogous to that of a person attempting to pull himself up by his own boot-strap), and, as we have seen, any statement which does this is vacuous in so far as it is interpreted self-referencing. It is therefore not the kind of statement which should be used as the initial premise in a philosophical argument.

The same charge cannot be made against those other self-guaranteeing statements whose truth we are forced to take for granted in all our assertions, for whatever is presupposed by a meaningful assertion must itself be meaningful. Thus, these operationally self-guaranteeing statements would appear to be the only ones which could be used as axioms in a transcendental deduction.

The question now arises whether the range of operationally self-guaranteeing statements which could perform such a function

---

1 See Part II, section 3. 2 Ibid (esp. pp.91-93).
is limited by yet other requirements, or whether any operationally self-guaranteeing statement at all could be used as the basis of a transcendental deduction.

If there were no other requirements, there would be no reason why the conclusion of a transcendental deduction should be treated as a conclusion rather than as an axiom, because in a transcendental deduction one simply moves from one self-guaranteeing statement to another. For if the initial premise is self-guaranteeing, any statement which is logically derived from it must also be self-guaranteeing. Why, then, should any particular operationally self-guaranteeing statement be granted axiomatic status in preference to any other?

At least part of the answer to this question is that some statements are more obviously and straightforwardly self-guaranteeing than others. Indeed, there would be no point in a transcendental deduction if this were not so. For example, a statement like 'a language of some kind exists' is obviously self-guaranteeing, and a statement like 'there is some regularity and uniformity in the way things happen' unobviously self-guaranteeing. Thus there would be no point in attempting to derive the former statement from the latter (assuming that this were possible), whereas there would be some point in attempting to derive the latter statement from the former. For this would show that, in spite of appearances, it was self-guaranteeing.
It should be remembered, however, that whether we judge a truth to be obvious or not depends in part on our own perceptiveness. What is obvious to some may not be obvious to others. Thus, it might be argued that, in principle, any operationally self-guaranteeing statement could be used as the initial premise in a transcendental deduction, although for practical reasons our choice is limited. However, such a conclusion cannot be established so easily.

For instance, it might be argued that ultimately all operationally self-guaranteeing statements can be traced back to a single principle, and that this should be regarded as the true starting-point of any transcendental deduction.

The argument might run as follows: Only those assertions whose contradictories are operationally self-refuting can function as transcendental axioms. Such statements are self-guaranteeing only in so far as they state necessary conditions of the possibility of making true assertions.* Thus, the possibility of making true assertions is the ultimate basis of any genuine transcendental deduction. In other words, every member of the class of operationally self-guaranteeing statements can ultimately be traced back to the principle that true and meaningful assertions are possible (or words to that effect).

---

*This is true even of referentially self-guaranteeing statements, as we shall see.
Moreover, this principle (qua actual assertion) must itself be regarded as a member of the class, and therefore able to function as a transcendental axiom. In this sense, it has logical priority over all other operationally self-guaranteeing statements.

It is interesting to compare the view that this principle lies at the root of all transcendental deductions with Kant's own views on the matter. For Kant maintains that transcendental deductions are ultimately based on the principle of the possibility of experience:

The transcendental deduction of all a priori concepts has thus a principle according to which the whole enquiry must be directed, namely that they must be recognized as a priori conditions of the possibility of experience.¹

It should be noted that this claim is in no way inconsistent with Kant's other claim that the transcendental unity of apperception (viz. that the 'I think' must accompany all my representations)² is 'the highest principle in the whole sphere of human knowledge;'³ for this latter principle merely spells out what is already contained in the (kantian) concept of experience. Thus, by 'experience' Kant simply means 'representations which are thought' (i.e. brought under concepts).

¹A74. ²B131. ³B135.
It might be objected that the 'I' at least is not present in the mere concept of the possibility of experience. However Kant explicitly states that the 'I' in the 'I think' is not to be taken as referring to an existent of some kind; that whether it does so or not is strictly speaking irrelevant.

The proposition 'I think' is here taken only problematically, not in so far as it may contain perception of an existent but in respect of its mere possibility, in order to see what properties applicable to its subject (be that subject actually existent or not) may follow from so simple a proposition. 1

As we discovered in Part I, 2 the statement 'experience is possible' must be self-guaranteeing if it is to function as a transcendental axiom. But it can only be self-guaranteeing in the required sense by stating a necessary condition of the possibility of statement-making in general. Thus the principle of the possibility of experience should be derived from the principle that true and meaningful assertions are possible (that is to say, all the conditions necessary for making true and meaningful assertions exist) and not, as Kant would have it, the other way round.

The claim that all genuine transcendental deductions are ultimately based on the principle that true and meaningful assertions are possible does not of course mean that every

1A347/B405. 2pp.59 & 60 of this thesis.
transcendental deduction must necessarily start from such a premise if it is to be valid, but merely that it always could start from such a premise. In other words, all transcendental deductions can ultimately be traced back to this same principle, although they are not all bound to take it as their actual starting-point. The reason for this is simply that any transcendental proposition which is derived from this principle is ipso facto operationally self-guaranteeing, but the mere fact that it is operationally self-guaranteeing is prima facie adequate reason for treating it as an axiom.

However, it does not follow that all statements in this category could equally well function as axioms in a transcendental deduction. There are, for instance, at least two kinds of operationally self-guaranteeing statement: those which cannot be coherently denied by a particular person or group of persons (referentially self-guaranteeing), and those which cannot be coherently denied by anyone (universally self-guaranteeing). The question therefore arises 'does it make any difference which of these two classes the transcendental axiom is taken from?' I shall try to show that it does.

As we have already seen,¹ the transcendental method is parallel in at least one respect to the geometrical method; that is to say, it attempts to demonstrate the truth of certain

¹pp. 77 & 78 of this thesis.
theorems by logical inference from a set of axioms. Unlike the axioms of geometry, however, the axioms of transcendental philosophy are not at all arbitrary, for they cannot be replaced by axioms with which they are logically inconsistent. In other words, they are undeniably true.

There is, of course, nothing new or original in this approach, for Descartes also attempted to derive an entire metaphysical system from a single indubitable premise; and a premise, moreover, which looks as though it might be referentially self-guaranteeing.

It is, I think, possible to throw a great deal of light on the transcendental method in general by comparing it with the Cartesian method, for they are both attempting to achieve the same ends, namely, to ascertain what, if anything, can be known with complete certainty about the self, the world, and God. Thus, by comparing what Kant held to be the most fundamental principle in philosophy with what Descartes regarded as the true starting-point of his philosophical system, it is possible to learn a great deal about the kind of premise which could stand as the basis of a transcendental deduction.

We may also hope, by this means, to shed some light on the various claims which Kant makes on behalf of the transcendental method; for example, that it provides us with our only access to metaphysical truth (that is to say, a priori-
synthetic truth), and that the method is entirely different from any hitherto employed.

The starting-point of the Cartesian and transcendental methods is, in a sense, the same. For both Descartes and Kant begin with the concept or judgement 'I think', although their treatment of it is quite different. I hope to show that whereas, in general, Kant's approach is the correct one, there is at least one respect in which he should have adhered more closely to the Cartesian mode of argument.

Section 2. 'I do not think' as a self-refuting statement.

In his preface to the first edition of the Critique, Kant says:

As to certainty, I have prescribed to myself the maxim, that in this kind of investigation it is in no wise permissible to hold opinions. Everything, therefore, which bears any manner of resemblance to an hypothesis is to be treated as contraband; it is not to be put up for sale even at the lowest price, but forthwith confiscated, immediately upon detection.

Descartes takes an exactly similar line in his Meditations, and resolves to 'withhold belief no less carefully from things not entirely certain and indubitable than from those which appear to me manifestly false'. He realizes that his first task, therefore, is to find at least one thing whose existence is

1 Av. 2 Meditations on the First Philosophy, 'Meditation I'.

1.52
'certain and indubitable'.

In the second meditation, he decides that the existential proposition "'ego sum, ego existo' is necessarily true every time I propound it or mentally apprehend it". And he uses this principle as the basis of his philosophical system. The argument by which he backs up this claim may be summarized as follows: in order to doubt my own existence I must already exist as a being which is capable of doubt (and therefore of thought). In this sense, my existence as a thinking being is indubitable.

This is clearly a self-refuting argument of some kind, since Descartes is claiming that any attempt to deny one's own existence would be self-refuting. A self-refuting argument is an argument to the effect that a particular statement should be accepted as true because its contradictory is, in some sense, self-refuting.

Obviously, not just any kind of self-refuting argument would serve Descartes' purpose, but only one that was capable of showing that some statement or other was indubitable, even though the situation it described might not have existed. In other words, it must give rise to a self-guaranteeing statement. Descartes would not therefore be justified in using an argument which was based on the concept of indirect or ad hominem self-refutation in order to establish his initial
premise, because a statement cannot be shown to be self-guaranteeing in this way.

This is obviously true in the case of a standard ad hominem self-refutation (e.g. 'I sometimes write that I cannot write') where the first-order statement is false if, and only if, the second-order statement is true, and where the truth of the second-order statement can only be determined empirically.

An absolute ad hominem self-refutation, such as 'I have no conscious beliefs', is equally though less obviously prevented from playing a part in Descartes' system, for its contradictory is neither analytic nor self-guaranteeing. At first sight, it might appear to be self-guaranteeing since I cannot assert that I have no beliefs without implicitly denying the truth of my assertion. For in asserting anything, I implicitly commit myself to the claim that I believe my assertion to be true. I cannot therefore claim that I have no beliefs without committing myself to the further claim that I believe that I have no beliefs. Nevertheless, as we saw before, this does not mean that I cannot coherently assert that I have no beliefs, but only that I cannot truthfully assert it. Thus, I can correctly assert that I have no beliefs as long as I neither believe nor disbelieve in the truth of my assertion, and have no other beliefs.

1See Part II, section 8.
As Professor Passmore has pointed out,¹ it is possible to represent Descartes' famous cogito formula, 'I think therefore I am', as being based on a form of ad hominem self-refutation. The first step would be to show that a person who says that he has no thoughts, is committed to the additional claim that he thinks that he has no thoughts. For, as Passmore says, if Descartes could show that a person who says that 'I do not think' is admitting that he sometimes thinks that he does not think then he could use this fact to prove that such a person can be compelled to admit that what he thinks cannot be true.

This first step is easily taken since, as we saw in the previous paragraph, in making any assertion I commit myself to the claim that I believe that what I assert is true, but since believing is a form of thinking, it follows that in asserting something I commit myself to the claim that I think that what I say is true.

The next step would be to show that I can coherently say of myself that I am not thinking, just as I can coherently say of myself that I have no beliefs, although in both cases I could be forced to withdraw my claim by someone drawing my attention to the fact that I could not consistently believe in the truth of what I was saying. In other words, it must be shown that I can truly assert that I am not thinking without

¹Philosophical Reasoning, Ch.4, pp.63 & 64.
actually thinking, just as I can truly assert that I have no beliefs without either believing or disbelieving in the truth of the assertion.

On the other hand, it might be argued that in order to assert anything I must think it, and that therefore I cannot truly assert that I am not thinking, just as I cannot truly assert that I never make assertions. This would mean that the statement 'I do not think' could be classified as operationally self-refuting, which is what Mackie claims.1

Whether the sentence 'I do not think' expresses a self-refuting statement, and if so, whether it expresses an operationally self-refuting statement rather than an ad hominem self-refutation, depends of course on how the key term 'think' is interpreted here.

The Oxford English Dictionary suggests a number of possible meanings for the term. Two very common uses are: 'to reason or reflect about something' as in 'give me time to think', and 'to be of the opinion that something is the case' as in 'I think that Mary is sick'.

If, in accordance with the first of these interpretations, the statement 'I do not think' is taken to mean 'I do not reason' then it is not self-refuting at all, since an assertion is an assertion whether it is preceded by a process of reasoning or not.

---

1See Philosophical Quarterly, July 1964, pp.197-198.
not. If, in accordance with the second of these interpretations, it is taken to mean 'I do not hold any opinions' then it is an example of an absolute ad hominem self-refutation. For an assertion is an assertion whether the person who makes it has any opinion about it or not, although if he does not have an opinion he can be forced to withdraw the assertion.

In a more technical or philosophical sense 'to think' can mean something like 'to employ concepts' or even 'to be conceptually aware of something'. Thus, we could take 'I do not think' to mean 'I do not employ concepts'. According to this interpretation, the statement is operationally self-refuting, because a statement just is a way of using concepts. There is therefore no way in which a person can coherently assert that he does not employ concepts, even though it may in fact be the case that he never employs concepts.

Operationally self-refuting statements may, as we have seen, be divided into those which are universally self-refuting, and those which are referentially self-refuting. A statement is universally self-refuting if a necessary condition of its truth is that it remain unasserted; e.g. 'no language of any kind exists'. A statement is referentially self-refuting if a necessary condition of its truth is that it remain unasserted by the very person who is asserting it; e.g. 'I never make assertions'. The distinction is really between what one must
accept as a basic truth about the world in general if one is to say anything at all, and what one must accept as a basic truth about oneself in particular if one is to say anything at all.

Clearly, if the statement 'I do not think' is to be regarded as a form of operational self-refutation, it must be classed as referentially, rather than as universally self-refuting, since my ability to think can only be a necessary condition of my making assertions, and not a necessary condition of discourse in general.

Descartes would not be justified in basing his entire metaphysical system on the cogito principle if the statement 'I do not think' gave rise only to an absolute ad hominem self-refutation, but since, on at least one interpretation, it gives rise to a form of operational self-refutation, this objection does not apply. The question remains, however, whether Descartes would be justified in basing his metaphysical system on the cogito principle, even if this latter interpretation were accepted.

This question really amounts to asking whether a referentially self-guaranteeing statement could be used as the basis of a transcendental deduction, because, as I shall now try to show, a transcendental deduction is the only kind of self-refuting argument which would serve Descartes' purpose.
Section 3. The concept of self-evidence.

In the synopsis to his Meditations Descartes says:

Since it has been my endeavour to write in this treatise nothing of which I cannot give exact demonstration, I have found myself obliged to adopt an order similar to that used by geometers, viz. to state all the premises on which the proposition in question depends, before coming to any conclusion regarding it.

Since Descartes' conclusions are for the most part concerned with matters of fact, the axioms from which they are derived must be both factual and completely certain. Unlike the geometer, therefore, Descartes is faced with the problem of explaining how he can know that his axioms are certain.

However, in demonstrating the truth of his axioms, Descartes must take care not to deprive them of their axiomatic status. In other words, he must show that, in a sense, they do not need to be proved, and that he is justified in taking them for granted; i.e. treating them as axioms. For clearly, if he tried to ground each of his proofs on a prior factual premise he would be caught in an infinite regress, and would never be able to prove anything absolutely.

Thus, whereas Descartes cannot afford to take his axioms for granted without first showing that he is justified in taking them for granted, a geometer can afford to regard his axioms as completely arbitrary.
An obvious way of defending the axiomatic status of a premise is by arguing that it is self-evidently true, and therefore in no need of proof. In effect, this is what Descartes' doctrine of clear and distinct perceptions really amounts to. Thus, at the end of the second paragraph in the third meditation, he says 'everything I apprehend in a genuinely clear and distinct manner is true'. In other words, whatever appears to me to be patently self-evident must be true.

He seems at times to regard this subjective test as being ultimately the only sure criterion of truth. The fallacy inherent in this view lies, of course, in the underlying assumption that there is a way of apprehending or intuiting an idea which is qualitatively different from all other modes of apprehension and intuition, and which somehow guarantees the truth of the idea. If, however, we eliminate this notion, all that we are really left with, is the bare concept of self-evidence. But clearly a statement can appear to be self-evident without in fact being true, and therefore could scarcely be used as a criterion of truth.

To say that a statement is self-evident is to imply that any attempt to prove it would be superfluous, since it neither is, nor appears to be, less certain than any other statement; that is to say, its certainty for us could not be increased by reference to other allegedly true statements. It is therefore
just the kind of statement which we would be justified in taking for granted in any of our proofs.

If, however, Descartes is to abide by his resolution to doubt everything which is capable of being doubted, he cannot afford to leave even his axioms unchallenged. Thus, although there is no question of his being able to prove his axioms, he must at least explain why he holds them to be self-evident; in other words, he must explain why he is justified in taking them for granted.

Descartes could not of course justify his axioms by appealing to other facts because, as we have already seen, this would mean depriving them of their axiomatic status. And he would then be landed with a new set of axioms, equally in need of justification.

If the axioms are to be justified, therefore, they must be justified on the basis of logical criteria alone; that is to say, their justification must be treated as an internal matter. But there is only one way in which a statement of fact can be shown, on logical grounds, not to be in need of any proof, and that is, by showing that it is in some sense self-guaranteeing or self-validating.

A statement of fact may be said to be self-guaranteeing or self-validating if its truth is presupposed by its contradictory. In such a case the statement cannot be coherently
denied, because in denying it one is forced to take its truth for granted; e.g. 'a language of some kind exists'. One is justified in taking such a statement for granted (i.e. treating it as an axiom) simply because one is forced to take it for granted if one is to say anything at all. This does not of course mean that the axiom is true, but only that we are obliged to regard it as true. In any case, to demand that nothing be taken for granted, and that even the axioms be proved, is absurd, since every proof must start from somewhere; i.e. presuppose something.

Thus, an axiom is just the kind of statement which we either cannot prove or do not need to prove. There can be no stronger justification of an axiom than to show that without it, we could never arrive at any kind of knowledge. This is the point which Aquinas was making when he said 'that through which all the rest can be known must itself be self-evident'.

A self-guaranteeing statement, then, is a statement which makes no sense to doubt, even though it is only contingently true. Any conclusion which is logically deriveable from such a statement will be similarly indubitable.

Since Descartes proposes to exclude from his philosophical system any statement which is less than indubitable, he ought to derive his theories only from those axioms which can be shown

---

1Summa Contra Gentiles, Book I, Ch. 10. section 6.
See also section 4 of this thesis.
to be self-validating or self-guaranteeing. Spinoza was perhaps the first philosopher to realize this,¹ but the same point has recently been reaffirmed by Bernard Williams in his article on Descartes in the Encyclopedia of Philosophy:²

Descartes' task will not have been satisfactorily carried out so long as, for anything I claim to know, it is possible to ask whether I know that I know it; to eliminate this question, genuine knowledge has to be self-guaranteeing.

According to Bernard Williams, statements may be described as self-guaranteeing if, and only if, their contradictories 'could not possibly be true because their truth would defeat the conditions of their own assertibility'.³

Another way of putting this, would be to say that the only type of argument which would serve Descartes' purpose is a transcendental deduction. For a transcendental deduction has already been defined as an argument in which a statement of fact is shown to be indubitable on the grounds that any attempt to deny it would be operationally self-refuting.

It is interesting to note that one of Kant's chief claims on behalf of the transcendental method is that it provides us with our only access to metaphysical truth.⁴ If we take the

¹Treatise on the Correction of the Understanding, section VII.
²Volume 2, p.346, section entitled 'the method of doubt'.
³Volume 2, p.347, section entitled 'cogito ergo sum'.
expression 'metaphysical truth' to mean 'a logically indubitable but non-analytic truth about the universe', then we must agree with him.

Section 4. Some self-refuting arguments.

It is possible, as we have seen, to interpret the 'I think' as a self-guaranteeing statement. Thus, it might be thought that Descartes at least succeeds in getting off on the right foot. But even if this is the case, he goes astray almost immediately thereafter, when he says:

What then is it that I am? A thinking being. What is a thinking being? It is a thing which doubts, understands, affirms, denies, wills, abstains from willing, that also can be aware of images and sensations. ¹

The transition from the general to the particular here is clearly invalid, since although the statement 'I think' is entailed by any particular statement about the way in which I think, it does not itself entail that I think in any particular way. For example, it might be said that if I am aware of sensations then in a very general sense of 'think' it is true that I think. But it has not been shown that if I think then I am necessarily aware of sensations.

Descartes' most serious mistake is that he introduces new concepts into his metaphysical system without first

¹'Second Meditation', paragraph 10.
justifying them by means of a transcendental deduction. The correct procedure to be followed in a transcendental deduction is first, to establish the indubitability of the axiom by means of a self-refuting argument, and then to show how other indubitable truths can be derived from it. This may be done either by showing that certain other propositions are logically entailed by the axiom or, more interestingly, that they are logically presupposed by the axiom. This latter course is the more fruitful, I believe.

Descartes starts off in the right way by first establishing the indubitability of his initial premise, but goes wrong almost immediately by introducing into his system other axioms which he makes no attempt to justify. For instance, he assumes, without argument, that he cannot possibly be mistaken about the contents of his own consciousness. Thus, he thinks that if a statement of 'immediate experience' is recognized as being true then it is indubitably so.

This assumption is vital for Descartes, because all his proofs of the existence of God start from his awareness that he already possesses an idea of God. However, the statement 'I have an idea of God' is in no sense self-guaranteeing, and could quite easily be false. Why, then, should it be regarded as indubitable? There is surely nothing inherently absurd in the notion that I might be wrong in thinking that I have an idea
of God, for might I not be confusing the idea of God with the idea of something else? In some cases, the task of classifying an idea or sense-datum correctly might be very difficult (e.g. where states of mind are concerned) but even in the simplest and most straightforward cases (e.g. classifying coloured images) there can be no built-in guarantee that the idea or sense-datum is being classified correctly.

On the other hand, the theory that we might always be mistaken in our classification is self-refuting, since if we did not know how to use general concepts correctly we could not make meaningful assertions.

Even statements which appear to consist wholly of proper-names linked by the copulative 'is' (e.g. 'Hesperus is Phosphorus') presuppose general concepts, as Anscombe and Geach have shown.¹ For we can only apply proper-names correctly if we can identify the individuals to which they refer, but there is no way in which we can identify an individual without identifying it as an individual of a certain kind. In other words, in order to identify anything at all, we must identify it under a particular description. Thus, the statement 'I sometimes use class-words correctly' is a necessary condition of my engaging in discourse, and is therefore self-guaranteeing. But this does not imply that we always use class-words correctly,

¹Three Philosophers, p.10 (Blackwell, 1963).
and so does not entitle us to infer that a statement like 'I have an idea of God' is self-guaranteeing.

The only time that Descartes makes use of a self-refuting argument is when he tries to demonstrate the indubitability of his own existence as a thinking being. But if he had attempted to demonstrate the indubitability of God's existence in the same kind of way he would not have been the first philosopher to do so.

In Book I, chapter 10, of the Summa Contra Gentiles, Aquinas deals with the arguments of those philosophers who claim that the existence of God, being self-evident, cannot be demonstrated by reference to anything else. One of the arguments he describes is an obvious, but unsuccessful attempt at a transcendental deduction. I will therefore quote it in full:

There is also the consideration that that through which all the rest are known must itself be self-evident. Now, God is of this sort. For just as the light of the sun is the principle of all visible perception, so the divine light is the principle of all intelligible knowledge; since the divine light is that in which intelligible illumination is found first and in its highest degree. That God exists, therefore, must be self-evident.

To be valid, this argument would have to be interpreted in the

^Book I, Ch.10, the fifth argument. (Doubleday, 1955).
following way: If God's existence is a necessary condition of our knowing or understanding anything then we cannot meaningfully assert that he does not exist, for if we do not understand what we are saying we cannot be said to be asserting anything. In other words, the statement that God exists is self-evident in the sense of being self-guaranteeing or self-validating. If it was false, we would not be able to state that it was false.

The argument as it actually stands is of course fallacious, since it derives whatever force it might appear to have from an ambiguity in the phrase 'that through which all the rest is known'. For, as Aquinas says:

God is indeed that by which all things are known, not in the sense that they are not known unless He is known (as obtains among self-evident principles), but because all our knowledge is caused in us through His influence. ¹

A further attempt at a transcendental proof of God's existence is to be found in the First Part of the Summa Theologiae, Question 2:

It is self-evident that truth exists, for even denying it would admit it. Were there no such thing as truth, then it would be true that there is no truth; something then is true, and therefore there is truth. Now God is truth itself; 'I am the way, the truth and the life'. That there is a God, then, is self-evident. ²

¹Book I, ch. 11, section 7. ²a, question 2, article 1, section 3.
The fallacy in this argument, as Aquinas points out, is that although it is self-evident* that there is at least one truth (i.e. one true statement) 'it is not self-evident to us that there exists a First Truth'.

Both these arguments are designed to show that we are justified in treating the statement 'God exists' as a self-evident axiom or first-principle, since any attempt to deny it would be self-refuting. But this is very similar to the way in which Kant tried to justify the fundamental principles of Newtonian physics. Thus, his claim that the transcendental method 'is entirely different from any hitherto conceived' must be rejected.

Although the statement 'God exists' is clearly not self-guaranteeing, some of Descartes' other statements are. For instance, several of his claims about what it is for him to exist as a thinking being are self-guaranteeing, even though he does not attempt to show that they are; e.g. his claim that he is a being who understands, affirms and denies things. Thus, he cannot coherently deny that there are things which he understands, because this would imply that he could not

*It is not, however, a necessary truth that there is at least one truth, as I have already tried to show. See pp.130-133.

understand the meaning of his own utterances, in which case they would not count as genuine assertions or denials. Again, he could not coherently say of himself that he was a being who never affirmed or denied anything, because to make an assertion just is to affirm it, and to affirm something just is to deny its contradictory.

Descartes also says that he is a being who is aware of images and sensations.\(^1\) Kant would presumably regard at least the first part of this claim as a priori-synthetic (or, on our interpretation, self-guaranteeing) on the grounds that the awareness of images (where 'image' is taken to cover all sensory data) is a necessary condition of language; for as he says 'concepts without intuitions are empty'.\(^2\) On this view, to say that one is never aware of images is to say that one cannot attach any meaning to one's own utterances.

Apart from the cogito principle itself, Descartes does not attempt to establish the truth of any of his statements by showing that they are either self-guaranteeing or else presupposed or entailed by statements which are self-guaranteeing. Thus, whereas Kant attempts to prove the existence of the external world by means of a transcendental deduction\(^3\) (that is, by showing it to be a necessary condition of experience, and therefore of language), Descartes attempts to prove it by

---

\(^1\) Op. cit. (above, p.164, note 1). \(^2\) A51/B75. \(^3\) B274ff.
means of the singularly weak argument that since God is the
author of our nature he would not suffer us to be deceived
about so important a matter.¹

Section 5. The cogito principle as transcendental axiom.

I suggested earlier that in adopting the cogito principle
as his starting-point, Descartes at least manages to get off
on the right foot; in other words, such a principle could
quite legitimately be used as the basis of a transcendental
deduction. I shall now try to show that not only is this claim
true, but that it is also one to which Kant himself is logically
committed.

Although Kant is highly critical of what he terms
Descartes' problematic idealism, he shares with him the view
that there is no room in philosophy for speculation or
hypothesis, and that no statement should be accepted as true
unless it can be shown to be indubitable. At B275, he says:

problematic idealism is, in so far as it allows
of no decisive judgement until sufficient proof
has been found, reasonable and in accordance with
a thorough and philosophical mode of thought.

Further, he fails to find fault with Descartes' choice of the
cogito principle as his starting-point, for he goes on to say
that if Descartes is to prove the existence of the external
world, he must show:

¹See the sixth Meditation.
that we have experience, and not merely imagination of outer things; and this it would seem, cannot be achieved save by proof that even our inner experience, which for Descartes is indubitable, is only possible on the assumption of outer experience.¹

In other words, the existence of the external world can be proved on the basis of the cogito principle, although not in the way that Descartes thought. In short, Kant is not objecting to Descartes' having used the principle for this purpose, but to his having misused it.

Earlier I suggested that there are at least three possible ways in which the statement 'I think' can be interpreted.² I tried to show, however, that only if it is taken in a very general sense to mean something like 'I am conceptually aware of something' can it be regarded as self-guaranteeing. The term 'conceptually aware' here refers to any form of awareness which involves the use of concepts. It therefore covers a wide range of mental processes: from bringing items in our sensory field under concepts, to judging that one proposition is entailed by another.

When interpreted in this very general way, the statement 'I think' is self-guaranteeing, in the sense that its truth is a necessary condition of its own assertibility. In other words, it is a necessary condition of my being able to say anything

¹B275. ²See pp.156 & 157.
at all, and therefore any attempt on my part to deny it, would be operationally self-refuting. Thus, it appears to be just the kind of statement which could be used as the basis of a transcendental deduction.

Kant realizes that the 'I think' principle belongs to a very special class of statements for, in the first chapter of the 'Paralogisms of Pure Reason', he says:

We come now to a concept which was not included in the general list of transcendental concepts but which must yet be counted as belonging to that list, without, however, in the least altering it or declaring it defective. This is the concept or, if the term be preferred, the judgement 'I think'. As is easily seen, this is the vehicle of all concepts, and therefore also of transcendental concepts, and so is always included in the conceiving of these latter, and is itself transcendental.

Clearly, Kant is taking the expression 'I think' here to mean something like 'I am conceptually aware'. We are therefore entitled to regard it as operationally self-guaranteeing.

By describing the 'I think' as a transcendental concept or judgement Kant places it amongst the necessary conditions of the possibility of experience. It must be remembered, however, that for Kant experience is always conceptualized experience:

\[1\text{A34/399}. \quad 2\text{e.g. A783/B811}.\]
Experience is itself a species of knowledge which involves understanding; and understanding has rules which I must presuppose as being in me prior to objects being given to me, and therefore as being a priori. They find expression in a priori concepts.¹

Unconceptualized experience he calls intuition; 'thoughts without content are empty, intuitions without concepts are blind'.² In other words, for Kant, anything which is a necessary condition of discourse (i.e. concept-employment) is a necessary condition of experience. Since the 'I think' is necessarily involved in this process of conceptualization it is, for Kant, a necessary condition of the possibility of experience. This is why he describes it as a transcendental concept or judgement.

But if Kant is committed to saying that anything which is a necessary condition of discourse is a necessary condition of experience, he is also committed to saying that anything which is a necessary condition of experience is a necessary condition of discourse. For, he not only maintains that intuition must be conceptualized to become experience, he also maintains that concepts without intuitions are empty; 'thoughts without content are empty'.

On this view, the statement 'there is no such thing as experience' is operationally self-refuting, since it implies

¹Bxvii. ²A51/B75.
that any kind of meaningful discourse is impossible. Its contradictory 'there is such a thing as experience' must therefore be self-guaranteeing. Thus it too is the kind of statement which could stand as the basis of a transcendental deduction.

It should be remembered, however, that when Kant describes a statement as transcendental he means no more than that its truth is a necessary condition of the possibility of experience, and not that it is thereby self-guaranteeing. Nevertheless, Kant, like Descartes, needs to base his philosophical conclusions on self-guaranteeing axioms if they are to be factual (i.e. non-analytic) as well as indubitable. In other words, only if his transcendental propositions are self-guaranteeing is he entitled to regard them as a priori-synthetic.

In his 'Deduction of the Pure Concepts of the Understanding' Kant claims that it is the 'I think' which is the true starting-point of philosophy. Our immediate task, then, is to examine the grounds for this belief, to see how it differs from the Cartesian approach which it superficially resembles, and to see whether the principle has any special claims to be treated as the ultimate premise in a transcendental deduction.

According to Kant, when the 'I think' is interpreted as a transcendental judgement, it is 'free from any empirical
admixture\textsuperscript{1} and is therefore in itself quite vacuous. It is therefore impossible to arrive at any interesting or important truths about the nature of the self by analyzing the constituent concepts of this proposition. But, as Kant points out, this is precisely what the rational psychologists, including Descartes, are trying to do. Indeed, it is their 'sole text' from which the whole of their teaching has to be developed.

An obvious objection to the view that the 'I think' is free from any empirical admixture is that it 'contains an inner experience'. Kant's answer to this is that 'inner experience in general ... in which no special distinction or empirical determination is given, is not to be regarded as empirical knowledge but as knowledge of the empirical in general'.\textsuperscript{2}

Knowledge of the empirical in general is, for Kant, knowledge of the necessary conditions of the possibility of experience; that is to say, transcendental knowledge.

Nevertheless, Kant could scarcely deny that the fact that he actually does think is a purely empirical one, for other people could coherently say of him 'he does not think'. He admits this:

It must ... seem strange that the condition under which alone I think, and which is therefore merely a property of myself as subject, should

\textsuperscript{1}A342/B400. \textsuperscript{2}A343/B401.
likewise be valid for everything which thinks, and that on a seemingly empirical proposition we can presume to base an apodeictic and universal judgement, namely, that that which thinks must, in all cases, be constituted as the voice of self-consciousness declares it to be constituted in my own self.¹

According to Kant, the reason why statements about thinking in general can be validly derived from statements about my own thinking is that 'we must assign to things, necessarily and a priori, all the properties that constitute the conditions under which alone we think them'.² In other words, whatever is a necessary condition of my being able to think is a necessary condition of thinking in general.

This should not of course be taken to mean that some particular fact about me is a necessary condition of thinking in general. For instance, although my existence is a necessary condition of my being able to think, it is clearly not a necessary condition of anyone else's being able to think. But rather, it means that the condition can always be stated in such a way as to be universally true; for example, in this case, that in order for any being to think it must exist.*

*For this reason, a universally self-guaranteeing statement (e.g. 'there is thinking') can always be derived from a referentially self-guaranteeing statement ('I think'). ¹A343/B401. ²A346/B404.
However, if Kant is not so much concerned with his own thinking as with the possibility of thinking in general, why does he take as his starting-point the seemingly subjective statement 'I think' rather than a more general statement like 'there is thinking' or 'something exists'.

The reason why the epistemologically correct starting-point is the 'I think' rather than the idea of thinking in general is, in Kant's words, that 'I cannot have any representation whatsoever of a thinking being, through any outer experience, but only through self-consciousness'.⁴ In other words, if I was not aware of my own thinking or my own states of consciousness I would not know what thinking or consciousness was, and therefore would not be able to ascribe thoughts or states of consciousness to other people. In short, I must be aware of my own existence as a thinking being before I can be aware of other thinking beings.

This could certainly be regarded as a reason for taking the 'I think' as one's starting-point rather than the idea of thinking in general, but as it stands, it is by no means a conclusive one. It could be attacked, or at least undermined, in a number of ways. For example, it might be argued that one's capacity for self-awareness is equally dependent on one's capacity for being aware of other people, and that

⁴A347/B405.
therefore the former has no logical priority over the latter.

This is more or less the line that P.F. Strawson takes in chapter three of Individuals. He argues that the ability to ascribe states of consciousness to others and the ability to ascribe states of consciousness to oneself are not different in kind, as has sometimes been thought (e.g. by the classical empiricists) but are merely different sides of the same coin. But if, as this view implies, awareness of oneself as a thinking being, and awareness of others as thinking beings are mutually dependent notions, then Kant's reason for taking the 'I think' (self-consciousness) as his starting-point is no longer a compelling one. Other reasons, however, can be adduced. But before examining these we must look more closely at Kant's own interpretation of the 'I think' principle, for it differs significantly from Descartes.

Having explained how we can base an apodictic and universal judgement on a 'seemingly empirical proposition', namely the proposition 'I think', Kant goes on to say that:

> the proposition 'I think' is ... here taken only problematically, not in so far as it may contain perception of an existent (the Cartesian cogito ergo sum), but in respect of its mere possibility, in order to see what properties applicable to its

\*i.e. the one involving direct (by acquaintance) knowledge, the other indirect (inferential) knowledge.

\[A347/B405\]
subject (be that subject actually existent or not) may follow from so simple a proposition. Thus, although Kant agrees with Descartes that the 'I think' is the true starting-point of philosophy, he regards the question of whether or not the 'I' stands for an existent of some kind as strictly speaking irrelevant. As we saw in Part I, this view is derived from Kant's theory of the nature of philosophy, and is fully in accordance with his actual procedure. Thus, he defines philosophy as 'knowledge obtained by reason from concepts' and claims that his own philosophical conclusions are 'deduced wholly in abstracto out of concepts'. Unlike Descartes, therefore, he does not attempt to base his philosophical system on a contingent truth about himself, namely that he exists as a thinking being, but on what he regards as a purely analytic truth, namely that the 'I think' must accompany all my representations.

He calls this the 'principle of the necessary unity of apperception' and claims that it is 'the highest principle in the whole sphere of human knowledge'. Kant regards this principle as 'an identical and therefore analytic proposition' because, for him, a representation is by definition any kind of inner state or process of which one is aware, whilst to be

---

1 Part I, section 5. 2 A713/B741. 3 Prolegomena, Lucas' edn. p. 34. 4 B135. 5 ibid. 6 ibid.
aware of something just is to think it.

If, however, Kant's conclusions were arrived at in this way, he would have no right to regard them as synthetic, for a synthetic truth cannot possibly be derived from an analytic truth.* Thus, if Kant is going to establish the indubitability of various statements of fact (the fundamental principles of Newtonian physics, for example) by means of his transcendental deductions, he must include statements of fact among his original axioms. And so, if he takes the principle of the necessary unity of apperception as his starting-point, he must show not only that its constituent concepts are instantiated in some way, but that they are indubitably instantiated. He is therefore wrong to dismiss the question of whether or not the 'I' in the 'I think' stands for an actual existent, as irrelevant.

As we have seen, there is only one way of showing that a concept is indubitably instantiated, and that is by showing that any attempt to deny that it was instantiated would be operationally self-refuting. It would be operationally self-refuting if the assertion 'there is nothing which is an x' necessarily presupposed the proposition 'there is something which is an x'. Broadly speaking, this is the kind of argument which Descartes uses in the second meditation to show that

*I have already suggested a number of reasons why Kant may have become confused on this point. See pp.62-64.*
the 'I think' is indubitable, and it is precisely the kind of argument which Kant should have used.

If the initial premise of a transcendental deduction should be established by means of a Cartesian-style argument, the conclusions, on the other hand, should be arrived at by means of a Kantian-style argument. In other words, in a transcendental deduction at least one of the axioms should be self-guaranteeing (the rest could be analytic), whilst the mode of inference should be presuppositional rather than implicational in form.

The reason for this is that the 'I think' is virtually devoid of content, as Kant realized,\(^1\) and therefore entails almost nothing. And the same is true of every other obviously self-guaranteeing statement; e.g. 'a language of some kind exists', 'some sentences are intelligible', 'I sometimes make assertions'. On the other hand, if a statement is not obviously self-guaranteeing it should not be treated as an axiom, but should first be derived from one which is.\(^2\)

Admittedly, the 'I think' is not totally devoid of content in the way that a 'degenerate'\(^3\) statement like 'I exist' is.* For 'think' is at least a descriptive concept, whereas 'I' and 'exist' are not. But 'I think' certainly does not entail

\(^*\)In so far as it is indubitable, although in an empirical context 'I exist' could be quite informative.


\(^2\)See pp. 146-147.

anything about the way in which I think (i.e. employ concepts), or about the nature of the world in general. It does not, for instance, entail the existence of a soul substance, as Descartes seemed to think.

On the other hand, a statement may entail virtually nothing, but presuppose a great deal. This is because the presuppositions of a statement are not part of its meaning, and therefore the fact that it says very little does not also imply that it presupposes very little.

For instance, some of the more immediate presuppositions of the statement 'I think' would include: the existence of a language containing general concepts and referring expressions, and the existence of a being capable of using words in a rule-governed way, of making assertions, behaving intelligently, and of saying things about himself.

Thus, there is no way in which I could say that I never used referring expressions without using one. Any denial of the existence of general concepts would have to contain general concepts. If I did not use words in a rule-governed way, nothing I said would make sense. If my assertion could not be classified as a piece of intelligent behaviour it could not be regarded as a genuine assertion, as opposed to the utterance of a sentence which might be used in a genuine assertion (parrots may say things but they do not make assertions). And finally, in saying that I think I am saying something about myself.
With this mode of inference special care should be taken not to read into the conclusions more than is strictly warranted by the argument. For instance, one should not take the self-guaranteeing statement 'I am capable of intelligent behaviour' as entailing 'I am conscious', unless it can be shown that consciousness is itself presupposed by the distinction between genuine assertions and mere parrot-like utterances.

In a transcendental deduction, one moves from one set of presuppositions to another; for instance, from the immediate presuppositions of the self-guaranteeing statement 'I think' to the presuppositions of these presuppositions. For example, as we have just seen, the 'I think' presupposes the existence of referring expressions which can be used in sentences to make true and meaningful assertions. But the existence of these referring expressions, in their turn, presuppose criteria for identifying individuals. Thus, I could not say anything of myself (e.g. 'I think') unless I was able to identify myself. Or, to put it another way, I could not talk intelligibly about myself unless I was distinguishable as an entity of some kind. Therefore, although the 'I' is not itself a descriptive concept, its use presupposes descriptive criteria of some kind.
One might go on to argue, as Strawson does,¹ that the criteria must be of a certain kind, and that certain other conditions must be satisfied if they are to be applicable; although it is at this point that Kerner's objections become relevant.²

Section 6. Referentially self-guaranteeing axioms.

We come now to the question of whether the 'I think' principle has any special claim to be regarded as the ultimate premise in a transcendental deduction, and as a corollary, whether it makes any difference if a reflexively self-guaranteeing statement is chosen in preference to one which is universally self-guaranteeing.

I will begin by dealing with Russell's objection that "'I think, therefore I am' says rather more than is strictly certain";³ that, in fact, we are only entitled to say that 'there is thinking, and not that 'I think' since the existence of a self which thinks is open to doubt. I shall then go on to show that there are perfectly sound logical, and epistemological reasons for using a referentially self-guaranteeing statement like 'I think' as the basis of a transcendental deduction, rather than one which is universally self-guaranteeing.

A statement is referentially self-guaranteeing if it cannot be coherently denied by the person making the assertion, but can (in the form of a logically equivalent statement) be coherently denied by someone else. For example, I cannot coherently deny that I think, although someone else could say of me 'he does not think'. A statement is universally self-guaranteeing if it cannot be coherently denied by anyone at all. For example, the proposition 'someone sometimes thinks' is undeniable in this sense, because asserting is a form of thinking, and therefore whenever anyone makes an assertion, it is true that someone thinks.

According to this account, the basic difference between a referentially self-guaranteeing statement and a universally self-guaranteeing statement is that whereas the former refers specifically to the person making the assertion, the latter does not. A referentially self-guaranteeing statement must therefore contain a referring expression of some kind. But, according to some,¹ the statement cannot be indubitable if it contains such an expression.

In order for the inclusion of a referring expression in a self-guaranteeing statement to be justified, it must somehow logically guarantee the existence of the thing referred to. For if it did not, the statement would be contingently rather

¹e.g. Bertrand Russell, Op.cit. (above, p. 185, note 3).
than logically self-referring, and a statement which just
happens to be self-referring cannot be self-guaranteeing.¹
For example, a statement like 'the most isolated person in
Australia sometimes makes assertions' would not be self-
guaranteeing even if the descriptive phrase 'the most isolated
person in Australia' happened to refer to the very person
making the assertion.

Another point about this example, is that there is no
way in which the descriptive content in a referring expression
could guarantee the existence of the thing referred to. For,
as Kant himself realized, we cannot prove the existence of
something merely from the fact that we have a concept of it.
Such a move would only be possible if 'existence' was itself
a descriptive concept, but it is not. As Kant says 'it is
not a concept of something which could be added to the concept
of a thing'.² Thus, if a referring expression is to guarantee
the existence of the object referred to, it must do so in some
other way than by describing it.

In his Enquiry into Meaning and Truth, Bertrand Russell
recognizes the existence of a class of referring expressions
which, as he puts it, 'merely designate an object without in
any way describing it'.³ (This is a somewhat misleading
characterization as we shall see later on). Included in this

¹See Part II, section 1. ²A598/B626. ³Ch.VI, 'Proper Names', p.102.
class are such terms as 'this', 'that', 'I', 'you', 'we' and so on.

All these words have a demonstrative rather than a descriptive function. They point but do not classify (and so in that sense are not concepts). Unlike definite descriptions 'their denotation is relative to the speaker'.¹ In other words, they fulfill their function by using the speaker himself as a reference-point. For this reason, Russell describes them as 'egocentric particulars'.²

Not every member of the class of egocentric particulars can be used as the subject of a self-guaranteeing statement, although only members of that class can be used. The reason is simply that in most cases the question of whether or not an egocentric particular refers to an actual existent is a purely contingent one. Thus, terms like 'you' and 'this' may sometimes not refer to anything at all. For instance, a child might address an imaginary companion as 'you', or an unclothed emperor might say 'this is a fine garment I have on'. Statements in which demonstratives are used in this way cannot be self-guaranteeing; e.g. 'you are able to make assertions'.

The situation is somewhat different in the case of exclusively self-referring statements (e.g. 'this is a true proposition'), for here the term 'this' is used to refer to

¹ibid, p. 102. ²ibid.
the very statement in which it occurs, and it might therefore be argued that, in this case at least, it cannot fail to have a reference. However, as we have already seen, there is good reason for supposing that such 'statements' assert nothing, and are therefore meaningless, in which case they could not of course be self-guaranteeing.

On the other hand, some egocentric particulars cannot be used at all (in direct as opposed to reported speech) unless the thing they refer to actually exists. Thus, in a sense, their use guarantees their reference. This is the case whenever the speaker refers to himself by means of an egocentric particular; e.g. 'I', 'me', 'myself'. If, for example, I use the personal pronoun 'I' it follows that I must exist, whereas if I use the pronoun 'you' it does not follow that you must exist. I shall term this sub-class of egocentric particulars 'ego-indicators'. A referentially self-guaranteeing statement may therefore be defined as a self-guaranteeing statement in which the speaker refers to himself by means of an ego-indicator of some kind; e.g. 'I think'.

If the ego-indicator is omitted from a referentially self-guaranteeing statement, it becomes universally self-

"'we' and 'us' may also, on occasion, be classified as ego-indicators. Thus, 'some of us think' is referentially self-guaranteeing, when 'us' refers to a limited class of individuals of which I am a member.

1 See Part II, section 3.
guaranteeing. Thus, a universally self-guaranteeing statement (e.g. 'true and meaningful assertions are possible') can always be derived from a referentially self-guaranteeing statement (e.g. 'I am able to make true and meaningful assertions') but never the other way round. In other words, a universally self-guaranteeing statement does not presuppose the existence of any particular person in the way that a referentially self-guaranteeing statement does.

On the other hand, both types of statement are validated in precisely the same way, namely by the mere fact of their being asserted. Thus, although the proposition 'true and meaningful assertions are possible' cannot be coherently denied by anyone, and is necessarily presupposed in all our assertions, it is self-guaranteeing only because it is actually asserted by someone. But as a proposition it could well be false.

One reason for using a referentially self-guaranteeing statement as the basis of a transcendental deduction is to draw attention to the fact that it is only by asserting the proposition that I am entitled to regard it as self-guaranteeing.

However, if this were the only advantage that a referentially self-guaranteeing statement had over a universally self-guaranteeing statement, the difference between them would be so slight as to be virtually irrelevant from a philosophical point of view.
However, in at least some cases the distinction is far from being irrelevant. For example, a person who tries to prove that the external world exists must assume, at least provisionally, that its existence is somehow open to doubt, otherwise the question of how it could be proved would not even arise. This means that he cannot afford to treat any statement about the external world as axiomatic. Thus, as Descartes realized, he is left with only one alternative, namely, to start from a statement about himself; not, of course, as an object in the external world, but as a subject of inner experience.* The Cartesian 'I think' is a good example of this.

We may therefore conclude that if one is going to prove the existence of the external world by means of a transcendental deduction (assuming for the moment that this is possible), one must take as one's starting-point a statement which is not only self-guaranteeing, but also, as in the example above, subject-referring. Obviously, in this case, the referential form is the most appropriate.

One might add that since one is not entitled to take anything for granted in a transcendental deduction unless it can first be shown to be self-guaranteeing, one should always start by doubting the existence of the external world, for its existence is, at best, unobviously self-guaranteeing, and, at

*There are, of course, problems with this, as we shall see later on.
worst, not self-guaranteeing at all. Statements about one's own existence, on the other hand, are clearly self-guaranteeing. One's initial premise, therefore, should always be referential in form."

In using a transcendental deduction to justify our belief in an external world, the trick, as Kant describes it, is to turn the game played by idealism against itself by proving that 'even inner experience, which for Descartes is indubitable, is possible only on the assumption of outer experience'. In other words (according to Kant) I can only be aware of myself as determined in time through my awareness of objects existing in space, and therefore the fact that I am indubitably aware of myself as a subject of inner experience proves that a spatially determined world of objects also exists.

In order for this argument to count as a genuine transcendental deduction it must first be shown that statements like 'I am aware of myself as determined in time' and 'there is such a thing as inner experience' are self-guaranteeing, and therefore able to function as transcendental axioms.

*Barry Stroud has argued (see pp.73-76 of this thesis) that the conclusions of all transcendental deductions should be of the form 'I believe that ...' or 'it seems to me as if ...'. If he is right, then not only the axioms, but also the conclusions of a transcendental deduction should be referential in form.

'B275, B276 'Note 1'.
If these statements were not self-guaranteeing the argument might still be valid; not, however, as a transcendental deduction, but as a transcendental exposition. For a successful transcendental exposition would show that we could only accept one class of statements as true (e.g. statements about our inner experience) if we were prepared to accept another class of statements as true (e.g. statements about the external world). But it would presumably still be possible to doubt the existence of inner experience and therefore of the external world. In other words, a transcendental exposition is not designed to refute the sceptic's conclusion that we cannot know that the external world exists, but only the argument on which it is based. A transcendental deduction, on the other hand, is supposed to refute the conclusion, and must therefore rely on the indubitability of its initial premise.

The same is true of Strawson's argument in chapter 3 of *Individuals*, where he tries to prove that since we can ascribe states of consciousness to ourselves, we must also be able to ascribe them to other people. In order for the argument to count as a transcendental deduction rather than as a transcendental exposition it must be shown that we cannot coherently deny that we are able to ascribe states of consciousness to ourselves. If this were not done, the most

---

1 See Part I, section 1.
the argument could show is that all predicates ascribing states of consciousness were meaningless unless they could be applied to other people. This might trouble the solipsist, but it would not trouble either the behaviourist or the physicalist, for their aim is to eliminate such concepts altogether.

It may be that transcendental deductions, as we have described them, could not be used to refute any of the sceptic's theories, in which case their philosophical importance would be considerably diminished. However, it is not one of the aims of the present work to determine the scope of the transcendental deduction as an instrument of philosophical enquiry. Its chief aim has been to find out whether a transcendental deduction, as conceived by Kant, could even get off the ground; in other words, whether any sense could be made of the notion of an a priori-synthetic principle. Nevertheless, it is worth considering for a moment whether there is any possibility of refuting the sceptic by means of a genuine transcendental deduction.

Kant's refutation of idealism is based on the axiom 'I am conscious of my own existence as determined in time.' If the argument is to count as a transcendental deduction this axiom must be self-guaranteeing. But is it?

As we have seen, the self-guaranteeing statement 'I think' (i.e. employ concepts) presupposes an ability to talk about

\[^{1}\text{Op.cit. (above, p.192, note 1).}\]
myself (i.e. make assertions about myself), and so it could be said to presuppose an awareness of myself, in a somewhat attenuated sense of 'aware'. We are not, however, entitled to assume, at this stage, that I am consciously aware of myself, or that I am aware of myself 'as determined in time'.

If it could be shown that the concept of time was built into the very concept of an assertion, we would at least be entitled to take its existence for granted, for we could not coherently assert that time did not exist.

I could not assert anything if I was unable to identify my assertion as an assertion. A parrot does not assert because it does not know that it is asserting. The criteria for identifying an assertion is part of the very meaning of 'an assertion'. Thus, an assertion must be identified as a temporal event just because it is a temporal event. A non-temporal assertion would not be an assertion but something completely different, just as an unlocatable sensation would not be a sensation but something completely different. Thus, one can only meaningfully talk about an assertion if one is prepared to accept the existence of temporal events. Hence the assertion 'temporal events are possible' (i.e. the conditions of their possibility exist) is operationally self-guaranteeing.

At this point, one could use a Kant-Strawson style argument to show that the occurrence of temporal events presupposes
(the representation of) a world of persisting objects locatable within a single spatio-temporal framework. In other words, in order to identify temporal events (like assertions) I must be able to identify persisting objects in space.

In order to identify persisting objects I must myself persist, for I must be able to re-identify these items at different periods of their existence. In other words, the statement 'I am (non-consciously) aware of myself as determined in time' is operationally self-guaranteeing.

I must confess that I see no way in which the concept of consciousness could be built into this statement, but since it is not in any way essential to the argument, its omission is not serious. On the other hand, if the statement was to be used as the basis of Strawson's argument in chapter three of *Individuals*, it would be necessary.

Although the ego-indicator 'I' merely 'designates an object without in any way describing it', its use does presuppose descriptive criteria of some kind. If this version of Kant's refutation of idealism is accepted, it follows that its use presupposes criteria for identifying the self as a persisting object; an object, moreover, which is capable of identifying objects and events in an external world.

---

Even if the argument is not accepted, it at least serves to illustrate how the 'I think' might be used as the basis of a transcendental deduction.

Section 7. The conclusion.

The revised notion of a transcendental deduction at which we have now arrived has much in common with Kant's original conception.

In the first place, it is a method of justification* rather than of proof. For it deals with a class of statements which, in a sense, cannot be proved, and, some might add,† do not need to be proved. It is designed to show that we are logically justified in taking certain concepts and principles for granted on the grounds that, although they might not apply to the world as it really is in itself, we have no alternative but to accept them; that is, if we are to have any explicitly formulated beliefs at all. There can surely be no stronger justification of a principle than to show that without it we could never assert anything.

*For Kant, the term 'deduction' has a legal or juristic connotation. It is concerned with the vindication of a right, rather than with establishing a fact; namely, our right to use certain concepts or principles. See A84/B117.

†e.g. A.J. Ayer argues that the principle of induction not only cannot be proved, but does not need to be proved, for it is itself a canon of reason. See The Problem of Knowledge, p.75.
It also looks as though Kant is right in claiming that the transcendental method provides us with our only access to metaphysical truth: at least, if we take 'metaphysical truth' to mean 'synthetic truth which can be established on logical grounds alone.' For the only way we could guarantee the truth of an ultimate axiom without depriving it of its axiomatic status would be to show that it was in some sense self-guaranteeing. As Aquinas says 'that through which all the rest can be known must itself be self-evident'. But this is surely the only internal proof of self-evidence that we could have.

The only obvious alternative would be to abandon the quest for a logical guarantee altogether, and to adopt the view that ultimate principles and axioms must be accepted or rejected on pragmatic grounds alone. For instance, it might be argued that if a viable scientific system can be based on a particular set of axioms then we have good reason for accepting them, even though the system may one day be replaced by an equally viable system based on a quite different set of axioms.

To say that the transcendental method provides us with our only access to metaphysical truth does not mean that it provides us with the key to the solution of all so-called metaphysical problems. For these problems could have arisen in any number of ways. For instance, they could, as

Wittgenstein claims, \(^1\) simply represent a state of conceptual confusion, in which case to clear up the confusion would be to clear up the problem.

Even if the traditional problems of philosophy are non-empirical it does not mean that they must be solved by non-empirical means. It might, for instance, be argued that doubts about the reliability of the principle of induction are neurotic and therefore do not deserve to be taken seriously. Thus, the philosopher who claims that we can never be certain that the sun will rise tomorrow even though it has always risen in the past, may be compared to the man who keeps getting up in the night to make sure that the back-door is locked, even though it always is. Such doubts, it has been said, are irrational, and therefore cannot be dispelled by rational argument.

There is, however, an important difference between the two cases, for the philosopher's doubts would certainly be resolved if it could be proved to him that any attempt to deny the principle of induction would be operationally self-refuting, whereas in the case of a genuine neurotic they would not.

Whether or not any of the sceptic's theories could be refuted in this way is questionable, to say the least. On the other hand, it has not yet been shown that transcendental deductions of this kind are impossible. Korner may well be

right in claiming that the uniqueness of a particular categorial schema cannot be demonstrated on internal evidence alone. But he is not entitled to claim that it is impossible to demonstrate the indubitability (for us) of any statement of fact on internal evidence alone, for operationally self-guaranteeing statements are indubitable in precisely this sense.

Transcendental deductions, as I have described them, are concerned with those states of affairs which must exist if statement-making in general is to be possible. It is not, therefore, exclusively or even mainly concerned with the indispensability of certain concepts or conceptual schemas. Thus, the presuppositions of discourse need not always be explicitly formulated.

In other words, a transcendental deduction may sometimes only show that we are entitled to use a certain concept or principle, not that we must do so. For instance, we could perhaps do without a concept of causality, but if the actual conditions which make a concept of causality possible did not exist, statement-making in general would be impossible. For if there was no degree of regularity or uniformity in the way things happened, all our 'assertions' would be meaningless.

This is because in a world of total Heraclitean flux, items could never be brought under general concepts and hence could

\[1\text{Op.cit. (above, p.26, note 2).}\]
not be identified as items of a certain kind, or referred to in any way. And this means that statement-makers could not be identified as statement-makers and assertions could not be identified as assertions. Indeed, the twin concepts of a persisting object and a temporal event would be inapplicable in this context.\(^*\) Transcendental deductions of this kind are designed to show that certain states of affairs are unconceptualizable in the sense that if they existed they could not be brought under concepts, and therefore could not be coherently asserted to exist.

Barry Stroud would object that such arguments are impossible because self-guaranteeing statements are too vague to lead to any interesting conclusions.\(^1\) But, as we have seen,\(^2\) this might not be true. For a statement may say (i.e. entail) very little, but presuppose quite a lot, and may even presuppose the falsity of some of the sceptic's theories. At any rate, even if transcendental deductions prove less than one would like, they prove (or are capable of proving) more than Stroud was prepared to admit.

There have been many attacks on Kant's claim that the transcendental method is 'entirely different from any hitherto

\(^*\)The argument has been deliberately abbreviated, since it is only being quoted for purposes of illustration.

\(^{1}\)Op.cit. (p.72, note 1).  \(^{2}\)p.183.
some of which have been examined in this thesis. Norman Kemp-Smith, for instance, has argued that the transcendental method was 'really identical in character with the hypothetical method of the natural sciences. It proceeds by enquiring what conditions must be postulated in order that the admittedly given may be explained and accounted for.' But the conclusions of a transcendental deduction, as we have defined it, are categorical (i.e. apodeictic) in form rather than hypothetical. Kemp-Smith's description only applies to transcendental expositions. 

Strawson and Bennett have argued that the transcendental method is really only a form of conceptual analysis, and therefore in no sense unique. But the transcendental propositions are, on our interpretation, a priori-synthetic, and not, as their view implies, analytic.

Barry Stroud has claimed that transcendental arguments are just another way of applying the verification principle, and that therefore 'there was nothing special or unique, and certainly nothing new' about them. But again, we have shown that it is possible to construct transcendental deductions which do not invoke the verification principle.

---

On the other hand, it must be admitted that Kant's claim that no philosopher has ever attempted a transcendental deduction before is not altogether justified. We have already seen that at least two of the arguments which Aquinas describes in the *Summa Contra Gentiles* are primitive, but recognizable attempts at a transcendental deduction of the concept of God.

A more sophisticated type of transcendental deduction can be found in Book K, chapter five, of the *Metaphysics*, where Aristotle tries to show that we are justified in taking the law of contradiction for granted. He begins by explaining that although certain truths cannot be proved absolutely, because there is no more certain premise from which they may be inferred, they may nevertheless be proved *ad hominem* against anyone who makes contradictory assertions. He goes on to say that 'the method is to wring from our opponent some admission which is identical with the truth in question without appearing so to him'.

In the course of a fairly long argument Aristotle points out that the law of contradiction is a necessary condition of the intelligibility of all our assertions, and indeed of the very words we use, and so Heraclitus can only deny the law of contradiction by taking it for granted. Thus, 'if his assertion

\[\text{\textsuperscript{1}Op.cit. (p.167, note 1 & p.168, note 2).}\]
that the same thing can at the same time be and not be, were true, even it itself would not be true'. Aristotle concludes that 'if one had argued thus with Heraclitus, one might have obliged him to recognize the law of contradiction; for he evidently adopted his view without understanding what it involved'.

Aristotle, I believe, over stresses the ad hominem nature of the argument. For if it could be shown that the law of contradiction (or some generalized version of it) is a necessary condition of the intelligibility of any utterance, then Heraclitus would be conclusively refuted. For there would be no way in which he could coherently assert that the law of contradiction was invalid or untrue.
BIBLIOGRAPHY

Books:


*Summa Theologiae*, Blackfriars, 1964


Austin, J.L. *Philosophical Papers*, O.U.P. 1961


Bennett, J. *Kant's Analytic*, Cambridge University Press, 1966

Collingwood, R.G. *Essay on Metaphysics*, O.U.P. 1940

Descartes, R. *Meditations on the First Philosophy*, Everyman’s library, 1962

Kant, I. *Critique of Pure Reason*, Norman Kemp-Smith, Macmillan 1963


Kemp-Smith, N. *A Commentary on Kant’s ‘Critique of Pure Reason’*, London, 1918

Korner, S. *Kant*, Pelican Books, 1960


Passmore, J. *Philosophical Reasoning*, Duckworth (paperback) 1961

Quine, W.V.O. *From a Logical Point of View*, Harvard, 1961


Russell, B. *The Problems of Philosophy*, (Home University Library) O.U.P. 1962

Strawson, P.F. *The Bounds of Sense*, Methuen, 1966


*Tractatus Logico-Philosophicus*, Routledge & Kegan Paul, 1961

**Articles:**


Jorgensen, J. 'Some reflections on reflexivity', *Mind* 1953

Korner, S. 'The impossibility of transcendental deductions', *The Monist*, July 1967

Passmore, J. 'The idea of a philosophical method', *The Encyclopedia of Philosophy* (p. 218), 1966

Popper, K. 'Self-referring statements', *Mind*, 1954


Stroud, B. 'Transcendental arguments', *May* 1968

Williams, B. 'Descartes', *Encyclopedia of Philosophy*, 1966