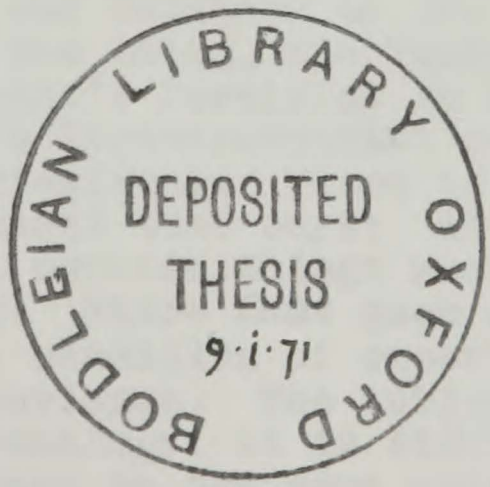
KANT'S PHILOSOPHICAL METHOD

Ralph C.S. Walker

D. Phil. Thesis
and Abstract

Merton College

(MT 1970)



A B S T R A C T

Introduction pp. 1-4.

We are faced with a choice between three philosophical methods: the examination of natural languages to find what features they share; piecemeal conceptual analysis; and conceptual analysis which justifies the interest of its subject-matter by setting out conditions for experience. The third method is Kant's.

I. Transcendental Arguments pp. 5-206.

A. "Transcendental" pp. 5-21.

There are difficulties over Kant's use of "transcendental"; but it is clear that he holds a transcendental argument to be one which sets forth necessary conditions of there being any knowledge, or experience, at all. This use of the word grows out of Wolff's.

B. The Transcendental Deduction of the Categories pp. 21-67.

Transcendental deductions show that particular concepts can legitimately be used since otherwise there could be no knowledge. The transcendental deduction of the categories is Kant's most important transcendental argument, so it is worth considering, to see how he proceeds. It has nothing to do with criteria for personal identity. Its aim is to show that experience must be subject to the categories, and one can distinguish three strands of argument. The latest version turns on the notion of a judgement; it is straightforward, but rather trivial. Another proceeds by showing how the concept of an object, or of a subject, requires synthesis: if the categories are forms of synthetic judgement we have the same triviality as before, and if not their status, and relation to the forms of judgement, remains unclear. The third, the "subjective deduction", investigates the mind's faculties in knowing; it involves what Strawson calls "transcendental psychology" but should not therefore be dismissed. There is nothing in principle wrong with an argument that says: if knowledge/experience is to be possible, certain things must be true about the mind of the knower. Where Kant goes wrong is in supposing that any necessary condition of experience must be imposed by the mind on experience. The subjective deduction is not really faculty-psychology; it is similar to the second line of argument, and open to the same criticism.

C. General Criticisms of the Deduction pp. 67-78.

Kant is analysing the concept of experience; his argument proceeds by analytic steps. - It is hard to see how he could have believed them synthetic a priori; there is no evidence that he did, and it would not fit his account of synthetic a priori judgements.

Yet his argument is not circular: the conclusion is not simply built into the notion of knowledge/experience from which he starts.

Transcendental arguments do not require us to get outside our own conceptual scheme. They show that experience without certain conditions would be impossible, and do not claim that it would be possible with no other conditions.

But the most serious objection to transcendental argument is that its conclusions are always trivial. Some plausible transcendental arguments have recently been put forward, purporting to establish non-trivial conclusions. Are they valid?

D. Some Transcendental Arguments pp. 78-159.

Given that we must be aware of persisting objects, Strawson argues that experience must be causally ordered. We must be able to make counterfactual-supporting regularity-generalisations; but it has not been shown that these must support predictions, so they may not have to be causal generalisations. (And no limitation is imposed on the world, for since what we see as similar is a matter of psychology, what we see as regular is so too). Strawson also has an argument that in any experience we must be aware of persisting objects, in order to distinguish objective from subjective time-order. But could not patterns of events fulfil the same function?

Experience must be temporally ordered; it has not been proved that it must be spatially ordered, or ordered on any analogous dimension. A purely temporal world could be far from chaotic.

The argument against private languages is transcendental. Five objections to it are considered. Three are rejected: the conclusion is not trivial; terms cannot have public sense and private reference; nor can the argument be avoided by introducing private terms into a public language. Two are accepted: to get evidence independent of what seems to one to be the case it is not necessary to apply to others, it can be found in natural regularities; nor need we even appeal to these, for our own memories can check one another in any sense in which other people's can check them. Wittgenstein replies that the meaning of a word is the way it is used, and if there can only be one speaker there is no possibility of misuse. - But a linguistic community may come to use a word in a new way; in understanding records one must know what the

words meant, for which one has to rely on remembered and not current usage, just as the speaker of a private language does. This memory must not merely correlate words with words, but either carry out the correlation in such a way as to preserve the majority of contingent truths, or - and this seems the better account - correlate words with memory-images.

Does the need for an is/seems distinction show that one must have the idea of an unperceived existent? I must think of myself as persisting, and admit the logical possibility of there being other subjects independent of me; but I might never be aware of any such subjects, and the idea that I might be might never occur to me. Nor can it even be argued that factual, as well as verbal, mistakes are always possible. So the conclusion that we must apply to our experience counterfactual-supporting regularity-generalisations cannot be drawn, for it depended on the assumption that we had to apply the idea of an unperceived existent.

Can we get more interesting conclusions by building in that the language/experience in which we are interested is that of a self-conscious being? Different things can be meant by "self-conscious", but they do not get us very far. To have an intersubjective communication-system in which one can refer to one's interlocutor one must make counterfactual-supporting regularity-generalisations.

Such conclusions as we have reached have been disappointing, largely because we expected too much. They are not all entirely trivial, for they have been denied by competent philosophers. But perhaps more exciting results can ~~by~~ be got by looking for the conditions of experience, not of any (self-conscious) being, but of beings of some further specified sort.

E. Schematism pp. 160-206.

The most important thing Kant does in the chapter on schematism is to introduce a further restriction into the premises of his transcendental arguments. The Transcendental Aesthetic showed that our experience is spatio-temporal, but not that experience of any other sort would be impossible; so now, instead of looking for the conditions of any experience, he examines the conditions of spatio-temporal experience. This might prove fruitful, but Kant's arguments are not very successful. The idea, however, that a necessary condition of all knowledge or experience will still be a necessary condition of knowledge or experience of a more restricted kind, and may apply differently in different specific kinds, is valuable.

Transcendental arguments can be used against sceptics to show that their position undermines its own intelligibility; we cannot build too much into the characterization of

knowledge/experience from which we start, or the sceptic can deny that his knowledge/experience is of that sort. We can counter this denial if it is the only sort of knowledge/experience available to human beings: "our conceptual scheme". For if his doubt presupposes some conceptual scheme closed to us, it is idle. But even if we could change our conceptual scheme, the only grounds for doing so could be pragmatic; and our own seems preferable on these grounds to any alternative. The sceptic can reply that the alternatives might seem simpler to other beings; but such different ways of classification really are not accessible to us.

So transcendental arguments from restricted starting-points are of interest when the sceptic can only avoid admitting the unintelligibility of his own position by claiming to retreat to a conceptual scheme not available to us, so that his doubt loses interest. And they exhibit how the general conditions of any experience work in human experience.

II. Metaphysical Arguments pp. 207-249.

Kant's use of the term "metaphysics" fluctuates. The Metaphysische Anfangsgründe der Naturwissenschaft analyses the concept of movable matter. Metaphysics exhibits synthetic a priori relations between concepts, and so must rely on arguments about conditions under which experience is possible; but it deals with empirical concepts (unlike KdrV). By calling the concept of matter empirical Kant must at least mean that it cannot be transcendentally deduced. It is, certainly, often said that Kant was attempting to prove that (an amended) Newtonian physics must apply to any spatio-temporal experience. But he did not even deny that there might be experience of a spatial world in which there was nothing movable. Those who think he did give no adequate account of what he meant by calling the concept of movable matter empirical. Nor do Kant's Critical writings support their interpretation.

One might say that Kant should have denied this possibility, because such a world could have no place for unperceived existents. We can avoid this difficulty by introducing a new parameter (intensity, e.g.); but would such a world count as both static and spatial? Enough of the appropriate similarities to our familiar world remain for us to say so.

But how can an argument employing an empirical concept have a synthetic a priori conclusion? Kant's conception of the synthetic a priori splits apart here, and we have a concept acquired by abstraction, not mind-imposed, which is nevertheless necessary for experience like ours. And this perhaps marks the difference between metaphysical and

transcendental arguments, for Kant: transcendental arguments show that certain forms are mind-imposed and so unavoidable for us, metaphysical ones only exhibit general conditions for experience.

The MAdN thus really schematises further the schematised categories; no radical change of method is involved. It is no surprise to find Kant's attempted proofs here again invalid.

The metaphysical deduction and expositions in KdrV are of a different type. Metaphysical expositions show that human experience must (as a matter of psychological fact) be of a certain sort; the metaphysical deduction of the categories exhibits their a priori origin.

Besides arguing for synthetic a priori truths, Kant also tries to establish regulative principles, which only set out rules for the understanding. He does not regard these as justified by fruitfulness: they must be unavoidable for reason. He tries to show this in various ways; sometimes it is unclear whether a transcendental argument is meant to support a constitutive, or only a regulative, principle.

III. The Flight from Extension pp. 250-326.

The most general criticism of this whole method of doing philosophy is that the notion of analyticity lacks sense. It belongs to a circle of interdefinable terms, out of which it seems impossible to break. Such circles are not vicious if concepts within them are transcendentally necessary. In this case, no doubt there could be language-users without these concepts, but perhaps they are indispensable for any account of language-use. Quine's account of it tries to avoid them, but has difficulties. Can such an extensional account be given?

The word "extensional" can be used to mark three different, though analogous, features of systems: truth-functionality, referential transparency, and predicate-extensionality. It is not obvious that systems extensional in one of these ways must be so in the others.

An extensional account of language-use could employ the notion of (efficient) causality, or something very like it. It seems reasonably natural to take many of our causal statements as transparent and predicate-extensional; and the truth-functionalist can at least express a relation sufficiently similar to causality to do a lot of the same explanatory work.

Unless the extensionalist is to hold that our usual teleological, and hence intensional, explanations of human behaviour are always quite wrong - which seems altogether implausible - he must offer some reduction of them to (extensionally tractable) causal explanations. Nagel and

Noble argue that any teleological explanation is logically equivalent to a non-teleological one, since the state of affairs mentioned in the antecedent of the hypothetical can always be described without mentioning a goal. But it may not be possible to identify the set of states co-classified by the antecedent of a general teleological hypothetical by means of any finite set of intrinsic descriptions. Teleological explanation is not vacuous; nor reducible to backwards-causal explanation.

But it is possible that the same behaviour be explicable in both ways. This might be because two quite different generalisations both happened to apply to a particular item. But if the extensionalist admits that though strictly unintelligible our ordinary explanations of human behaviour are not entirely wrong, he must rather argue that the two generalisations should cover the same ground, the antecedents of each being true in the same circumstances. It is possible to draw up, in terms of negative feedback, an efficient-causal explanation which can bear this relation to any teleological explanation; but whether it is correct to give the causal explanation is an empirical matter. If human behaviour turns out to be explicable in this way the extensionalist cannot be accused of failing to explain it, and conceptual economy supports his rejection of the (equally useful but unnecessary) purposive way of talking. And if human language-use is explicable non-teleologically, our languages can't really contain intensional terms at all.

Quine himself grants that in translation we have to make assumptions about synonymy; cf. Kant on regulative ideas. There is some truth in his thesis of indeterminacy of translation; but we can sometimes interpret the desires and beliefs of others unambiguously, so that there should be no difficulty in principle in telling that two expressions are used for the same linguistic purpose.

To say that the remaining issue is empirical, and to express a faith that the facts will favour analyticity against Quine, seems the best one can do.

F O R E W O R D

This thesis began as an attempt to see how Kant's thought developed out of Leibniz's. It became increasingly clear, however, that there was too much to be said on this subject for a work of this length, and also that most of it would have to be fairly detailed and rather limited in its philosophical interest. Furthermore, a good deal of it has already been done, in one form or another, though not much is available in English. Kant's most important innovation was his way of doing philosophy, a method which could be said to reconcile rationalism and empiricism in that it proceeded by looking for the necessary conditions of experience. The historical antecedents of this method are interesting; there can be little doubt that he derived it from his reading of Wolff and of Leibniz's Nouveaux Essais, which were published by Raspe in 1765 and almost certainly provided the stimulus for the revolution in Kant's thought which occurred four years later and let to the Inaugural Dissertation and the development of the Critical philosophy. But more interesting than the historical question seemed the question of the value of Kant's method of doing philosophy. And it is with that question that this thesis is concerned.

Any adequate list of acknowledgements would be a long one. Therefore I shall not attempt to give one. But my thanks are due most of all to Mr. H.H. Cox, my supervisor; and I am also very grateful to Mr. N.J.H. Dent and Dr. J.D. Kenyon for a large number of valuable discussions.

Quotations from Kant's works I have generally left in German. Where I have quoted in English from the Critique of Pure Reason I have sometimes used Kemp Smith's translation, but most translations are my own.

The first part of chapter II, in a rather different form, is to be published in the proceedings of the Third International Kant Congress.

Introduction

Philosophizing in what is taken to be a Kantian way has been popular in Germany since the 1780's. In England it has been practised more sporadically, but at present it is very much on the increase. Very different opinions have been held as to what arguments, or what general types of argument, may fairly be ascribed to Kant; no common strand runs through the work of all who have called themselves Kantians. My aim is partly to adduce evidence for the ascription of certain general types of philosophical argument to Kant; but also, and mainly, to evaluate these types of argument and this manner of doing philosophy. In general I shall seek to defend Kant's method, though not his particular conclusions; but the defence may well be considered to result at times in a rather Pyrrhic victory, if indeed in a victory at all.

In speaking of Kant's philosophical method, I have in mind particularly the types of argument he uses in the Critique of Pure Reason and other works in speculative or theoretical philosophy in his Critical period - which I take to run roughly from the publication of the first edition of the Critique in 1781 to about 1796, and thus to exclude most of the Opus Postumum, which I would classify as Post-Critical and would be inclined to dismiss as senile. The types of argument used in the same period in practical philosophy

are the same, as Kant's language would suggest; but I shall not be specially concerned with examining the forms in which they appear there.

There are two main sorts of arguments which Kant considers philosophical: transcendental arguments and metaphysical arguments. The status and function of these is less clear; roughly, they are different sorts of argument to the effect that something is a necessary condition of there being experience or language. Such arguments were used occasionally before Kant[1], but he was the first man to think self-consciously about his use of them and deliberately to build his philosophy with them. Contemporary interest in Kant in England is largely due to the hope that important conclusions may be established by an examination of the necessary conditions of there being any language at all - conclusions like, for example, that of the argument against private languages; in part also to a feeling that this is the only way in which philosophers can proceed if they are to do more than examine particular concepts which may be dispensable, or features which as a matter of fact natural

1. Most clearly perhaps by Christian Wolff: Philosophia Prima sive Ontologia (new ed., Frankfurt and Leipzig 1736), sects. 77, 499, 500.

languages share.[1] Roughly, three different philosophical methods present themselves, and it is not easy to decide between them: one is the Kantian method using transcendental arguments; the second is the method of conceptual analysis; the third is the method of those who, like Katz[2], reject as unintelligible the notion of a necessary condition, which is central to both the other methods, and concern themselves with questions about how words are actually used and what general features natural languages happen to share. I shall be examining some considerations relevant to deciding between these three alternative ways of doing philosophy.

The notion of analyticity is fundamental to Kant's work. Metaphysical and transcendental arguments proceed by means of a series of propositions which are, or should be, analytic. It is disputed whether Kant himself recognized this, but I am strongly inclined to think he did. He was not the first to draw the analytic/synthetic distinction, though the

1. It stands therefore in the tradition of the neo-Kantians rather than of those contemporary Germans who, following the lead of Wundt, Heimsoeth, and Heidegger, claim that the interest of Kant's work lies in his "ontology", i.e. his conclusions regarding metaphysics, and who consequently see the Analytic as merely preparatory to the Dialectic. The difference is one of emphasis, but transcendental arguments are worth more attention than the "ontologists" are prepared to give.
2. J.J. Katz, The Philosophy of Language, New York, c. 1966.

terminology is his; nor can he legitimately claim, as with transcendental arguments, to be the first to use it self-consciously and make it fundamental to his work. Leibniz had preceded him. But Leibniz so misused the distinction he drew that perhaps we can call Kant the first person to handle it properly.

In what follows, I shall first discuss transcendental arguments, and then metaphysical arguments as they appear in the main theoretical work that uses them, Die Metaphysische Anfangsgründe der Naturwissenschaft. I shall try to show that these arguments are of some use, though perhaps they do not support all the conclusions they have sometimes been thought to establish. They are a weapon against scepticism, in a way that the simple analysis of a given concept can never be: the sceptic cannot simply shrug his shoulders and talk in some other way. Finally, I shall return to the notion of analyticity which underlies both sorts of argument, to consider the objection that it is unintelligible. And here, I shall argue, philosophical defence must give way to empirical faith: the question of its intelligibility rests on an empirical issue we are not in a position to settle.

I. Transcendental Arguments.

A. "Transcendental".

In the Critique of Pure Reason the term "transcendental" is introduced with these words:

> Ich nenne alle Erkenntnis transzendental, die sich nicht sowohl mit Gegenständen, sondern mit unserer Erkenntnisart von Gegenständen, insofern diese a priori möglich sein soll, überhaupt beschäftigt. (B25.)[1]

This requires clarification. From this definition it looks as though Kant has a psychological study in mind: an examination of the faculties whereby we are capable of a priori knowledge. Similarly in the Prolegomena he speaks of

> Das Wort transzendental ..., welches bei mir niemals eine Beziehung unserer Erkenntnis auf Dinge, sondern nur aufs Erkenntnisvermögen bedeutet .... (Ak. IV: 294.)

But further light is thrown by A 56/B 80:

> Und hier mache ich eine Anmerkung, die ihren Einfluss auf alle nachfolgenden Betrachtungen erstreckt, und die man wohl vor Augen haben muss, nämlich: dass nicht eine jede Erkenntnis a priori, sondern nur die, dadurch wir erkennen, dass und wie gewisse Vorstellungen (Anschauungen

1. "B" indicates the second edition of the Critique of Pure Reason, "A" the first. "Ak." refers to the collected edition of Kant's works put out by the German Academy.

> oder Begriffe) lediglich a priori angewandt werden, oder möglich sein, transzendental (d.i. die Möglichkeit der Erkenntnis oder der Gebrauch derselben a priori) heissen müsse.

This suggests that what is essential to transcendental knowledge is that it is concerned with the possibility of knowledge, and that it might be wrong to think of it as necessarily a sort of psychology. For it might turn out that not everything which was a condition of the possibility of knowledge, in the required sense, was a psychological phenomenon. Kant may not be clearly aware of this; he may want to make both its psychological character and its concern with the possibility of knowledge defining features of transcendental knowledge, without seeing that it is not obviously necessary that the two should go together. In common with the neo-Kantians and most present-day English users of the term, I shall use it, except when reporting Kant, in such a way as not to make it a matter of definition that transcendental knowledge is in any sense psychological in character.

Is it concerned with the possibility of any knowledge or only of a priori knowledge? A 56/B 80 seems rather unclear about this, and A 11 appears to support the latter alternative. But A 11 was replaced in the second edition by B 25, which supports the former; as do the last quoted clause of A 56/B 80

and other passages in the Prolegomena[1] and the Critique of Judgement.[2] One must admit that Kant does sometimes speak of his task as being to find the conditions for a priori knowledge.[3] In the Prolegomena he says repeatedly that his aim is to investigate the possibility of synthetic a priori knowledge. But there is no conflict here; he considers this more limited aim to be necessarily included in the wider one of finding conditions for knowledge in general:

> Weil nun alles theoretische Erkenntnis mit der Erfahrung zusammen stimmen muss: so wird dieses nur auf eine oder die andere Art möglich, nämlich dass entweder die Erfahrung der Grund unserer Erkenntnis, oder das Erkenntnis der Grund der Erfahrung ist. Gibt es also ein synthetisches Erkenntnis a priori, so ist kein andrer Ausweg, als es muss Bedingungen a priori der Möglichkeit der Erfahrung überhaupt enthalten....
>
> Die höchste Aufgabe der Transcendentalphilosophie ist also: Wie ist Erfahrung überhaupt möglich? (Ak. XX. 274f.)[4]

1. Sect. 21; and Anhang, first fn. (Ak. IV: 302, 373n.).
2. Introduction, sect. V. (Ak.V: 181). Cf. A 94/B 127; B 168; A 719/B 747; A 722/B 750; A 783/B 811; Ak. VIII: 194.
3. E.g. Reflexion 5667 (Ak. XVIII: 323f.); B 151; Ak. VIII: 214; Ak. XX: 272.
4. On the relation between Erfahrung and Erkenntnis see below.

It is clear that it is not just any sort of condition that is in point: human beings can only know things when they are at temperatures of (perhaps) between -50°C and 100°C, and this would hardly be relevant. The conditions must be a priori. Now he admits space and time as conditions of the right sort, and they are themselves a priori for Kant so far as human beings are concerned, since without them we could not have any knowledge. Yet he apparently derives the proposition that space and time are required for experience from the a posteriori premise that we cannot imagine a world without them. How then can transcendental knowledge of that proposition itself be purely a priori, as he says in the A 56/B 80 passage that it must?

Other questions raise themselves as well. What about knowledge of the a priori conditions, not ~~in~~ of knowledge in general, but of some specific sort of knowledge? He is prepared to call this transcendental in some cases; for to give the transcendental exposition of a concept is to proceed in the manner he calls "analytic" in the Prolegomena, and to argue that the possession of that concept is a necessary condition of having knowledge of a particular kind - it being assumed that knowledge of that particular kind is possible. He introduces the idea at B 40:

> Ich verstehe unter einer transzendentalen Erörterung die Erklärung eines Begriffes, als eines Prinzips, woraus die Möglichkeit anderer synthetischen Erkenntnisse

a priori eingesehen werden kann. Zu dieser Absicht wird erfordert, 1) dass wirklich dergleichen Erkenntnisse aus dem gegebenen Begriffe herfliessen, 2) dass diese Erkenntnisse nur unter der Voraussetzung einer gegebenen Erklärungsart dieses Begriffs möglich sind.

This raises the question of how closely a type of knowledge must be specified if transcendental arguments regarding its conditions are to remain interesting. An argument to something as a necessary a priori condition of religious knowledge lacke the force of an argument to something as a necessary a priori condition of there being any knowledge at all.

Again; it is conditions of the possibility of knowledge that Kant has in mind, i.e. necessary conditions. This is reasonable; it would be tedious to examine possible sufficient conditions for knowledge unless they or a disjunction of them were also necessary. But what status does "necessary" have here? Is it necessity in the sense in which analytical propositions are necessary (as I shall argue it is), or is some other sense involved (as has often been supposed)?

Towards the end of the Critique, in the chapter called "The Disipline of Pure Reason", Kant tells us a little more about transcendental knowledge. After distinguishing between philosophy and mathematics in that philosophical knowledge proceeds simply from concepts but mathematical from the construction of concepts in intuition, he points out that

if the philosopher is to have synthetic a priori knowledge about something he must go beyond the concept he starts from, "to properties which are not contained in this concept, but yet belong to it."[1] The mathematician does this by construction in pure intuition; the philosopher can only appeal to the general conditions under which the intuition of things is possible, not to any intuition.

> Synthetische Sätze, die auf Dinge überhaupt, deren Anschauung sich a priori nicht geben lässt, gehen, sind transzendental. (A 720/B 748)

The proof of such propositions starts from the possibility of experience: unless they were true there could be no experience.[2]

But in this chapter as elsewhere Kant sometimes uses the word "transcendental" rather more widely than his earlier definitions would suggest, in such a way that the alleged synthetic a priori claims of metaphysicians count as transcendental even though not susceptible of any such proof. Thus he speaks of the illegitimacy of "transcendental hypotheses" except for polemical purposes, meaning by this hypotheses which go beyond all possible experience and invoke ideas of reason.[3] At other times he would call these

1. A 718/B 746; A 721/B 749.
2. A 783/B 811; A 737/B 765.
3. A 772/B 800.

not transcendental but <u>transzendent</u>, <u>überschwenglich</u>. It is a commonplace that Kant frequently uses "transcendental" where one might expect him to use "transcendent".[1]

Only one proof can be found for each transcendental proposition, he goes on[2]: it consists in showing how the concept under examination affects the possibility of an object in general, i.e. exhibiting it as required for experience of objects.[3] And transcendental proofs must always be direct, not indirect ("apagogical").[4] The reason for this is not that an indirect proof would be formally invalid - which it could hardly be if direct proofs were valid - but that, as in the case of the antinomies, it is very easy in philosophy to confuse subjective features of our representations with objective features of things and so to think one has a genuine contradiction when one does not.[5]

From what has already been said one can see that Kant sometimes speaks of transcendental arguments as concerned with conditions for <u>knowledge</u>, and at other times as concerned

1. Thus A 614/B 642.
2. A 787f./B 815f.
3. Cf. <u>Prol</u>.sect. 17 (Ak. IV: 296).
4. A 789ff./B 817ff.
5. A 791/B 819.

with conditions for experience, or for the objects of experience. These come to much the same: he means by experience the sort of self-conscious experience which involves knowledge of objects. The more fashionable formulation of the same sort of argument as to do with conditions for language or communication comes to much the same as well; there could not be a language unless the speaker knew some things, and the sort of self-conscious experience Kant has in mind is the sort to which linguistic expression is appropriate.[1]

It is an interesting question why he chose the word "transcendental", and it deserves a longer discussion than I have space for.[2] The word was used by mediaeval philosophers

1. See further below, pp. 154f. Kant dismisses animals rather vaguely as having an analogon of reason (e.g. Ak. V: 464n).

2. It has come in for some discussion in Germany. Leisegang holds that Kant gives an "entirely new meaning" to the word. Knittermeyer suggests that one can see in German school-metaphysics a tendency to treat transcendentals as criteria of objectivity; his case I find unconvincing. Martin argues that Kant's use grows out of the traditional use, the central difference being simply that his transcendental philosophy is concerned not with the most general properties of things as things, but with those of things as they can be known. He makes the point that Kant uses the word "transcendent", previously used as synonymous with "transcendental", to apply to properties of things as things. This is no doubt true; but it must be Wolff's use of "veritas transcendentalis" which enables Kant to treat his term "transcendental" as relatively familiar, in something like his usage, to his readers, and which provides the immediate occasion for his assigning to it the sense he does. Cf. H. Leisegang, "Über die Behandlung des scholastischen Satzes...", Kant-Studien 1915, and his review of Martin's Wilhelm v. Ockham, Philosophische Studien II, 1950-1; H. Knittermeyer, "Von der klassischen zur kritischen Transzendentalphilosophie", Kant-Studien 1953-4; and G. Martin, Wilhelm von Ockham (Berlin, 1949), sect. 55.

to refer to those terms (notably "one", "true", and "good") which may be predicated of anything regardless of what kind of thing it is - regardless of what Aristotelian category it belongs to. This use seems entirely different from Kant's. The connection is to be found in Christian Wolff's remarks on veritas transcendentalis: these resemble some of the things Kant says about transcendental logic, which we have still to consider.

Leibniz took two primitive principles to be fundamental to rational knowledge, the principle of identity or non-contradiction and the principle of sufficient reason. Propositions which can be shown to be true by the principle of non-contradiction alone he calls truths of reason; all other truths are truths of fact, and the principle of sufficient reason states of them that there must be some reason why they are true. We might know them to be true without knowing why they were true, but such reasons there must always be; it is only because of our finitude that we have to rely on empirical sources of knowledge and cannot infer the nature of the world in every respect from the perfection of its creator.[1] Wolff's greatest innovation was

1. Nouveaux Essais book IV ch. 2 sect. 1; Théodicée, Discours Préliminaire. Both are in Die philosophischen Schriften von G.W. Leibniz, ed. by C.J. Gerhardt (7 vols., Berlin 1875-90).

to offer a supposed proof of the principle of sufficient reason from the principle of non-contradiction.[1] But he did not follow out the consequences of this very far; and he gave a different argument, which Kant would have called transcendental, for the principle of sufficient reason as well. For he says firstly that unless that principle were true we could not distinguish the world of experience from the world of dreams[2]; and later, that without it there could be no knowledge of anything not immediately present[3], and apparently no determinate propositions even about that.[4] And he presents this as an argument about the transcendental "true". Transcendental truth is different from logical truth: logical truth is a property of propositions, transcendental truth of things; and he says that if things were not transcendentally true propositions could not be logically true.[5] He does not mean what we mean by "logically true": here he is saying that no proposition, of reason or of fact, could be true at all (or indeed formulable).[6] Transcendental truth is defined as follows:

Veritas adeo, quae transcendentalis appellatur et rebus

1. Ontologia sect. 70.
2. Ibid. sect. 77, note.
3. Ibid. sect. 499.
4. Ibid. sect. 501 note.
5. Ibid. sect. 499.
6. Logica (3d ed., Frankfurt & Leipzig, 1740), sect. 505.

> ipsis inesse intelligitur, est ordo in varietate eorum, quae simul sunt ac se invicem consequuntur, aut, si mavis, ordo eorum, quae enti conveniunt. (Ontol. sect. 495.)

This order is given by the applicability to things of the principles of non-contradiction and sufficient reason.

> ... principium contradictionis et principium rationis sufficientis sunt fons omnis, quae datur in rebus, veritatis, hoc est, datur in ente ~~veritatis~~ veritas, quatenus ea, quae insunt, per ista principia determinantur. (Ontol. sect. 498.)

So Wolff is taking very seriously the idea that transcendentals can be predicated of things of any category, and saying that things in the world are true, i.e. real and not illusory, if they are ordered coherently. Coherence had of course been used by others to distinguish dreams from reality, but so far as I know Wolff was the first to connect this with transcendentals and with the idea that in an incoherent world no propositions could be true - no knowledge would be possible.

Kant's distinction between formal and transcendental logic recalls Wolff's distinction between these two kinds of truth. Formal or general logic abstracts from all content; it deals with the most general forms of thought.[1] It is, in

1. A 52/B 76; A 76/B 102; Ak. VIII: 214.

fact, what is ordinarily called logic; so far as it establishes propositions as true these propositions must be analytic, for the principle of non-contradiction is a principle of formal logic but the principle of sufficient reason is not. Transcendental logic takes care of the other sort of a priori truths, those which Leibniz and Wolff thought demonstrable by the principle of sufficient reason. So it is concerned not with relations between propositions but with relations between things, or at least between things and our knowledge of them. Kant even uses the expression "transcendental truth" in this context: it is equated with the objective reality of a concept, and consists in conforming to the conditions of possible experience.[1]

Kant says transcendental logic has to do with the possibility of things and not merely with the conditions for knowledge of things.[2] It is surprising that he should say this, after having spoken (in the very section on transcendental logic) of transcendental knowledge as concerning the a priori conditions for knowledge.[3] Because transcendental

1. A 146/B 185; A 222/B 269.
2. B 113ff.; Ak. VIII: 194. At Ak. IX: 15 he says that general logic has to do with objects in general, transcendental logic only with objects of the understanding.
3. A 56/B 80.

logic has to do with the possibility of objects, he apparently thinks no investigation of the relations between propositions, which are the business of formal logic, can help us with it.[1] The muddle he is in here may be a special case of his general confusion as to what is meant by "knowledge of objects". Sometimes he means by this objective knowledge, i.e. knowledge of propositions which have truth-values (so, any propositional knowledge at all); at other times, he means knowledge of items which are themselves objects, independent of the perceiving subject. If we take transcendental logic to be an enquiry about knowledge of objects in this latter sense, we might perhaps regard as the concern of formal logic conditions so general as to be conditions of any knowledge at all (i.e. conditions of objective knowledge in the former sense). Though even so it is not clear that there is much to recommend this; after all, any condition of knowledge as such will also be a condition of knowledge of objects in the restricted sense, and it may be an interesting condition despite its generality. And Kant does not always take his transcendental logic to have to do with conditions only for knowledge of objective items; sometimes, as we shall see, it is conditions for any

1. B 116.

knowledge at all that interest him.

Alternatively, we might keep the idea that general conditions for any knowledge at all are a matter for formal logic, but add that they are also transcendental. Indeed, the only anti- sceptical defence that can be offered for the laws of logic, e.g. the law of non-contradiction, is of the transcendental type: without them language/knowledge would be impossible. But this would be un-Kantian. Kant does not argue for the law of non-contradiction in this way (which is surprising, because Wolff does, though he does not give his own argument the importance it deserves[1]). And he considers the propositions of formal logic to be provable by appeal to that law: so they must be analytic; but he takes the conclusions of transcendental arguments, even the most general ones, to be synthetic a priori. And it is difficult to see how some could be analytic and others not.

But the distinction between formal and transcendental logic is not always drawn in a manner so difficult to make sense of. Sometimes he draws it by saying that formal logic considers only the form of judgements while transcendental logic is a degree less abstract, taking pure intuition into account.[2]

1. Ontol.sect. 55.
2. A 55/B 80; A 76f./B 102; cf. Ak. VIII: 245.

This fits in with his speaking of pure intuition as the "third thing" which can make the connection between subject and predicate in a synthetic a priori judgement. And it also fits in with what he said about transcendental knowledge: transcendental knowledge becomes knowledge of the propositions of transcendental logic. For to speak about intuition like this is really to say, in a slightly mysterious way, that transcendental logic has to do with the possibility of experience, or that the possibility of experience is the "third thing" - it is because the synthetic a priori judgement must be true if experience is to be possible that we can know it to be true a priori.[1] For it is because our experience involves intuition (awareness of particulars) that not all truths are analytic; all relations between concepts (general terms) can be expressed in analytic propositions. So the problem of the possibility of experience is a problem about the possibility of knowledge of what is given in intuition.[2]

Kant's greatest difference from Wolff in this area is that he does not regard the principle of sufficient reason as providing the touchstone for synthetic a priori truths.

1. A 55/B 80; A 155ff./B 194ff.; A 782ff./B 810ff.
2. Cf. B 136.

He replaces it by this appeal to the possibility of experience: the "highest principle of synthetic judgements" becomes "Every object stands under the necessary conditions of synthetic unity of the manifold of intuition in a possible experience".[1] But the alteration is partly superficial; for in view of Wolff's use of a transcendental argument to defend the principle of sufficient reason one may say that he regarded that principle, together with the principle of non-contradiction, as providing the necessary condition of the unity of the manifold. After all, his argument is just that unless these principles applied experience would be chaotic; he differs from Kant in not thinking of the unity of experience as being synthetic, due to the perceiving mind. And his principle of sufficient reason - Nihil est sine ratione sufficiente, cur potius sit, quam non sit[2] - is not so very different from the principle of causality, which could (if we put aside architectonic matters) be said to be Kant's main principle of the possibility of experience. Kant recognized that he stood in the Leibnizian tradition; in the polemic against Eberhard in 1790[3] he calls the Critique of Pure Reason the true apology for Leibniz. He suggests (how

1. A 158/B 197. Cf. Ak. VIII: 241.
2. Ontol. sect. 70.
3. Über eine Entdeckung ..... Ak. VIII: 250.

seriously it is hard to say) that Leibniz had in mind something like his own views but which still had to be worked out.[1]

Kant uses the term "transcendental" in plenty of other contexts than the ones I have mentioned. He speaks not only of transcendental knowledge, transcendental proofs, transcendental logic, and so on, but also, and in rather a bewildering way, of such things as transcendental synthesis, transcendental apperception, and transcendental imagination. These other expressions are however to be considered in the light of the previous ones: when Kant calls something transcendental, it is because he considers it a condition of the possibility of experience. But the role that all these terms play in his arguments can best be understood, and the questions raised earlier in this section can best be answered, by looking at the arguments themselves. The most important place to look is the transcendental deduction of the categories.

B. The Transcendental Deduction of the Categories.

Kant calls several of his arguments transcendental deductions. His definition of a transcendental deduction is not very helpful:

> Ich nenne .... die Erklärung der Art, wie sich Begriffe a priori auf Gegenstände beziehen können, die <u>transzendentale</u> <u>Deduktion</u> derselben, und unterscheide sie von

1. Ak. VIII: 248; XX: 363, 366, 376.

> der <u>empirischen</u> Deduktion, welche die Art anzeigt, wie ein Begriff durch Erfahrung und Reflexion über dieselbe erworben worden, und daher nicht die Rechtmässigkeit, sondern das Faktum betrifft, wodurch der Besitz entsprungen. (A 85/B 117.)*

It becomes clear however that it amounts to explaining that a concept is "justified", i.e. that it has to be used and can validly be used, because the possession of that concept is a necessary condition of having any knowledge. Transcendental deductions provide "justifications" in a way that transcendental expositions do not, because although both can be regarded as weapons against scepticism[1] the deductions are much the better weapons. For a transcendental exposition requires that one agree, to start with, on the possibility of some specific sort of synthetic <u>a priori</u> knowledge, like geometry or arithmetic, whereas a transcendental deduction requires no such concession. In showing something to be a condition of <u>any</u> knowledge one shows it to be beyond reasonable doubt: one shows it to be a condition of the intelligibility of any such doubt. In fact Kant's transcendental deductions do not try to give the conditions for absolutely any sort of knowledge; but the limitations he imposes are very general, not the sort of thing the sceptic can call in question if he wants his doubt to be of interest

1. See below, pp. 180ff.

* Cf. also Ak. XI: 79 (To Kosmann, Sept. 1789) and B 159.

to us.[1] The deduction of the categories attempts to set out conditions of knowledge for any being without intellectual intuition. Kant is not as explicit as one would like, but I think he would agree that another sort of transcendental deduction can in principle be given as well, one which really is concerned with the conditions for absolutely any sort of knowledge; for, despite the passages where he says the deduction of the categories is valid only for beings with some sort of sensible intuition[2], he seems to think that the pure categories - as forms of thought, simply - govern the thinking even of a mind with intellectual intuition.[3]

To give a transcendental deduction of a concept is to give a transcendental argument - an argument about the necessary conditions for knowledge ~~of~~ experience - to justify the claim that objectively true or false judgements can be made using that concept. In the Critique of Pure Reason[4] he speaks of a transcendental deduction of the concepts of space and time, as well as of the deduction of the categories. In the Critique of Judgement[5] we have a transcendental deduction of the concept of finality in nature. In the

1. See below pp. 184 ff.
2. B 145; A 286/B 342.
3. Ak. XX: 272; cf. Adickes, Kant und das Ding an Sich (Berlin, 1924), ch.4. This is not to say that at B 145 he is speaking of schematized categories: rather he is thinking at an intermediate level of abstraction, at which categories are not pure forms of thought but forms of synthesis of what is given in intuition.
4. Abbreviated "KdrV".
5. Abbreviated "KdU".

Critique of Practical Reason[1] we are told that no transcendental deduction of the moral law can be given[2], but a deduction of the idea of freedom is given from the fact that rational beings are conscious of the moral law as binding upon them.[3] On this basis also justifications are given for the belief in the summum bonum, the existence of God, and the immortality of the soul. The deduction of the idea of the summum bonum is explicitly called transcendental.[4]

I shall not consider in detail the deductions in the second and third Critiques. They raise special problems, and I am interested here in Kant's general idea of a transcendental deduction. The arguments in KdpV differ from those of KdrV in that they do not take as premise that there is knowledge or experience, and show that in consequence it must be possible to use the concept in question in an "objectively valid" way. They start rather from the "moral fact" that we are conscious of obligations. It is never claimed that a being not conscious of obligations could not have knowledge or experience; so the

1. Abbreviated "KdpV".
2. Ak. V: 46f. In fact "transcendental" and "moral" are contrasted in KdrV at A 801 &n./B 829 &n. and A 805/B 833; though cf. also A 816/B 844.
3. Ak. V: 47. Cf. A 802/B 830. That when he calls this a fact he really means it is made clear by A 829f./B 858f, where he recognizes there may be people indifferent to the moral law.
4. Ak. V: 113.

deductions of the second Critique do not have the same generality as that of the categories. They are rather to be compared with the arguments of the Analytic of Principles, in which it is conditions for experience of a (supposedly) more specific sort, namely temporal experience, that are considered. But they differ also from these in that they justify the use of what Kant calls an idea rather than a concept: we can have no theoretical knowledge about freedom or the summum bonum, God or immortality, but only practical knowledge - we would not be justified in applying these notions to the world we experience, but we are justified nevertheless in acting in certain ways and having certain attitudes. It is noteworthy that though Kant calls the arguments of KdpV deductions, he hardly ever calls them transcendental deductions. They are certainly not supposed to be transcendental expositions: those start from the premise that there is synthetic a priori knowledge of a particular type and draw consequences from that, but KdpV assumes only that our experience includes the "a priori consciousness" of the moral law, and this is presumably not supposed to amount of itself to a knowledge of synthetic practical propositions. KdpV does contain also an "exposition of the highest principle of practical reason"[1], and the distinction is drawn very

1. Ak. V: 46.

clearly in the Grundlegung[1], where the method of the first two chapters is "analytic"[2], like that of the Prolegomena, and that of the third "synthetic": the first two chapters assume that we do possess synthetic a priori moral knowledge and derive the principle of autonomy from that, thus giving it a transcendental exposition, whereas the third gives it a deduction (like that of KdpV except that it starts from our awareness of freedom).

In KdU the argument does start from the general premise that there is knowledge or experience. It adds to the claim, made in the deduction of the categories, that experience must be unified, the further claim that we must approach nature with the assumption (which may not be correct) that the laws which govern nature are readily comprehensible to us: i.e., that nature is as it were designed to fit human cognitive faculties.[3] Since we must all feel pleasure when we come across phenomena which seem neatly to fit our cognitive faculties, and since judgements about the pleasure attaching to such phenomena are what we call aesthetic judgements, a transcendental argument can be given for the validity of such judgements.[4] The details of the argument are obscure; it would not be profitable to consider them here. Here again,

1. Ak. IV: 385-463.
2. Ak. IV: 274f., 276n. (Prol. §§ 4,5).
3. Einleitung sect. V.
4. Einleitung sects. VI-VII; sects. 12, 37-40.

however, Kant is not saying as in KdrV that we are justified in applying certain predicates ("beautiful" etc.) to the objects of experience, but only that we are justified in having certain attitudes towards them.[1]

Let us turn to the Critique of Pure Reason. Here two transcendental deductions are supposed to be given: one of the categories, and one of the concepts of space and time. The latter shows, according to Kant, that "the synthesis in them (space and time) has objective validity" because

> nur vermittelst solcher reinen Formen der Sinnlichkeit uns ein Gegenstand erscheinen, d.i. ein Objekt der empirischen Anschauung sein kann. (A 89/B 121.)

By saying "the synthesis in them has objective validity" he must mean that judgements involving these concepts can be objectively true or false. In a similar way the transcendental deduction of the categories is supposed to establish the categories' "objective validity" by showing

> dass sie als Bedingungen a priori der Möglichkeit der Erfahrungen erkannt werden müssen. (A 94/B 126.)

Strictly it is only the deduction of the categories which can be supposed to show the objective validity of certain concepts. For strictly space and time as forms of intuition are not concepts, though Kant often calls them "Begriffe"

1. Sects. 1, 34.

when he is using the word non-technically. But I shall postpone discussion of space and time until later.[1]

It is customary to distinguish, as Kant himself did, two elements in the transcendental deduction of the categories: the subjective and objective deductions. But there is little agreement as to what these are. I shall distinguish three arguments, without claiming that they are the only ones to be found; each is supposed to reach the conclusion that the categories must be applied throughout our experience, but they have different starting-points and take different routes. One could be called a subjective deduction, for it is concerned with "the subjective sources which form the a priori foundations of the possibility of experience"[2]; it starts from the premise that experience must not be "haphazard and arbitrary"[3], and has to do with what Strawson calls "the imaginary subject of transcendental psychology".[4] The other two may both be called objective deductions, being concerned to "present and make conceivable the objective validity of the a priori concepts of [pure Understanding]", and do this by showing "that by means of them alone an object can be thought".[5] One of the arguments, however, has to do with

1. Below, pp. 167 ff.
2. A 97; cf. A xvii.
3. Kemp Smith's rendering of "aufs Geratewohl, oder beliebig" (A 104).
4. Bounds of Sense (London, 1966), p. 97.
5. A xvi, A 97.

the question of how we come by the concept of an object, i.e. of a particular item capable of existing independently of the perceiver, and it does this by a discussion of synthesis; this argument comes in a number of different forms. The other, which is simpler and later, is concerned with knowledge of propositions with an objective truth-value - so, with all knowledge - and makes the simple point that it involves judgement and hence must proceed according to the twelve forms of judgement. The last argument is valid but rather trivial, or at least not very exciting nowadays; the other objective deduction, in its various forms, may be read as trivial or as claiming a rather strong sort of involvement of the categories in constituting objects for knowledge, in which case it is invalid. The subjective deduction may also be read either as trivial or as invalid, depending on how one thinks of the faculties of the mind whose interaction it maps out.

These arguments do not, of course, appear in neatly separated parts of the deduction. Where Kant himself distinguishes between the objective and subjective arguments, in the preface to A, he speaks of them as two <u>sides</u> of the argumen t; and certainly they get mixed together in his presentation. And Kant adds to them, especially to the parts properly belonging to the subjective deduction, a good deal of material which is strictly irrelevant to the argument, just because when he finds himself in the area he is tempted

to do a bit of conceptual mapping of nearby regions. For example, the idea of pure synthesis - the synthesis of a pure manifold of space or time - is not very necessary to the argument. This is not to say that his remarks on such matters are not interesting in themselves, or that the arguments involved are not transcendental in character.

Before proceeding further, I shall consider two recent interpretations of the transcendental deduction, put forward by Strawson and Bennett.[1] They succeed in getting philosophically interesting material from Kant on the readings they propose (and some of it will be discussed below), but the text of the deduction does not fully support their interpretations of Kant. For Strawson, the conclusion of the argument is supposed to be that if we are to have experience at all we must have experience of objective items, items whose _esse_ is distinct from their _percipi_. For Bennett, it is supposed to be that any experience I can conceive as mine must be such that I can apply concepts, and in particular the concept of my own identity, ~~to~~ in it; he agrees with Strawson that the argument _does_ contain the claim that we must have experience of objective items, but he regards this as less important and as an illegitimate attempt to anticipate the argument of the Refutation of Idealism.[2]

1. Strawson, _op.cit._, pp.85-117; Bennett, _Kant's Analytic_ (Cambridge, 1966), chs. 8 & 9.
2. At any rate in _Kant's Analytic_. In his review of Strawson in _Philosophical Review_ 1968 he says, "Strawson has satisfied me that I have erred in seeing the T[ranscendental] D[eduction]'s value as wholly residing in Kant's arguments for the claim "A self-conscious being must apply concepts"]." (p.342).

It seems to me that the main line of argument in the deduction has nothing to do with criteria for personal identity, and that the need for objective items is generally assumed, not proved. It is in fact the starting-point of the objective deduction in at least one of its forms, as I shall try to show. Strawson accepts that Kant put it forward as a premise, but claims that it is only a "provisional premise"; but Kant does not say this, nor is it clear why he should have thought it desirable to put forward his conclusion as a provisional premise. Kant does say that the first part of his argument in A is to be regarded as provisional, but it is not only there that experience of objects or knowledge of objects is taken as the starting-point: it appears in most accounts of what the transcendental deduction is doing[1], and as we have seen it even comes into his first account of the word "transcendental" at A11/B 25.

Strawson holds, like Ewing[2], that the initial premise of the deduction is that we have self-conscious experience. Certainly Kant does sometimes start from this, but he does not always. He has an argument to show that self-conscious experience must involve knowledge of objects; at least, he has an argument that self-conscious experience must be experience

1. E.g. A92 ff./B124 ff., A155 ff./B194 ff., Prol. §§ 18-22.
2. A. C. Ewing, Short Commentary on Kant's Critique of Pure Reason (London, 2nd ed., 1950), p.72.

of an objective world, which he takes to come to the same thing. But he also has another argument to show that knowledge of objects must involve self-conscious experience, on the grounds that it requires a necessary connection of past and present representations.[1] And the point of the deduction is not to prove either of these things: they are steps in the proof that the categories must apply throughout all experience. The need to argue that self-conscious experience must be experience of an objective world arises only because Kant seems to think he has not shown enough when he has shown that my experiences must be capable of being synthetically united in one consciousness: he wants there to be a thoroughgoing synthetic connection between appearances themselves.

Kant often says he is investigating the possibility of experience, simply. This really brings both the starting-points together: for it appears that "experience" ("Erfahrung") means "self-conscious knowledge of the objects of the senses". It is so defined in the Fortschritte[2]; and it is often explained in terms of perception (Wahrnehmung)-- as a synthesis of perceptions[3] or as knowledge of objects through

1. A 107, A 121.
2. Ak. XX: 274.
3. B 218; Prol. sect. 5; Reflexion 5928.

perceptions[1]; -- he takes perception to be the self-conscious awareness of something.[2] He does not clearly discuss whether I can still be said to be having experience when I bring to consciousness a manifold which does not provide knowledge of objects outside me -- e.g. when I am having an hallucination. This connects with his general unclarity about the relation between knowledge of objects, i.e. of things which exist independently of being perceived, and knowledge which has objective validity in the sense that it is expressible in propositions which are objectively corrigible and which can therefore have truth-values. (All knowledge is "objective" in this latter sense; in the former, knowledge about one's own, or other people's, inner states is not). Sometimes he uses the word "Erfahrung" as we would use "experience", and here it can only be a matter of "knowledge of objects" insofar as we treat the "objects" as being the intentional objects of the experiences; judgements about these will have objective validity, for it is objectively true or false that I had an hallucination at t, but they will not necessarily be about objects which can exist independently of being perceived. At other times he uses it in such a way that we

1. B 161; B 218f.; Ak. XX: 274.
2. A 120; B 160; B 207.

only have <u>experience</u> insofar as we have knowledge about things which exist independently of us -- cf. his distinction between <u>Wahrnehmungsurteile</u> and <u>Erfahrungsurteile</u> in the <u>Prolegomena</u>.[1]

In any case it cannot be true that Kant's purpose in the deduction is to show that there must be experience of objective items (in either sense), for Kant leaves us in no doubt that the aim is to establish that experience must be subject to the categories. That is done by showing that without the categories experience would be impossible; and an alternative way of putting this is to say that the argument goes some way towards showing how experience, i.e. knowledge of objects, is <u>possible</u>. But that is quite different from constructing a transcendental argument to show that there must be knowledge of objects. And it is difficult to find in Kant anything analogous to Strawson's argument about the recognitional component in knowledge and the is/seems distinction.

The idea that the Deduction is concerned with personal identity derives from connecting the "unity of apperception" with the requirement that there should be criteria whereby I recognise all my representations as mine. This is inconsistent with what Kant says in the <u>Paralogisms</u>, besides being of slight

1. Sect. 18. See also below, pp. 132 ff.

relevance to the purpose of the Deduction. In the Paralogisms, as indeed also in the Deduction, pure or transcendental apperception is carefully separated from empirical self-consciousness, and while it is recognised that criteria are required for the unity of empirical self-consciousness -- i.e. for the notion of myself as a persisting entity which has experiences $\underline{e}_1$ and $\underline{e}_2$ at $\underline{t}_x$, $\underline{e}_3$ and $\underline{e}_4$ at $\underline{t}_y$, etc., and now $\underline{e}_n$ and $\underline{e}_m$ -- it is held that it is mistaken to look for such criteria for the unity of transcendental apperception.[1] He is not saying - or need not say - that there is something (noumenal?) which persists and is such that I can know that certain former states are states of it and that certain present states are states of it, but without any criteria for personal identity being relevant; this is the view he ascribes to the transcendental psychologist.[2] I take his point to be rather that I must think of myself as one person persisting through time without invoking criteria of personal identity: I just must take certain of the claims I am disposed to make about my present and past states as true without further investigation. (It

1. A 362ff.
2. This is an instance of his using "transcendental" where we might expect "transcendent": for Kant, a transcendental psychologist is a metaphysician who makes the mistakes he points out in the Paralogisms. Thus there is a prima facie difference between his usage and Strawson's here -- though no doubt Strawson wishes to suggest Kant himself makes similar mistakes. See below, pp.41-50.

would not make sense to me to suppose that there is one person here who is writing this but does not have this feeling of warmth now and another who has the feeling of warmth but is not writing the paper; for the rather dull reason that I do not have to find out, at least in typical cases, which mental states belong to me. So I must think of myself as a unit, not just a disconnected set of Humean impressions and ideas, without having to apply criteria of personal identity). No doubt it is true that non-criterial first-person judgements can only have the meaning they do because there are criteria which could be applied, by others or even by oneself, and the result of the application of which could be, occasionally, to overset the non-criterial self-ascription. But the consideration that non-criterial self-ascriptions are in some way parasitic on ascriptions based on criteria is one of which I find no trace in Kant[1]; it cannot be presumed he had in mind such a move in the Deduction.

1. Bennett (Kant's Analytic pp.114-5 and notes) cites B 129-30, B 134, A 116, A 121-2. The first of these he interprets as involving the assertion "Existentially quantified conjunctions can be true only by virtue of the satisfaction of unity-criteria" (p.114), but Kant says nothing about criteria, only that combination is an act of the understanding. The understanding follows rules, but Kant never says these take the form of identity-criteria. Again B 134 only says that to think of my representations as mine I must be able to synthesise them; and the same goes for the passages in A. Synthesis, or combination, or as he later calls it Zusammensetzung, is just the putting-together of representations. It cannot be supposed without further argument that Kant took this necessarily to involve the application of criteria for personal identity.

It is an argument of Kant's that to have experience I must think of myself as a unity. But how would it help him towards the conclusion of the Deduction, that all experience must be ordered by the categories, to argue further that we require criteria for personal identity? Of course if we have to have criteria for personal identity it will be necessary for us to give work to certain concepts, namely those involved in expressing our criteria. But these concepts are not the categories; no reasonably plausible set of criteria for personal identity could have any particularly close connection with the Kantian table of categories. And they could in no sense be said to order our experience in the way Kant thinks the categories do. At most such a requirement would make it necessary that there be certain similarities between my experiences at one moment and those at the next; and if the criteria are something like the spatio-temporal continuity of human bodies, a view Kant seems to favour at A 363, it imposes no restriction at all on the content of my mental life- not even that I cannot have experiences of a non-spatio-temporal character.

The clearest and I think latest[1] of the three arguments

1. We may regard it as the final form of the transcendental deduction. Adickes claims there is an entirely new transcendental deduction in the Opus Postumum (see his Kants Opus Postumum, Berlin 1920), but this argument shows the categories to apply to our experience in virtue of considerations about the ways in which we can perceive things; which is not a purely a priori matter. The argument is limited in scope and would not apply to beings with sense-organs quite different from ours; it is what Kant calls a psychological, not a transcendental, deduction in his letter to Kosmann of September 1789 (Ak.XI:79). Cf. also A 85/B 117.

I am going to discuss is one of the "objective deductions". It is simple and straightforward, which is one reason for thinking it late in date - Kant would not have bothered with the confusions and obscurities of the first edition deduction if he had been clear about it in 1780. The argument is simply that all objective knowledge - all knowledge that has a claim to be considered valid or invalid - involves judgements; but it has been supposedly shown in the Metaphysical Deduction that there are twelve, and only twelve, sorts of judgement, and when we think of these twelve sorts of judgements as sorts of mental activity involved in knowing things we call them categories. So all knowledge involves the categories. This argument appears in sect. 19 in B, in a footnote to the preface of the Metaphysische Anfangsgründe der Naturwissenschaft[1] (1786), and in the Fortschritte (1793-4); something very like it appears in Reflexionen 5923 and 5934 in the Nachlass.[2] In the MAdN passage he does not so much give the argument as indicate it: he says he now sees that the deduction is very easy,

1. Abbreviated "MAdN".
2. Dated by both Erdmann and Adickes to the period between the Prolegomena (1783) and B (1786). (See notes in Ak.XVIII: 385f.) The Prolegomena also contains suggestions of this argument:

> [Ich sah] mich nach einer Verstandeshandlung um, die alle übrige enthält und sich nur durch verschiedene Modificationen oder Momente unterscheidet, das Mannigfaltige der Vorstellung unter die Einheit des Denkens überhaupt zu bringen, und da fand ich, diese Verstandeshandlung bestehe im Urteilen. (Ak. IV:323.)

> da sie beinahe durch einen einzigen Schluss aus der genau bestimmten Definition eines Urteils überhaupt (einer Handlung, durch die gegebene Vorstellungen zuerst Erkenntnisse eines Objekts werden) verrichtet werden kann. (Ak. IV: 475n.)

In Reflexion 5923 he fills this out, starting from the claim that all knowledge involves judgements and then saying:

> Die Form aber eines jeden Urteils besteht in der objektiven Einheit des Bewusstseins der gegebenen Begriffe, d.i. in dem Bewusstsein, dass diese zu einander gehören müssen, und dadurch ein Objekt bezeichnen, in dessen (vollständiger) Vorstellung sie jederzeit zusammen angetroffen werden.
>
> Es ist aber diese Notwendigkeit der Verknüpfung nicht eine Vorstellung empirischen Ursprungs, sondern setzt eine Regel voraus, die a priori gegeben sein muss, d.i. die Einheit des Bewusstseins, die a priori statt findet. Diese Einheit des Bewusstseins ist in den Momenten des Verstandes beim Urteilen enthalten, und nur das ist Objekt, worauf in Beziehung Einheit des Bewusstseins der mannigfaltigen Vorstellungen a priori gedacht wird. (Ak. XVIII: 386f.)

And this, together with the claim that

> Alles Erkenntnis, mithin auch das der Erfahrung, besteht... aus Urteilen (Ibid.),

is supposed to constitute a "Deduction der reinen Erkenntnisse

a priori".[1] The last sentence of the long quoted passage contains something of another argument as well; what is important for the present argument is that all experience involves judgements and these judgements have certain a priori forms. Though it is not a happy way of putting this to say that making a judgement involves making some sort of necessary connection between the subject- and predicate- terms. Judgements can be objective (in either sense of that word) without being necessary.

The argument here, and that in Reflexion 5934, starts from experience as knowledge of objective items - experience in the sense in which judgements of experience are contrasted with judgements of perception in the Prolegomena. In B, Kant regards the argument of sect.19, which also proceeds simply by considering the notion of judgement, as incomplete, apparently because he thinks it only shows that experience in this limited sense involves the categories[2]; it remains to be shown that we need the categories for all experience even of the most rudimentary sort - even the sort expressed in the Prolegomena's judgements of perception. This he shows by arguing that all synthesis is governed by the categories.

1. Ak. XVIII: 385.
2. Though perhaps this is not the reason. Paton (Kant's Metaphysic of Experience I: 526ff.) thinks what remains to be done is to show how the categories apply to specifically human experience.

> Folglich steht alle Synthesis, wordurch selbst Wahrnehmung möglich wird, unter den Kategorien, und, da Erfahrung Erkenntnis durch verknüpfte Wahrnehmungen ist, so sind die Kategorien Bedirgungen der Möglichkeit der Erfahrung, und gelten also a priori von allen Gegenständen der Erfahrung. (B 161).

This position is an advance on the Prolegomena, where judgements of perception were held not to involve the categories.[1] That work did however contain the insight that the synthesis involved in perception (Wahrnehmung) could be expressed in judgements, despite the fact that these judgements would be subjective in the sense, not of lacking truth-value, but of not being about things independent of the observer. So it is not surprising to find Kant in the Fortschritte simplifying his argument by claiming that all synthesis is a matter of judging and consequently must proceed according to the forms of judgement. "Der Verstand zeigt sein Vermögen lediglich in Urteilen."[2]

The idea that this argument is or should be the core of the transcendental deduction has certain advantages for the attempt to understand Kant; and certain disadvantages. A minimum of what Strawson calls "transcendental psychology" is involved, and the place of the metaphysical deduction is

1. Prol. sects. 18, 20.
2. Ak. XX: 271.

made clear, for its task is to establish what the a priori forms of judgement are; but the place of the Schematism and Kant's general account of dialectical error becomes obscure, since the pure categories are on this reading the forms of judgement under a pseudonym, or else simply the concepts of the different forms of judgement.[1] Or at least, the forms considered as applying to synthetic judgements. In the Fortschritte passage he goes on to say that the categories are the logical functions of judgement, so far as these represent the synthetic unity of apperception of a manifold in general[2]: so the categories, as forms of synthesis, must be forms of synthetic judgement. This of course makes only a trivial distinction between categories and forms of judgement: we are just prevented from calling the forms of judgement categories when it is analytic judgements we have to deal with.

More importantly, the conclusion of this argument is not nearly as exciting as Kant's remarks about the transcendental deduction give us a right to expect. Though it is not entirely trivial. It did need to be pointed out that since (or so far as) experience must be describable we cannot

1. Cf. also A 68f./B 93f; Ak. XI: 333; Refl. 5933 (Ak. XVIII: 392). Kant's description of the categories as "concepts of an object in general" does not fit this account very well (e.g. B 128).
2. Ak. XX: 272.

experience things which cannot be described using the sorts of judgement that logic allows us. And this classification can be said to be imposed by the mind on experience, because the different categories are just different ways in which the understanding can grasp things.

He is concerned with finding necessary conditions for knowledge or experience, for beings which do not have intellectual intuition.[1] Now his distinction between intuition and concept seems to be a distinction between particular and general terms: for his method of showing space and time to be intuitions and not concepts is to show that they have parts not instances.[2] For a being with intellectual intuition, as in Leibniz' system, there would be no ultimate distinction between particular and general terms; but for any other sort of being there is such a distinction, and synthetic knowledge must consist in the subsumption of particulars under general concepts. The faculty which performs this operation is the judgement, and it operates by forming judgements.

> Urteilskraft überhaupt ist das Vermögen, das Besondere als enthalten unter dem Allgemeinen zu denken. (KdU Einl. IV -- Ak. V: 179).

1. B 145.
2. A 25/B 39f.; A 31f./B 47f.

But it is quite obvious that any judgement must be either universal or particular or singular, either affirmative or negative or "infinite" (of the form "A is not-B"), etc.; though the alternatives within these triads are only exclusive if one allows him to make them so by his definitions of "particular" and "negative". He has not shown what would have been really interesting, namely that one could not have a language in which only some of his twelve sorts of judgement could be made[1]; for one might, I should think, have a language without disjunctive or infinite judgements, and it seems an interesting question -- which we shall have occasion to consider in chapter III -- whether or not one could do without modalities and operate with a purely extensional language. And what he has shown does not seem to have much salvageable bearing on the arguments of the Analytic of Principles, where the categories are supposed to be applied to temporal or spatio-temporal intuition. The principle of causality is supposed to arise out of the hypothetical form of judgement applied to temporal phenomena -- or as Kant would put it the corresponding pure category "schematized" in the pure intuition of time.[2] But one can make hypothetical

1. Though no doubt he thinks he has. Cf. Refl. 5932 (Ak. XVIII: 391f.). On this point see Strawson, Bounds of Sense, pp.79-82; Bennett, Kant's Analytic, pp.79-83.
2. See below, sect. D of this chapter.

judgements about events which have nothing to do with causality, or even with the objective sequence of events one upon another: for example, "If A will be happening at t then B will be happening at t". It is true, of course, that causal judgements can be expressed as hypotheticals about events, but a transcendental argument for having hypotheticals about events does not establish a need for causal judgements. Much less does an argument that there must be general, particular and singular judgements about things in space and time show that all intuitions are extensive magnitudes.

Kant does, of course, try to show for each of the Principles in turn that it is a necessary condition of (spatio-) temporal experience. Strictly this should not be necessary -- the Principles should be established by the arguments of the Deduction and Schematism, and further arguments can only serve to confirm these. Conceivably he recognized dimly that the Deduction just does not show that all the categories are necessary but only that some are, so that the indispensability of each has at this stage still to be argued for.

But the other arguments which can be found in the Deduction do not identify the categories so closely with the forms of judgement. If we follow them we can make the connection between the Principles and the judgement-forms less immediate, and so avoid running immediately into these difficulties; but on the other hand we are faced with a

problem as to what the pure categories are.

Kant never, or at least not until very late on, succeeded in separating completely the argument from the notion of judgement, which we have just been considering, from the other argument of the objective deduction; as for example the last sentence of Reflexion 5923, quoted above, suggests. This other argument is not concerned with what is involved in judging but (in one version) with the analysis of the concept of an object; and the point it makes is that objects cannot be regarded simply as representations: the concept of an object involves a synthesis or combination of representations. Nor are objects simply classes of representations, logical constructions out of nothing but sense-data: we have to go beyond what is or could be given in that we have to make use of the notion of a "transcendental object", something not itself a representation but a cause of ordered sets of representations, as well. But all that is actually given to the mind is representations; so the rest of the notion must be produced by the mind within itself through its own activity, which Kant in A calls synthesis and ascribes to the imagination and (usually) in B calls combination and ascribes to the understanding.[1] But

1. A 100ff.; A 104ff.; A 122ff.; A 250; B 306; A 93f./B 126. This last has particular importance in the light of Kant's remark at A xvii that A 92-3 contains all that is needed for the Deduction.

the mind can only act in an extremely limited number of ways; and what these are is indicated by the table of judgements. In judgement the mind links particulars and general concepts; in synthesis it links together particulars to form general concepts; but the sort of spontaneous mental activity will be the same in each case -- all that differs is what is acted upon.

> Dieselbe Funktion, welche den verschiedenen Vorstellungen in einem Urteile Einheit gibt, die gibt auch der blossen Synthesis verschiedene Vorstellungen in einer Anschauung Einheit, welche, allgemein ausgedrückt, der reine Verstandesbegriff heisst. (A79/B 104f.)

So here the categories are not just equated with the forms of (synthetic) judgement. But it is now difficult to see what they are -- what a pure, "unschematized" use of them could be. They are still forms of mental activity, nothing so specific as the principles of causality or real-possibility (the "possibility" of the Postulates); so a difficulty remains as to what the relation is between the conclusions of the Deduction and the findings of the Analytic of Principles.

This argument appears to improve on the last by reaching a less trivial conclusion than that all our experience must be expressible in judgements of a restricted number of logical forms. But it is not clear that it really

ought to reach a less trivial conclusion. The concept of an object, as something which can exist independently of us, is not the sort of concept for which one could give even a _prima facie_ plausible account in terms of what is _given_ alone, without bringing in some sort of mental activity which synthesises together different representations of the thing and adds in at least the idea of possible but non-actual representations of it. Nevertheless, why should the mind have to act here in those ways and only those ways in which it can act in judgement? Kant might reply that the synthetic activity of the mind just does consist in making judgements (though not necessarily conscious ones)[1]; but then the conclusion of this argument collapses into that of the argument from the notion of judgement. Alternatively he might deny that the synthetic activity consists in judgements; but now the nature of the parallel between the categories, as forms of synthesis, and the forms of judgement remains unclear. No reason is offered _why_ the sorts of spontaneous activity of which the mind is capable should be the same in synthesising as in judging. And it is difficult to see what a form of synthesis corresponding to the affirmative or to the hypothetical judgement could be like.

1. This seems a sensible thing to say. If I use a concept I must see certain items (to which it applies) as similar, and one might therefore say that in a sense I must _judge_ them similar.

The argument just considered does not appear on its own, except perhaps at A 104-6. Even there it is immediately followed, at A 107, by an argument that knowledge of objects also requires unity of consciousness, since it requires a synthesis of different representations at different times. This unity of consciousness is taken to be a unity of self-consciousness, perhaps on the grounds that if I am to think of objects I must be able to distinguish them from myself. It is this condition of unity of self-consciousness or apperception on which another form of the objective deduction, to be found in both editions, turns. This other version is by far the most prominent argument in the Deduction.

It starts by pointing out that if an experience is to be anything for me I must be able to ascribe it to myself, as my experience.[1] But I must also think of myself as a unity, not merely as the sum of a collection of diverse experiences, for otherwise nothing would be gained by ascribing an experience to me.[2] This notion of myself as a unity can only be produced through synthesis. The parallel with the concept of an object, in the other form of the argument, is exact: just as for the concept of an object we had to synthesise given representations in the light of

1. A 113; A 116; B 131f.
2. A 116ff.; B 134.

the idea, not itself given, of the transcendental object=x, so to get the concept of the subject we synthesise all our experiences together in the light of the idea of the transcendental subject=x.[1] So the unity of apperception involves synthesis carried out according to rules provided by the understanding. Now, all synthesis presupposes the synthesis of the pure manifolds of space and time which are the bases of any empirical manifold. This synthesis of pure manifolds involves no a posteriori element, so it cannot have any empirical concept as the rule governing it; therefore it must have a pure concept, i.e. a category.

This last bit of the argument seems to be stated only at B 160 (though perhaps it is suggested at B 150), but presumably Kant has it in mind when he bothers to show that we must synthesise the pure manifolds, at A 99. It is strictly unnecessary, for two reasons:

(1) Kant often (e.g. at A 111) replaces it by the consideration that the synthesis involved in transcendental apperception must be a priori, and so governed by the categories and not by empirical concepts. It must be

1. A 346/B 404; A 355. The other version of the argument -- the one from the concept of an object -- only showed that knowledge of objects in the narrow sense, as items independent of us, involved the categories. The present version -- from the concept of the subject -- reaches the more general conclusion that all knowledge or experience, whether of things independent of us or not, involves the categories. For all experiences must be such that we can ascribe them to ourselves if we are conscious of them at all.

a priori, because since all experience depends on one's being able to think of oneself as a unity this cannot itself be something one learns from experience. But

(2) not even (1) is necessary because after all any synthesis is a spontaneous mental act, and the table of judgements is supposed to show us what kinds of mental act there are. So that even if the synthesis is that involved in applying an empirical predicate like "red" -- and not just applying the pure categories to space and time, or entertaining the idea of the self -- it will have to proceed in one of these ways and so involve the categories.

In a rather fuller version of the argument, when he has shown that self-consciousness requires me to think of myself as a unity, Kant adds that if this is to be possible one must also be able to think of experience as a unity as well -- as not simply an aggregate of items but as necessarily connected. And this necessary connection can only be introduced by the mind's synthesis of the experiential data. But, as Strawson says, Kant leaves it obscure why the move must be made from thinking of oneself as a unity to thinking of one's experience as a unity.[1] Kant does not seem just to be saying that since I can ascribe

1. Bounds of Sense, p.96. Strawson attempts to fill this gap, and I shall consider his argument below (~~pp.~~ sect.D.).

all my experiences to myself they must all be synthesisable together as my experiences[1]; but it is perhaps because he confuses his more substantive point with this trivial claim that he thinks the matter does not require argument.

Since the argument about the concept of a subject has the same form as the argument about the concept of an object, it suffers from the same drawbacks; though its generality is greater, since it argues that the categories must be applied throughout self-conscious experience, and not only when dealing with objects independent of the perceiver. It is as implausible to think of the self as a series of actual perceptions as it is to think of objects in that way; at least some sort of synthesis of the different actual perceptions and other possible ones would be required. At any rate, it seems reasonable enough to say that, if I am to be capable of thinking of an experience as mine (and only so can I be said to be self-conscious), then I must be able to synthesis different representations. But once again what is the relation between the a priori forms of synthesis (the categories) and the forms of judgement? If the activity of synthesis is no more than that of making judgements (of

1. Though he appears to at A 108, A 113, A 117 and A 117n.

similarity etc.) then we know what the relation is -- the categories <u>are</u> the forms of synthetic judgement -- but then we are back to the conclusion of the argument from the notion of judgement. And if the activity of synthesis is something else there is no reason to suppose that the forms of synthesis correspond to the forms of judgement.

The argument from the concept of a subject is not what Kant calls the subjective deduction. This latter investigates the faculties of the mind involved in knowing; it

> geht darauf aus, den reinen Verstand selbst, nach seiner Möglichkeit und den Erkenntniskräften, auf denen er selbst beruht, mithin in subjektiver Beziehung zu betrachten ... (A xvi.)

He says it is strictly dispensable, and that

> im Fall meine subjektive Deduktion nicht die ganze Überzeugung, die ich erwarte, bei ihm [dem Leser] gewirkt hätte, doch die objektive, um die es mir hier vornehmlich zu tun ist, ihre ganze Stärke bekomme ... (A xvii.)

The course of the argument is roughly as follows. If experience is not to be haphazard and arbitrary, the manifold of intuition must be synthesised by the mind to introduce order.[1] The synthesis must be carried on by that mental

1. A 99; A 100f.; A 104; A 116; A 120; A 130; B 129f.; B 135; B 164.

faculty the function of which it is to bring together concepts and intuitions: this can be called imagination, but it is something more general than the faculty for reproducing mental pictures.[1] But besides sensibility and imagination there must also be a third faculty. For we must think of ourselves as continuing subjects; and this is not something we can imagine, for imagination has always to work on a sensible manifold, on intuitions. So we must have a faculty of pure thought, capable of being used (even though this use may be illegitimate) independently of any intuition.[2] This faculty of thought (which, when it is applied to intuition, is called understanding) provides the rules which govern synthesis, for all synthesis involves conceptual recognition -- otherwise it would be haphazard. Since they are supplied by the understanding these rules must be a priori concepts: i.e. categories.

There is no straightforward way of parcelling out passages as belonging to the subjective rather than to the objective deduction; which is not surprising, for if the above account is right the subjective deduction does not so much offer a new argument as put the same points as before

1. A 120; A 124; B 151f.; B 164.
2. A 107f.; A 113; A 116f.; A 122ff.; A 129; B 131ff; B 136ff.; etc.

in what seems a rather more objectionable way, going beyond what we had before only by making further claims of an apparently psychological sort. The question of our attitude towards it is the question of what is to be made of Kant's "transcendental psychology", as Strawson calls it. And here the first thing to do is to determine in what it really consists.

Kant sometimes speaks as if he is engaged in an entirely unverifiable investigation of the Ich an sich,[1] and he has frequently been interpreted in this way.[2] If this were what he is doing he ought to regard himself as refuted by his own arguments in the chapter on the Paralogisms. This does not prove anything, but it makes it very desirable to see if what he says is susceptible of some more harmless interpretation.

Even when deep in the study of transcendental synthesis and the transcendental unity of apperception -- when he sounds at his most psychological -- Kant makes it clear that he is still using the word "transcendental" in the same way as before; though now it is not just knowledge of the conditions of the possibility of experience that is called transcendental, but any of these conditions themselves.[3]

1. E.g. B 155, B 157.
2. Notably by Kemp Smith (Commentary to Kant's Critique of Pure Reason, London, 1918). Also Strawson, op.cit., p.32.
3. B 132, B 151.

Thus he calls the unity of apperception transcendental just because he considers it to be such a condition; the same goes for transcendental synthesis, and for the other cases. The purpose of the subjective deduction, and hence the purpose of talking about these features of the mind and its activity, is to show that if knowledge or experience is to be possible certain things must be true about the mind of the knower. One can call this a priori psychology: but that does not make it out of place in philosophical argument. Why should transcendental arguments be legitimate so long as they impose conditions on the world experienced, and illegitimate as soon as they start to place restrictions on what the mind of the knower can be like?

If we are to dismiss out of hand an argument to the effect that experience requires synthetic activity by the mind, we ought to dismiss equally swiftly the argument that if we are to be capable of applying predicates we must have an ability or tendency, which is not further justifiable, to treat things as similar or different in various ways. — This ability would not be an ability to detect similarities already objectively present in things, but just the readiness to make certain co-classifications and to refuse to make others - on no basis whatever. For if there could be objective similarities in things, apart from people's tendencies to co-classify them, these objective similarities could do no work: if a and b are "objectively similar"

in many respects, and a and c "objectively similar" in no respects, there is no logical bar to prevent me, or everyone, classifying a and c together as similar in many respects and quite failing to see any similarity, however thorough our examination, between a and b. And someone else, or some being of a different sort, might see similarities where I see none and differences where I see similarities; without either of us being wrong, or any more right than the other. — This argument, which is Wittgenstein's[1], is a transcendental argument reaching a conclusion about what the mind must be like: it must have a predisposition to classify some things together and others differently. But surely this argument cannot be dismissed as unworthy of attention merely because its conclusion is that the mind of the perceiver must be thus and so, and not that the world he perceives must be thus and so for him to be able to have self-conscious experience of it. Indeed, it is not clear that there is much difference between Kant and Wittgenstein here; to say that the mind must by synthetically active in forming concepts is not far from saying that we must have a predisposition, which cannot be further grounded or explained, to classify things in various ways.

1. Cf. e.g. Philosophical Investigations (Oxford, 2nd.edn., 1958), I sects. 189, 191, 206ff.; Remarks on the Foundations of Mathematics (Oxford, 1956), part I.

However, the fact that his arguments purport to show something about the mind of the knower raises a problem for Kant which he does not adequately solve, and which leads him to say extravagant things to which Strawson rightly objects. He considers he has shown that the mind must be spontaneously active (selbsttätig); yet it cannot intuit itself as active, since to intuit something is to treat it as given and as being acted upon.[1] His answer to this is that although we do not intuit ourselves as active we are nevertheless conscious of our activity in transcendental apperception.[2] But this consciousness of one's own activity is not knowledge, for it does not conform to the conditions for knowledge. Now, if the proposition "The mind is spontaneously active in synthesis" cannot be known to be true, how can it figure in the transcendental deduction? It will not serve Kant's purposes to say just that it is transcendentally necessary that one think of oneself as active; for his arguments to go through synthesis must really take place, not just be thought to take place. And if "The mind is spontaneously active in synthesis" does not conform to the conditions of knowledge, being a synthetic proposition for which no intuition can be given, its status

1. B 153.
2. B 157.

can only be "problematic", like that of the claims of the metaphysicians; and Kant's arguments cannot go through. So he can justly be accused of being a transcendental psychologist in his own meaning of the phrase: he makes assertions about the self which the arguments of the Transcendental Dialectic destroy.

He could have got out of this difficulty. For he could have admitted that we can know that synthesis takes place, and thereby know that the mind is active, just as we know that a man is acting on a ball by kicking it. The fact that the corresponding intuitions are "given" does not prevent us being aware that some things we experience are acting and others being acted on. What Kant cannot allow is not this, that we can experience active things, but the further claim that we can experience spontaneous, self-active, things. To know that the mind is spontaneous in synthesis would be to know that it, and not something external to it, is the source of its synthetic activity, i.e. to know that it is free; and the Second Analogy rules out the possibility of knowing such things. So it is open to him to say that it is transcendentally necessary that I, as a self-conscious percipient, should be active in synthesis; and transcendentally necessary that I should think of myself as self-active, free; but not transcendentally necessary that I should actually be self-active. And this, I think, would serve the purposes of his argument.

This difficulty over the mind's spontaneity does not arise because it is in general illegitimate to make transcendental arguments about what the mind must be like. It arises only given the thesis that we cannot experience completely free actions: if Kant permitted himself to say that the spontaneity of the mind can be intuited he would have no further problem about its intelligibility.

Strawson's objections to transcendental psychology are largely objections to transcendental idealism - the doctrine that what is a priori is imposed by the mind on reality.[1] But this doctrine plays only a very subsidiary part in the Deduction; even in the subjective deduction. Kant gets to it by a confusion over what it is for a condition of the possibility of knowledge or experience to be a priori, and so transcendental[2]: he thinks that whenever he shows something to be a necessary condition of knowledge or experience he thereby shows it to be a priori and consequently imposed by the mind on experience. He confuses four separate things here.

(1) It is an a priori truth that p's being the case is a necessary condition of experience.

This is what a transcendental argument establishes. In some cases p either has the form "The mind possesses concept x" or else entails a proposition of that form; this would be

1. Bounds of Sense, p.21.
2. B 25; Ak. XX: 274; see above, pp.4ff.

so for the categories. In such a case there is a good sense in which the concept x must itself be a priori:

(2) The concept x is not acquired by abstraction from experience.

For it cannot be derived from experience if to have any experience at all one must already possess the concept x. However, unless we assume transcendental idealism to start with, we cannot be sure that all cases will be of this sort; there may be cases where p in (1) has the form "The world is of such-and-such a sort" (e.g. not completely chaotic), and says nothing about what concepts a percipient must have. In these cases, the concept which p says applies to the world (e.g. orderliness) may, so far as the argument goes, be acquired by abstraction from experience if it is acquired at all; though one may still call it an a priori concept if all one means by that is that it must apply if experience is to be possible.

In the other class of cases, those in which p either is or entails a proposition of the form "The mind possesses concept x", p may go further than the mere ascription of a conceptual capacity to the percipient and say that the concept x applies to items in the world he experiences. Thus, it has often been held that transcendental arguments can show that we must not only possess the concept of an unperceived existent, there must actually be items in our experience to which we can apply this concept; and Kant

holds that we must not only have the pure concepts of the understanding which are the categories, they must also apply throughout experience. In these cases we can assert, as well as (2),

(3) We can know a priori that the concept x applies to the world.

But it is not obvious that whenever we can assert (2) we can also assert (3), for there may be transcendental arguments which only show that one must have certain concepts and not that one must be able to apply them to the world. For instance, Strawson argues in Individuals that one must be prepared to ascribe states of consciousness to others if one is to ascribe them to oneself, but he recognizes that there may not actually be other people to whom states of consciousness may be ascribed.[1]

(3) presents the problem about synthetic a priori propositions which led Kant into transcendental idealism. In a memorable sentence in the preface to his Nouveaux Essais sur l'Entendement Leibniz had said

> Car si quelques événements se peuvent prévoir avant toute épreuve qu'on en ait faite, il est manifeste que nous y contribuons quelque chose du nôtre.[2]

1. P. F. Strawson, Individuals (London, 1959), p.99 and fn.
2. Leibniz, Philosophische Schriften (ed. Gerhardt) V: 42f.

Kant takes over this solution - although, unlike Leibniz, he sees that it is not necessary in order to account for analytic truths, since these concern relations between concepts only and do not say anything about the world. Synthetic truths can only be a priori, he holds, if their truth is guaranteed by the mind's creating the experience in which they apply and impressing on it these a priori features. Otherwise they can at best be very highly confirmed empirical generalisations, lacking that necessity and complete universality which characterize the a priori.[1] And so according to Kant if we can assert (3) we can assert

(4) The concept x is imposed by the mind on experience.

Indeed, he thinks we can assert something like this whenever we can assert (2) or even (1), because he thinks we can always assert a statement like (3) under those conditions.

We have found him to be wrong on the latter point; he is wrong also on the former - (3) does not commit one to (4). One must distinguish between

(3a) We can know it to be necessarily true of the world that if anyone is able to experience it he must be able to apply the concept x to it

and (3b) We can know it to be necessarily true of the world that the concept x must apply to it.

1. A 2 (for example).

(3a) is an expansion of what is (or should be) meant by (3); it does not entail (3b), which raises the problem of which (4) is Kant's solution. If one asserts (3a) one can recognize that the world might change in such a way as to make experience impossible: it is only a contingent truth that it is arranged in a way that one can deal with; (3b) makes this contingent truth into a necessary one. Nothing licenses the move from (3a) to (3b).

So this way of getting to transcendental idealism will not do, and it seems perfectly possible that transcendental arguments should reach conclusions about the mind without thereby entailing such an objectionable doctrine. For a condition may be transcendentally necessary (1) and about the mind without being about concepts that the mind must have (2) - it may say, perhaps, that the mind must be receptive to impressions from outside. Even if this is denied, not all concepts that the mind must have (2) are concepts which must find application in experience (3), = (3a) - the world may contain nothing of the appropriate kinds. And even if this is denied, it can be a necessary truth that a concept must apply to the world if experience is to be possible (3a) without its being any the less a contingent truth that that concept applies to the world (not-3b) - it is only contingent that there is experience.

So Kant could say the psychological things he says in the subjective deduction without going on to commit himself to transcendental idealism. For it is possible to hold that the mind needs various faculties in order to experience things

without holding that these faculties create the world of appearances.

One might, however, say that transcendental arguments are an altogether insufficient basis for an elaborate faculty-psychology, which is in effect what the subjective deduction provides; this is another standard objection. I think it is unfair to Kant, because of the extreme fluidity of his terminology. If he really thought, for example, that there was an important distinction to be drawn between understanding and imagination, we should not expect to find him ascribing to them at different times the same role (of synthesising the sense-manifold) without a word of explanation.[1] If one sticks rigidly to an interpretation in terms of faculty-psychology, one will be forced to have recourse to a patch-work theory; it is partly because Vaihinger and Kemp Smith do read the subjective deduction in this way that they adopt the patchwork theory to account for the inconsistencies.[2] They are then faced with the picture of Kant as - in Paton's rather memorable words - "a pedantic old professor, armed

1. B130 vs. A 124 or B 151f. Cf. also (for example) the question whether reproductive synthesis can be pure - A 101 vs. A 118 - which Kemp Smith and Vaihinger puzzle over.
2. See esp. Vaihinger, "The Transcendental Deduction of the Categories,,,", reprinted in Gram, Kant: Disputed Questions (Chicago, 1967), p.26.

with an external architectonic, incompetently tacking together old notes of what he used to think."[1]

If we are not to read it as a gratuitous essay in faculty-psychology, the argument of the subjective deduction will be that experience involves synthetic mental activity which in turn involves the categories. So it is not very different from the other two arguments (those from judgement and from the concept of an object/subject). And the familiar dilemma recurs: here again we can either say that synthesis is a matter of making judgements, so that the forms of synthesis just will be the forms of judgement, in which case we have the same rather trivial conclusion as the argument from judgement gave us; or we can say it is <u>not</u>, or not <u>just</u>, a matter of making judgements, in which case the connection between the forms of synthesis and the logical functions of judgement is left obscure.

Perhaps the reason Kant fails to distinguish these two alternatives is that he is caught by an ambiguity in the expression "forms of synthesis": for this might mean "sorts of synthetic activity", in which use it is parallel to the use of "form" in "forms of judgement"; or it might have a

1. <u>Kant's Metaphysic of Experience</u> I:56. I must admit, however, that the fact that there are several different arguments for the same conclusion in the Transcendental Deduction is evidence for some sort of patchwork theory, particularly in the light of Kant's assertion that there can only be one transcendental argument to any particular conclusion (A787ff/B 815ff.) - though he does not say what criteria he has in mind for the identity of arguments.

more technical sense, whereby the concept - the rule according to which the synthesis is carried out[1] - is thought of as form and the intuitions which are synthesised are thought of as matter. Kant does use "form" in this way too, for he speaks for example of space and time as forms and Empfindungen as matter. These two uses of "form" do not have anything in common; it seems to be only by conflating them that Kant can appear to be putting forward a really exciting thesis.

So the transendental deduction of the categories is of interest to us here not only as providing the paradigm example of Kant's method at work. His arguments, so far as they are valid, establish only rather disappointing conclusions; and this may raise a doubt whether transendental arguments can ever show anything really substantive. I shall argue that they can, though much less than is usually supposed.

C. General Criticisms of the Deduction

Four general criticisms are made of Kant's procedure in the Deduction. It is claimed by de Vleeschauwer[2] that the argument proceeds by synthetic a priori steps, or confuses analytic steps and synthetic a priori steps; that it is circular; and more generally - the doubt raised at the end of the last section - that no argument of this sort could ever establish anything which it did not covertly import into its

1. A 112ff.; A106; B 145.
2. H.J. de Vleeschauwer, La Déduction Transcendentale (Paris Antwerp and the Hague, 1934-7).

premisses. The fourth criticism is that arguments like the transcendental deduction are too abstract: it makes no sense to talk about experience radically different from our own, and hence to ask questions about what things must be like if there is to be any experience at all.

In answer to the first, and in reply to one of the questions raised earlier about Kant's method, I want to argue that Kant is here performing conceptual analysis. His argument should therefore, if valid, be statable in a set of analytic propositions. His procedure is to see what is involved in the concept of the experience, or knowledge of objects, of a being without intellectual intuition. Perhaps he did not think about the logical status of his own argument, so that one ought to hesitate to ascribe to him the view that the transcendental deduction, and transcendental arguments generally, proceed by a series of analytic steps. But he does not hold that it proceeds by steps of any other sort[1], and it

1. I do not think that even B 23 shows that he holds them to be synthetic. The paragraph is about the analytic propositions made by the dogmatists within their metaphysical systems, which of course do not include those used by transcendental arguments. Cf. Ak. IV: 273f. (Prol. sect.4- properly sect. 2, as in P.G. Lucas' translation). But it may well be that Kant was muddled on the question.
   In Refl. 6386, Kant says "Die transzendentale Sätze a priori sind insgesammt analytisch..." (Ak. XVIII: 699); cf. also Refl. 4943. Adickes' argument (loc.cit.) that 6386 has to do only with propositions about the "transcendentals" unum, verum and bonum which are said to belong to formal logic at B 113ff. has an air of special pleading, for although these three terms appear in this Reflexion Kant is here trying to make them into transcendental predicates in his own sense, viz. as having to do with the possibility of experience, which is what he apparently thought them not to be in B.

is hard to see how he could have considered them other than analytic had he thought about the matter. The following three points are relevant.

(1) The first argument of the transcendental deduction discussed above proceeded, and was recognized by Kant as proceeding, from the definition of a judgement. Now this definition is what Kant calls an "analytic" definition, since it is a definition of a concept already understood prior to its being defined; but some such definitions are "real" definitions, to be expressed in synthetic propositions, while others are "nominal" definitions, to be expressed in analytic propositions. And this is a "nominal" definition. Real definitions are synthetic judgements because they involve predicates not obtained simply by the analysis of the concept defined: they show the Realmöglichkeit of the thing to which the concept applies, its consonance with conditions of possible experience; whereas nominal definitions do not.[1] Now the proposed definition of judgement -- as "an action, through which given representations first become knowledge of an object"[2] or as "the manner in which given modes of knowledge are brought to

1. Ak. IX: 143 (Logik sect. 106); cf. A 241-2n.; A 598/B 626; and L.W. Beck, "Kant's Theory of Definition", Philosophical Review 1956.
2. Ak. IV: 475n. (MAdN Vorrede).

the objective unity of apperception"[1] -- says nothing about whether judgement is consonant with the conditions of possible experience. One could hardly expect it to, since the aim of the argument is just to establish what the conditions of possible experience are. So it cannot be more than a nominal definition, obtained by analysis of the concept and expressible in an analytic judgement. And since, in this argument, the core of the transcendental deduction is provided by this definition, it does not proceed by synthetic a priori steps.

(2) The other versions of the deduction also appear to be pieces of conceptual analysis, not involving synthetic propositions. In B Kant himself emphasises the analytic character of the fundamental principle of the necessary unity of apperception.[2] And one would not expect to find the argument of the deduction proceeding in some versions by analytic, in others by synthetic a priori, steps.

(3) Although Kant does not make it perfectly clear what he means by "synthetic a priori", he does not leave it totally obscure either. A judgement is synthetic if the concept of the predicate adds something not already given in the concept of the subject; or, more generally, synthetic

1. B 141.
2. B 135, B 138.

judgements are those the truth-values of which are not determinable by means of the law of non-contradiction alone.[1] Now judgements of this sort can only be made validly when we can go beyond the interrelations of concepts, to intuition: for there must be a third thing besides the concepts of subject and predicate, to provide the connection between them.[2] This is a central doctrine of KdrV, for the mistake of the dogmatic metaphysician is to think he can say something substantive about things which cannot be intuited, whereas if he can say anything true about them at all it must be analytic. Synthetic judgements may go beyond cpncepts by being concerned either with the contents of our spatio-temporal intuitions or with the formal features of these intuitions; in the former case they are a posteriori, in the latter a priori. And Kant considers that there are two ways in which valid judgements of the latter sort can occur: either by the construction of concepts in pure intuition, as in mathematics, or by its being demonstrable that the judgement must be true if experience is to be possible.[3] -- It is important to distinguish here between the synthetic a priori judgement and the judgement that the synthetic

1. A 7/B. 10f.; A 150ff./B 189ff.; Ak. VIII: 228.
2. A 155/B 194; A 721/B 749. Intuition is the "third thing" only for theoretical, not for practical, synthetic judgements.
3. A 721ff./B 749ff.; A 736f./B 765f.; A 783/B 811; A 788/B 816.

a priori judgement must be true if experience is to be possible; the latter cannot, on this account, be a valid synthetic *a priori* judgement, and so presumably *must* be analytic. "Every event has a cause" Kant would call synthetic *a priori*; "It is a condition of the possibility of experience in time that every event have a cause" would, if true, be analytic. Moreover, Kant often speaks of his task as being to show how synthetic *a priori* judgements are possible; he can scarcely be said to have done this if his solution has itself to be expressed in judgements of that suspicious variety.[1]

The arguments of the deduction, like all Kant's arguments, will thus be directly open to attack from those who reject the analytic/synthetic distinction. They deny that it makes the right sort of sense to talk of necessary conditions. Instead, they allow us to speak of some assertions as more basic than others in our conceptual scheme in that we are less ready to abandon them to suit the demands of an economical account of experience; or they may allow us to formulate general linguistic hypotheses to

1. Passages like A 10/B 13 and those in which metaphysics is said to be synthetic *a priori* are not relevant. For Kant is presumably speaking there not of what he is doing in the *Critique* but of that speculative knowledge out of pure reason which the transcendental arguments of the *Critique* are supposed to make possible.

the effect that certain features are to be found in all actual languages. This would let us produce arguments a bit like transcendental deductions, and these might have interesting conclusions; but they would not justify the use of particular concepts in the way transcendental deductions do. For they could never show that without a particular concept we could have no experience at all.

I shall discuss this attack at length in chapter III. It marks out a possibility, and charts some consequences of this possibility's being actual; and the possibility is, that all human behaviour should turn out to be fully explicable by means of efficient causes. This possibility, I shall argue, cannot be excluded without a completed science; but having said that, I think we can put it to one side. One could put the point by saying that if those who deny the analytic/synthetic distinction are right there is no language at all and hence no experience, as we usually conceive them, for we think of the use of language as essentially rule-governed and intentional. Arguments about the necessary conditions for experience of a certain sort are of no great use against people who deny that there can be experience of that sort anyway; against this position one can only argue either that it is incoherent or that it doesn't fit the facts. In this case it does not appear to be the former, and whether it fits the facts is what we can't know for certain without a completed science.

The second general criticism of the Deduction was that Kant's argument is circular. He wants to know how synthetic a priori judgements can be possible, and his answer is that they must be if experience is to be possible. But often experience seems to be defined as involving synthetic a priori judgements: experience of objective items is marked off from the chaos of mere sensation by the fact that it involves synthetic a priori judgements.[1] This further suggests that Kant cannot maintain the distinction he draws in the Prolegomena between analytic and synthetic method, whereby the analytic method -- that used in the Prolegomena -- takes it as given that some synthetic a priori judgements are possible and investigates what must be true as a result of this, in the hope of shedding light on the question of the possibility of synthetic a priori judgements of other sorts; while the supposedly more satisfactory synthetic method of the Critique takes "nothing as given except reason itself", and, by investigating how reason works and presupposing nothing about the validity of any synthetic a priori judgements, comes to show that such judgements are possible.[2]

But Kant does not build it definitionally into his initial premise (sc. that there is knowledge, or experience)

1. See H. Vaihinger, Kommentar (2nd ed., Berlin, Stuttgart and Leipzig, 1922) I 440; also de Vleeschauwer, op.cit. II 395f.
2. Prol. sects. 4 and 5.

that there should be knowledge of the truth of synthetic a priori judgements.[1] The connection is not definitional, though it is meant to be analytic -- for the argument of the Deduction is analytic; and since in some versions the Deduction is not very complex it should be fairly immediately analytic. For the aim of the Deduction is to establish synthetic a priori propositions by showing that unless they were true experience of any sort would be impossible.[2] So if there is a circle it is not so obvious a one as this. As to the two sorts of method, it is true that in a way the synthetic method is not really so different from the analytic as Kant seems to suggest, since both investigate necessary conditions for knowledge or experience; but nevertheless there is an important difference, since the method of the Critique takes nothing for granted that a sceptic could at all plausibly doubt. For the doubt whether there is any experience at all is self-defeating in a way in which the doubt whether there is any pure natural science -- whether any propositions like "Every event has a cause" can be known true a priori -- is not. The sceptic who has the

1. At any rate, not normally. Vaihinger (op.cit. I 435f.) suggests that in asking in the Prolegomena "How is nature possible?" Kant conflates two questions, one about how certain synthetic a priori judgements are possible, the other the usual question of the Deduction about the possibility of experience.
2. There is of course a step from here to the thesis that we must know (some of) them to be true, but Kant does not notice this; presumably because of his idea that whatever is a priori is mind-imposed, so that the mind cannot be altogether ignorant of it.

former doubt must doubt everything and cannot regard even the expression of his own doubt as intelligible.

But it is sometimes argued -- and this was the third general criticism of the Deduction -- that all transcendental arguments are circular, or that if they are not their conclusions are uninteresting because trivial. It is said that all such arguments covertly build their conclusions into their premisses by making it a definition of "knowledge" or "experience" that they must be of the sort in question; and that if transcendental arguments ever do establish anything at all what they establish is entirely unexciting, e.g. that our thought about the world must obey the laws of logic, while none of the really interesting claims, e.g. that experience must contain causal sequences, could ever be established by this means. There is something in this objection, in that most of the transcendental arguments that have been put forward do establish their conclusions, if at all, only by building them into the idea of knowledge or experience from which they start. But I think perhaps some results can be salvaged which, though less sensational than those sometimes claimed (particularly by Kant), are yet worth having. In the next section I shall consider some of the transcendental arguments that have been put forward in recent years in order to show this.

Before going on to that, I should like to consider the fourth general objection to the Deduction. This was that we

can't use transcendental arguments at all because it makes no sense to talk about an experience radically different from our own: so that if the question which features of experience are essential to it is different from the question which features are present throughout all experience like our own about which we can ever find out anything, this question has no answer. From within our own conceptual scheme we can show how certain concepts function and that some are, perhaps, basic or central; but we can never get outside our own conceptual scheme to compare it with other possible ones, so that we can never say a concept is essential to <u>any possible</u> conceptual scheme -- the idea does not make sense.

But we <u>don't</u> have to imagine getting outside our experience in order to carry on transcendental arguments. For these do not affirm that there <u>are</u> or even could be beings capable of experiencing in relevantly different ways from ourselves; it is just that sometimes the best we can do with them is to show that (at least) a few features of our experience cannot intelligibly be thought away, which leaves us (so far) at liberty to suppose that those features for which this has not been shown <u>can</u> intelligibly be thought away. And it may be that on this basis we have no reason to <u>deny</u> that there could be beings radically different from ourselves.[1] The transcendental argument simply points out

1. Cf. K.d.U. sect. 77.

that unless something were the case we could have no experience. — An objector now going to the other extreme might suggest that transcendental arguments are still useless, because it is possible that there might be beings who did not fulfil the conditions for having a language or having experience, but who nevertheless had something similar. But then, if the transcendental arguments are right, this something cannot be similar in the relevant respects; beings for whom the law of non-contradiction does not obtain could not have a language. (And it doesn't make any difference to say, "But their signals would be intelligible to <u>them</u>": to suppose this true one would have to evacuate all content from the idea of intelligibility).

D. Some Transcendental Arguments.

Let us return to the argument that the only way in which we can perform transcendental deductions is by building so much -- perhaps covertly -- into the notion of a language or an experience that the conclusion follows trivially from it. Now it is to some extent a matter of taste what propositions we consider trivial. The nerve of the deduction has to be an analytic proposition, but presumably not all analytic propositions are to be condemned as trivial: only those which are ~~not~~ obvious. And even if an analytic proposition is obvious immediately it is stated, we might consider it non-trivial if philosophers or ordinary men were inclined to affirm the antecedent and deny the

consequent until it was clearly pointed out to them what they were doing.

In this section I shall examine a few transcendental arguments which have enjoyed some popularity in recent years. They are rather similar to Kant's deduction of the categories in that they attempt to show that certain things must be the case if experience is to be possible at all. I cannot consider all the transcendental arguments that enjoy some current vogue, but I think the ones I shall consider are fairly representative and have claims to be considered as among the most interesting. We found with the transcendental deduction of the categories that a conclusion could be reached, though a less interesting one than Kant originally thought; we shall find something rather similar with many of the arguments I shall discuss in this section.

Strawson on Causal Order.

We can start by examining Strawson's argument to the effect that our experience must be largely, although not necessarily entirely, causally ordered.[1] This argument is itself a replacement of the argument of Kant's Second Analogy, which seeks to establish the stronger conclusion and which I shall not discuss, since its deficiencies are well-known, at least in outline.

1. Bounds of Sense, pp.143-6.

Strawson's view is that the things we experience must for the most part be causally ordered, since otherwise we should be unable to think of them as enduring and re-identifiable objects. Let us grant, for the moment, that we have to be able to apply the notions of enduring and re-identifiable objects if we are to have experience at all; how does Strawson's conclusion follow? His answer is that things must change in regular ways if we are to be able to think of them as <u>persisting</u> things.

> Objects may change; but they must not, so to speak, change out of all recognition. If they did, we could not know they had; for we could not recognize <u>them</u> as having changed. (p.144.) ... our concepts of objects, and the criteria of re-identification which they embody, must allow for changes in the objective world subject to the limitation that change must be consistent with the possibility of applying those concepts and criteria in experience. How is this requirement satisfied? The answer seems to lie in the fact that our concepts of objects are linked with sets of conditional expectations about the things which we perceive as falling under them ... concepts of <u>objects</u> are always and necessarily compendia of causal law or law-likeness, carry implications of causal power or dependence. (pp.145f.)

It is, I think, incontestable that if we are to be able to apply to experience the concept of a persisting object, experience must exhibit some degree of regularity -- it cannot be totally chaotic. It is also true that we must

make generalisations about these regularities, and generalisations capable of supporting counterfactuals: the enduring object is envisaged as continuing to exist when unperceived, and because we think of it as continuing to exist we think of the same, or roughly the same, regularities as applying to it. The concept of an independent persisting object is the concept of something about which counterfactual propositions are true; we can only reasonably believe in the truth of counterfactuals when we can derive them from some law or lawlike generalisation. So, if we were right to grant that we must apply the notion of a persisting object in our experience, we are committed on transcendental grounds to making lawlike generalisations -- despite the fact that these always go beyond the available evidence in their favour.

However, although it may be transcendentally necessary to make such generalisations, it has not been shown to be necessary that they should apply to the future. The counterfactuals they support may all relate to the present or the past. And surely this is as it should be; I am not the less able to use a language and to know things now if all the laws of nature are about to break down, though then every one of the predictions I consider warranted on the basis of past evidence may be falsified. It would be reasonable for me in the present circumstances to make these predictions, since it is always reasonable to assume that the future will be like the past; but we are not concerned here with what would

be reasonable. Is it transcendentally necessary, i.e. a necessary condition of my having experience, that I should regard my generalisations as covering future instances as well as present and past ones? Strawson's argument has not shown that it is. So it would seem possible to get by with counterfactual-supporting generalisations which were not causal because they did not license any predictions. - I cannot prove that it <u>is</u> possible; generally speaking, it is very difficult to establish satisfactorily that something is not transcendentally necessary, for a new transcendental argument <u>may</u> always turn up; one would have to show that the supposition contains no latent inconsistency.

In this particular case, there is another transcendental argument which is sometimes supposed to fill the gap. It is that we must use generalisations which are quite unrestricted as to temporal conditions if we are to have a language: for communication involves treating things as instances of general characters, as being of certain sorts, and this involves abstracting from features like the particular times and places in which things are.

> ...we could not talk at all if we could not describe two objects as being of the same sort, merely on the grounds that they were different in date or position, or that the speaker and hearer thought differently about it. Language depends essentially on the type-token distinction, and would be impossible if the world were not correspondingly conformed to the rules of sorts and

instances.[1]

But although of course predicates like "is a horse" and "weighs 200 pounds" carry no built-in restriction to particular times, and although Strawson is right to say that many predicates of objects (like these ones) apply only because certain generalisations hold, we do not have to assume that these generalisations will continue to hold. For these predicates may cease to apply to anything. We should refuse to call anything a horse or a table unless it reacted to tests in certain regular ways -- there are generalisations which something must satisfy if it is to count as a horse or a table; but the possibility that the generalisations cease to obtain is simply the possibility that there should no longer be any horses, tables, things weighing 200 pounds, etc. The fact that such predicates are not semantically restricted to particular times does not show that they must be instantiated at all times, or even that we must think them to be.

So I do not think that these arguments show it to be transcendentally necessary that we should make regularity-generalisations which support predictions. And it does seem an important feature of causal generalisations that

1. J.R. Lucas, "Causation", in R.J. Butler (ed.), Analytical Philosophy, First Series (Oxford, 1962), pp.47f. Cf. also R. Harré, "Particular and General", in Williams and Montefiore (eds.), British Analytical Philosophy (London, 1966).

they do support predictions; they raise problems about induction which do not arise for the limited though counterfactual-supporting generalisations which we must have if we are to be able to apply the concept of a persisting object. Strawson's argument, even as added to, does not establish the transcendental necessity of making causal generalisations.

Regularity

It may be felt, however, that the argument is nevertheless valuable because although it does not establish the need for causal generalisations it does establish that for regularity generalisations which support counterfactuals, even though they may not support predictions. I do not wish to deny that this is an interesting conclusion. But part of its interest may be thought to lie in the idea that experience must itself be regular: a totally chaotic experience would be something no being could grasp.[1] This, however, would be wrong. To grasp experience at all we must see it as containing recurrent features, and if we are to apply concepts of enduring objective items in our experience we must see it as exhibiting regularities. But this imposes no limitation on what experience itself may

1. Cf. Bennett, Kant's Analytic, pp.33ff.

be like - on what the data presented to us may be. It is a feature of my psychology (or, to put it another and less fundamental way, of the entrenchment of predicates in my language) that I see as similar the things I do see as similar - that I see two blades of grass as both green rather than fail to see anything in common to them both. Of course, given that I see similarities and differences as I do, the world must have a certain sort of order for me to be able to experience it. But since whatever could happen would have to happen in some order in any possible world, there might always be a being for whom any given set of experiences which we consider totally chaotic was such that he could see similarities and regularities in it sufficient to apply concepts of enduring objects.[1] Such a being would co-classify - see as similar - items which we see as having nothing in common, and might quite fail to see any similarity at all between the two blades of grass. Its mind would thus work differently from ours, but it doesn't seem possible to say that we are right and it wrong in the detection of similarities; all we can say is that we and it proceed differently here. And that is no reason for supposing that it cannot have self-conscious

1. Cf. Wittgenstein, Philosophical Investigations (Oxford, 1953) §§ 185ff., 217ff., 239ff.; Remarks on the Foundations of Mathematics (Oxford, 1956) ad init.

experience of persisting items which it can express in language - though no doubt we could not understand its language, nor it ours.

We are left, then, with the conclusion that if we are to have experience we must make regularity generalisations which support counterfactuals. But we reached this conclusion by granting, for the sake of argument, the premise that if we are to have experience we must be able to apply the notion of an enduring and re-identifiable object. Now I am inclined to doubt whether this premise is true, unless it is made true by definition. Strawson does, of course, have an argument for it: so here is another transcendental argument to examine. He says that any experience involves placing items in an objective time-order, and that if we are to be able to distinguish between subjective and objective time-order we have to have the notion of objects persisting unperceived through time. Setting aside for the moment the need for an objective time-order, let us consider the other part of the argument. The distinction between a subjective and an objective time-order, he says,

> is effectively employed only if we think of objects actually encountered in experience, objects which we actually perceive, as existing not only when we perceive them, but also at other times, when we perceive, not them, but other objects.[1]

1. Bounds of Sense, p.141.

Let us grant that we must admit items which can exist unperceived if we are to have an objective time-order; this will be examined later. Strawson is claiming that these items must be persistent objects. And this is not obvious. If one supposes that events occur in regular patterns one can imagine having good reason to believe later members of a pattern are occurring when one is not perceiving them (perhaps because something else now occupies one's perceptual field) and one has just perceived the earlier members of the pattern. Strawson fills out his argument by saying that if my idea of the time-order is such as to allow it to include more than just my own experiences I must have the idea of something which exists unperceived and which hence occupies some other place than does what I do perceive.[1] But, he goes on,

> For this to be possible we must see objects as belonging to, and events as occurring in, an identical, enduring spatial framework. For this in turn to be possible, we must have empirically applicable criteria of persistence and identity, embodied in concepts under which we bring objects of non-persistent perceptions.[2]

1. Op.cit. pp.125, 128, 132.
2. Op.cit. p.132.

But why could not patterns of events give us the idea of a spatial framework (or a framework like the spatial one[1]) just as readily as persisting objects can? The idea of an event does not in fact play this role in our experience, and Strawson is no doubt right in Individuals to say that in general for us the identification of events depends on the identification of persisting things[2]; but that does not show that it might not be otherwise if the character of our experience was different.

Let us consider a world in which the only things of which I am aware are flashes of white light; I discern no regularity in their occurrence. There seems to be a good reason for saying that I cannot know such a world, namely that the lack of regularity makes it impossible to provide a foothold for the idea of something which exists unperceived.[3] But suppose we introduce some sort of order: we can imagine that the flashes come in regular patterns at different places and times, that brighter ones obscure less bright ones, etc.,

1. This, of course, is all Strawson really demands: see p.128.
2. Individuals (London, 1959), pp.46ff.
3. Whether it is even necessary to make use of this idea will, of course, be considered shortly.

and in this way give content to the idea of things going on unperceived. For example, there may be a series of dim flashes in the middle distance moving gradually across my field of vision, which is for several flashes obscured to me by a much brighter and more rapid series which suddenly appears at the top of my field of vision and soon disappears at the bottom. If the percipient can alter his field of vision he can easily check on the presently unperceived continuations of these series beyond his present field of vision, and even if not he can surely derive the idea of unperceived existence from cases of brighter series obscuring less bright ones. The flashes occur in series or patterns, and one might say that these series have continued existence, may come into and go out of existence, may move about and may remain stationary (though they must not all move so fast that we are unable to detect the regularities), and can undergo qualitative change - though under some circumstances there will be needed a decision (perhaps unconsciously built into our percipient's language) as to whether the old series has altered or a new one taken its place. And since they occupy different places at the same time, the percipient will be able to speak of them as nearer or farther away in space, and their temporal extendedness will enable him to say the same about them as regards time. And so it would seem he is in just the same situation as we are, in our world, regarding space and time: for him, as for us, the identification

of spatio-temporal points and intervals presupposes the identification of things in space and time (only for us persisting objects, for him event-patterns), and the identification of these presupposes in its turn the spatio-temporal matrices.[1]

It may be said that this only works because the world of flashes is made so regular that we really have room to introduce the idea of a flash-series as a persistent object. And this is true; but there is no reason why we should treat it as a persistent object - as a substance rather than as a collection of events.[2] The idea of something persisting is only needed to make it possible for us to carry out tests on the objective world to discover whether or not we have described it correctly; and this can be done equally well if it is patterns of like events, rather than substances, which persist.

However, the fact that we have to make the world of flashes regular in this way shows that Strawson's conclusion that we have to be able to make regularity-generalisations is

1. Cf. Strawson, Individuals, p.37. If this account of the relation between the identification of things and the identification of places were rejected, still one would expect whatever view replaced it to be stateable in terms of patterns of events as easily as in terms of persisting objects.
2. If he is self-conscious the percipient will think of himself as a continuant. So if he comes across other percipients and recognizes them as such he will have to consider them continuants as well. But it might be that he never does come across any others, and so does not have to apply his idea of a persistent substance. See below, pp. 135 f., 146 ff.

justified: not just because it follows from his claim that we have to use the concept of a persisting object, but because otherwise there would be no room for the idea of anything (even a series of flashes) existing unperceived. And I think the conclusion is indeed justified - provided we don't think regularity is an inherent feature of the world, and raise questions about how much regularity is needed. What sort of ordering, and how much, he needs to get the idea of an item existing unperceived will depend on the percipient - in particular on his intelligence. But the conclusion can only be tentatively asserted until we have examined the other assumptions that were made en route. I shall for a while postpone the question of whether experience would be possible in which one could not use, or did not have, the idea of an item existing unperceived. But Strawson spoke of an objective time-order, and a spatial order or something analogous to it; and I used these in describing the imaginary world of flashes. Are spatial and temporal ordering essential to any experience - or only, as Kant says, to human experience?

Time and Space

The need for experience to be temporally ordered follows from the claims that for a self-conscious being to have experience (in the required sense) it needs language, and that to have a language it must be capable of following a

rule, and hence of making and correcting a mistake. These claims will be examined later on; I think they are right, although neither so obvious nor so unambiguous as they are generally taken to be. But the making and correcting of mistakes can only go on in time: to recognize that you made a mistake is to recognize that you were wrong and are right. One could not know only mathematical propositions for one would still also have to acknowledge that one could be mistaken about them, and in order to know this one has to give a sense to the proposition "I made a mistake about them" as objectively true or false - it must also be something I could be mistaken about. And this involves treating the event of my making the mistake at least as an objective item, and clearly it takes place in time. So if I am to have any knowledge or experience at all I shall have to have the concept of something objective in time.

The case of space is less clear. Strawson and Bennett[1] have discussed the possibility of having instead of space an auditory locating matrix, in which sounds are located by means of a master-sound of varying pitch heard together with them or by means of a regular order in which sounds are heard. This seems to me quite successful. Although Bennett

1. Strawson, Individuals, ch. 2; Bennett, op.cit., ch. 3.

does not recognize this, his view that it is sufficient for the sounds to be heard in a temporal order suggests, what I take to be the case, that nothing really analogous to a spatial matrix is necessary: there is room to apply the concept of an unperceived particular in a world which is only temporally ordered (like a solipsist's world of private sensations[1]). It may be argued against this that with only one dimension (like the temporal one) there must be insuperable difficulties in providing criteria for the identification of particular items, because we normally do this by means of the rule that no two items of the same general kind can occupy[2] the same place at the same time. This rule saves us from having to rely on a Leibnizian principle of the identity of indiscernibles; which is undesirable, because if we have to rely on the identity of indiscernibles, then so long as we know only an n-membered sub-class of the properties of an item we can never have any reason to suppose that we have only one item on hand and not two or perhaps a million, all sharing n properties but having

1. One might object to the solipsist that he has no room for the concept of a subject of experiences, or that he has no adequate checks on whether he has remembered the past rightly: these points will be discussed below.
2. i.e. snugly fit into; or, wholly occupy (i.e. be present at all parts of the spatio-temporal region).

different and incompatible properties thereafter.[1] The rule's analogue in a purely temporal world would be that no two items of the same general sort can occupy the same temporal position; but this seems unduly restrictive. We do want to be able to say that I may have two pains simultaneously; and we do not want to have to say that if I am aware of an item of k-kind at t there can be no other item, qualitatively different but also of k-kind, at t as well, existing unperceived. But this difficulty can be got round by building more into the phrase "of the same general sort". In our familiar spatio-temporal world, two items can differ very greatly and still be logically unable each wholly to occupy the same region of space-time. This would be true of a lion and a man, a quantity of air and a heap of coal; though one might dispute whether it was true of any of these and a collection of molecules.[2] But in the purely temporal world we could drastically restrict the extent to

1. D. Wiggins, "The Individuation of Things and Places", Proceedings of the Aristotelian Society, Supplement, 1963. Though one might perhaps get round this by suggesting that we could form hypotheses of uniqueness whenever we wanted to talk about individuals. But not all our knowledge can consist of hypotheses - there must be categoricals to give the hypotheses sense: perhaps it could not be the case that none of these categoricals were singular statements about objects which could exist unperceived.
2. One might, of course, identify the collection of molecules with the body it constituted. I shall not discuss the justice of this; it would involve considering the use of sortals in identity-statements. But cf., for example, Wiggins, "On Being in the Same Place at the Same Time", Philosophical Review, 1968.

which items had to differ if they were to be logically unable wholly to occupy the same temporal region (i.e. be present at all parts of it). For instance we might allow that two pains could occupy the same time; only not allow perhaps that two sharp, stabbing pains should occupy the same time.

The argument which is used by Strawson[1] for a spatial dimension or something analogous to it is that if I am to think of things as unperceived I must think of them as located elsewhere than in my perceptual field. I must recognize that my experiential route through the world is not coextensive with the world itself; my experiential route occupies all the temporal points in question, so there must be some other dimension on which things can exist if they can be unperceived. But this doesn't prove much. I don't perceive the bacteria that are doubtless on this page, but not because they are somewhere else than in my perceptual field. One might think that if I don't perceive something which exists there must be some reason why I don't, and the commonest reason, for middle-sized material objects anyway, is that they are in the wrong place. But another perfectly good reason would surely be (for instance) that my vision was obscured by a blinding flash even though the object was in the best possible position for me to see it.

1. Bounds of Sense, p.128.

Or if the thing were a buzzing sound, it might be obscured by a larger grating sound. Strawson says that the only reason we have for thinking that loud sounds can obscure softer ones is that we know this to happen in our own world, where we can have independent evidence that the softer sounds go on -- "the visible but inaudible scrapings of the street violinist as the street band marches by"[1]; but this is unpersuasive. One might as well argue that we should only believe in the existence of the unperceived tree in the quad because we can get independent evidence of its causal consequences -- radar pictures, perhaps. We say that the tree continues to exist because whenever anyone is in the right situation he does perceive it, which leads us to believe that at any time if there <u>were</u> anyone in the right situation he <u>would</u> perceive it. Exactly the same goes for sounds and for bright and dim lights: if whenever the louder or brighter one is absent the softer or dimmer one can be perceived, this leads one to hold that at any time the softer sounds <u>would</u> be audible if the louder ones ceased, the dimmer lights <u>would</u> be visible if the brighter ones were extinguished. And it seems reasonable to speak (whether or not we normally do) of a headache continuing unperceived while one listens to a piece of music

1. <u>Individuals</u>, p.71.

if the headache was there before, is there afterwards, and would have been felt at any time during the performance if one had stopped attending to the music.

So the world of sounds considered by Strawson and Bennett, or that of flashes considered above, does not require spatial arrangement or any analogue of spatial arrangement. The sounds or flashes will occur in (temporal) patterns; I may observe one pattern for a time, and then a louder or brighter pattern, and then another louder or brighter still; but if when the third goes the second continues, and the first continues when the second has gone, and this sort of thing happens fairly regularly, I may have good grounds for thinking of them as going on all the time unperceived. Such a world could be very far from chaotic, if for example every ten minutes any white flash-pattern were obscured by a red flash-pattern for one minute, and similar regularities held. Life under these conditions might be a little dull; one could not travel, as one could in Strawson's auditory world, nor get lost, as one could in Bennett's -- unless by improving, or spoiling, the acuteness of one's hearing or the sharpness of one's eyesight. But I do not see that experience of such a world would be impossible. Conversation might to some extent alleviate the dullness. For one might be able to think of oneself as having a particular pattern of sounds or flashes as one's body, this being a pattern over which one could,

within certain limits, exercise control, and one might be aware of other rather similar patterns of sounds or flashes which one could think of as the bodies of other people. And it might be possible to communicate with these, using perhaps the Morse code.[1] At least it has not been shown that this could not be done.

I conclude, then, that although experience does have to be temporally ordered, it has not been proven that it must be ordered spatially or in some other but analogous way. It may be said that I have just introduced alternative dimensions, of loudness or brightness, and in a way this is true; but their analogy with spatial dimensions is exceedingly slim[2] -- slimmer than that of Strawson's and Bennett's analogues -- and I am not convinced that anything more analogous is necessary. But none of this upsets the argument that if one is to be able to use the concept of an unperceived existent one must be able to make regularity-generalisations about one's experience. For if we are to have reason to speak of objects, or pattern of events, as existing unperceived this must be given us precisely by such a regularity-generalisation. It is because we make generalisations which support counterfactuals about the

1. Cf. Strawson, Individuals, pp.84f.
2. See also below, pp. 225f.

behaviour of things that we can come to assert claims about how things would be if one perceived them.

Private Languages: Following a Rule

Another, and even more important, transcendental argument is the Wittgensteinian one against the possibility of a private language on the grounds that the speaker of such a language could never appeal to evidence independent of his own memory claims to correct his mistakes. I shall consider the argument in two stages: first the claim that having a language involves understanding what it is to make and correct mistakes, and then the argument that one can only understand this if one can appeal to other people so that one does not have to rely just on one's own memories. The first claim I shall not dispute, but I want to examine what justifies it; I shall dispute the validity of the argument for appeal to other people.

It is a commonplace that to follow a linguistic rule is not just to behave in a regular and predictable fashion. But if a purely efficient-causal (e.g. stimulus-response) account of linguistic behaviour can be given it is not clear that anything other than complex regularities of a causal sort can be involved in our linguistic use of words. In that case it does not seem possible to justify the claim that having a language necessarily involves understanding

what it is to make and correct mistakes. It will certainly involve rule-following, if that just means behaving in a regular way; but there is no reason why it should also involve understanding about mistakes. We could of course build it into our definition of language that any speaker of one must understand this; perhaps we could appeal to the way the word is used in support of our definition.[1] But then no argument can be given to show that having a language involves this, it is simply a matter of how the words are used.

If, on the other hand, we give an account of linguistic behaviour not in efficient-causal terms, but in terms of intentions, purposes, etc.[2], the rule-following involved in having a language can involve the speaker's intentions - typically, and very roughly indeed, his intention to produce a propositional attitude in his hearer.[3] This seems to be the sort of thing people have in mind when they talk of rule-following as involving the capacity to correct mistakes; this sort of rule-following does seem to be more than just regularity. If it can also be shown that this sort of

1. Cf. Bennett, Rationality (London, 1964).
2. The relationship between these sort of account, and their respective adequacy, will be discussed in Chapter 3.
3. Cf. Grice, "Meaning", Philosophical Review 1957; and "Utterer's Meaning and Intentions", Philos. Review 1969.

rule-following necessarily involves understanding what it is to make and correct mistakes, we shall have reached our conclusion without the arbitrariness of simply defining language in such a way as to make it involve understanding about mistakes.

If A utters p because he has some intention which is thereby furthered, he must not just utter p whenever some situation s obtains; he must utter it (roughly) whenever he thinks uttering it would be conducive to obtaining the contemplated goal - e.g. of getting B to do something. At any rate it must be true of him that he would utter it (roughly) whenever he thought this would be conducive to obtaining it. But then he may well find that 'p' does not have the hoped-for effect; so he may withdraw p and utter q instead. In a language a particular utterance-type is generally used for a particular purpose, and the reason 'p' does not have the hoped-for effect may well be that the speaker has used an utterance of the wrong type to produce this effect; if he replaces it by 'q', and q is of the right type, we can describe this as his correction of a mistake. Hence it seems correct to say that the following of a linguistic rule requires a preparedness to correct mistakes, in this sense- provided always that our account of language be in terms of purposes or intentions.

Incidentally, this has a consequence which is perhaps

not trivial. This is that in any language, indeed in any purposive symbolic system, there will be pairs of expressions such that both cannot be seriously asserted at once; i.e. something like pairs of contradictories. This impossibility of asserting both seriously at once will not be due to any conventions (over and above any that may be involved in investing a sign with its symbolic significance) but due to the fact that to do so would defeat one's purpose in using the sign. Following a rule involves following it consistently.

But it is important to appreciate what the argument does <u>not</u> prove. It does not show that the speaker must have any means of <u>saying</u> to himself "I said '<u>p</u>', but that was a mistake": he just comes out with '<u>p</u>', and then when that proves unsuccessful comes out with '<u>q</u>', which is compatible with his language having no means of referring to a particular event (his utterance of '<u>p</u>') or to himself. So it does not <u>show</u> that he must be self-conscious, in the sense of being able to formulate judgements about himself; nor that his language must contain tenses or dates. It does not show, either, that he must have the idea of an item capable of existing unperceived. And this is relevant to the arguments we were considering before; for the argument that we must be able to make regularity-generalisations about our experience started from the premise that we must have the idea of something which exists unperceived - a premise

consideration of which was deferred. I shall come back to examine whether there <u>are</u> any valid arguments to support it when I have discussed the other stage of the anti-private language argument - the argument that it must be possible to appeal to other people to support one's own memories.

Private Languages: Independent Evidence

The argument has sometimes[1] been put forward in a strong form, as excluding the possibility that as a purely contingent matter there might be only one language-user in the world. Here the suggestion may be that the isolated child could never <u>learn</u> a language, and this of course not as a matter of fact but as a matter of logic: on the grounds that the terms of a language can only acquire a meaning through the presence of other people, who provide the independent checks on one's use of words which must be available if there is to be a distinction between using a word correctly and merely seeming to oneself to use it correctly. But since the argument is not just genetic, but turns on a logical point about the necessity for independent evidence if the speaker is to understand the distinction between what is the case and what seems to be the case, it must if valid exclude more than the possibility

1. Rush Rhees, "Can There Be a Private Language?", <u>Proceedings of the Aristotelian Society</u> Supplement 1954; P. Winch, <u>The Idea of a Social Science</u> (London, 1958), ch.1.

of learning a language all alone. It must exclude the possibility that Defoe's Robinson Crusoe, or the sole survivor of some future war, should understand the is/seems distinction, and conclude that therefore they could not use language. And the same should be true of each of us when we are alone; for then we have no better evidence before us than our own memories provide. If it be said that we, and Crusoe, can understand the is/seems distinction because we know we have been in situations where there were other people, and so know what it is like to have better evidence than our own memories provide, the answer is that at best we can only remember being in such situations and we have no available checks for this memory: so, by the argument, we should be wrong to assert that we really ever were in such situations (for we could not distinguish between its being the case that we were and its seeming to ourselves to be the case). This reduces to absurdity the strong form of the argument against private languages.[1] For the is/seems distinction to be intelligible to me it is not necessary that I should actually have independent evidence on whether I am right, it is enough that I should be able to get this. Not necessarily able in the sense of physically capable, but able in principle: it should not involve any

1. For this reduction see Ayer, "Can there be a Private Language?", Proceedings of the Aristotelian Society Supplement, 1954; Strawson, review of Wittgenstein's Philosophical Investigations, Mind 1954.

contradiction to suppose that I should have such evidence. The only motive for holding otherwise would be that one held the principle of verification in an impossibly strong form: a statement being meaningful if, and only if, I actually have (or perhaps: am physically capable of getting) evidence for or against it. This is a much stronger form of the verification principle than even the first version in Language, Truth and Logic.[1]

In its weaker and more usual form, the argument against private languages just requires that it be in principle possible for the speaker to get independent evidence on whether he is right, but it still holds that the only acceptable form of independent evidence is that provided by other people. So the claim is now that one could not have a language unless the items picked out by the terms of the language were such that there logically could be more than one person aware of them. For if the terms are to be intelligible, one must recognize some distinction between using them correctly and seeming to oneself to use them correctly. But then it must be possible to have better evidence on the subject than what seems to oneself to be the case. I cannot rely simply on my own memory how the term is used, because then

1. Ayer, Language, Truth and Logic (2nd ed., London, 1946), p.36.

I could not know what it was to correct a false memory-claim: there would be nothing else in the light of which I could correct it. No other memory-claim of mine could provide the required evidence: to check one memory against another would be like checking two copies of the same newspaper against one another.[1] But if I can appeal to the memories of other people, then I do have an independent source of information which I can use to check my own memory-claims, and so I can draw the distinction between what really is the case and what just seems to me to be the case.

I shall consider five objections to this argument. The first three are unsuccessful, the other two seem successful.

I. One might suggest that in this weaker form the argument against the possibility of a private language excludes nothing that anyone might want to maintain. ~~It is~~ It is contingent that pains and bodily sensations can be perceived only by one person; the only things that are logically unshareable are states like being aware or perceiving. I could in principle have the very same pain as you do, e.g. in a shared limb, but I could not in principle

1. Wittgenstein, Phil. Inv. sect. 265. Presumably the two copies are of different editions and one has no independent knowledge about which is a correction of the other.

have your awareness of your pain. But awareness cannot itself be perceived, and all that can be said about it can be said by referring to the subject and the object of the awareness and saying (something like) that the subject can describe the object without going on any evidence. So I am not missing anything if I cannot have a private language the terms of which refer to my own states of awareness.[1]

This objection assumes that the conclusion of the anti-private language argument is correct. Only if it is correct is it true that all that can be said about my awareness of pain can be said by speaking of the subject and of an object which is only contingently private. Otherwise the possibility remains that there is a logically private phenomenon, the pain-feeling, of which I am directly aware and which can have no existence apart from my awareness of it. There would be no criterion for the truth of the ascription of this feeling to oneself other than the occurrence of the feeling. In this case we could call the feeling a mode, rather than an object, of awareness, to indicate that it is not something different from one's awareness of it and that it shares the logical privacy of awareness; though if pains and other feelings succeeded one another pretty regularly we could introduce a sense to

1. Cf. Don Locke, Myself and Others (Oxford, 1968), ch.1.

speaking of even such logically private phenomena as continuing to exist when one was not conscious of them, and in that case it would be natural to call them logically private objects. So this is the possibility which the anti-private language argument seeks to exclude; in advance of the argument it does not look incoherent.

In saying this, the proponent of a private language is not saying that my declaration that I am in pain is incorrigible. To start with he could point out that "pain" is a word which belongs to a public language, so that my aspription of pain to myself might well be subject to correction by other people even though my ascription of S to myself was not, 'S' being a term in my private language. But he need not -- and should not -- hold that even my self-ascription of S ~~to myself was~~ is incorrigible. He can accept that for any meaningful utterance there must be a distinction between its being correctly used and its only seeming to the speaker to be correctly used, but say that one cannot be wrong in saying "I feel S" in all the ways one can in saying "There is an elephant here". The point may be put roughly by saying that in the elephant case both factual and verbal mistakes are possible, but in the case of S only verbal mistakes[1]; and one must remember that

1. For a consideration of the Austinian view that factual as well as verbal mistakes must always be possible see below, pp. 139ff.

verbal mistakes are a perfectly good sort of mistake and should not be dismissed as trivial -- we must have grounds on which to correct our misdescriptions, just as we must have grounds on which to correct our misapprehensions. '_S_' can be used to describe the way I feel, and I can misdescribe my feeling -- by a slip of the tongue, or by forgetting how the word is used. In the case of seeing the elephant also I can misdescribe (in which case, if e.g. there is no elephant there but only a box of matches, it is only my English you have to worry about), but here it may also be the case that I describe correctly what I experience (seem to see) and my experience fails to match what is there. It is this distinction that is ruled out in the case of logically private feelings: there is no pain there over and above my awareness of it.

Thus the first objection, that the anti-private language argument excludes nothing, fails: it excludes no possibility only if we assume that it has already been invoked to show that what we would otherwise have taken for a possibility is not really one.

II. The second objection is made explicitly by Don Locke.[1] This is that the argument is only partly successful because it shows only that the _sense_ of any term must be public; it may still be possible to use it to _refer_

1. _Op.cit._ ch.5.

to logically private items. — But how? It is in virtue of its sense that a term can be used referentially: I can use the word 'table' in a referring expression only because I know what it means. A demonstrative by itself cannot be used to make a reference unless it be understood from the context roughly what sort of thing is being indicated, and 'that x', where x is a term in the public language, only picks out the public phenomenon. Of course, we can speak about states of awareness in a public language; but so far as we can do this our terms are public, other people can correct us when we use them. The situation is complicated by the fact that in my idiolect a term like 'pain' may have both private and public criteria: if these sets of criteria were to cease to give the same results, as it is logically possible that they should do, I should then have to replace my concept of pain with two, a public-language concept and a private-language concept. If this is true of the idiolects of many people it may explain why they find Don Locke's suggestion, that a term with public sense can have private reference, attractive.

III. The next objection, put forward by Saunders and Henze[1], is also an attempt to weaken the argument's conclusion. They concede that there could be no private

1. The Private Language Problem (New York, c.1967), last ch.

language which served only to describe logically private items, but suggest that once given a public language it may be possible to introduce expressions for private phenomena. They accept Strawson's claim in Individuals[1] that if I am to have the concept of myself as a subject of experiences I must also have the concept of other minds, but suggest that given a public language in which the concepts of a person and of experiences of various sorts are already established private terms for special sorts of experience can be introduced.

But if the argument is valid at all it is valid generally; and Saunders and Henze do not show that it is not valid at all. The question it raises is: what independent checks must there be for my claim that I am using the word correctly? And if in any case these checks have to be public purely in order to be independent of what seems to me, this must be true in all cases. One could not reply "But I know from other cases, where there are public checks, what being wrong is like, so I can extend the idea to this case", for this case is (by the argument) disanalogous to the others in precisely the relevant respect.

IV. The objections considered so far leave the argument intact. The fourth and fifth objections do not.

1. p.99. On this argument see the section on self-consciousness, below.

The fourth is made by Bennett[1], and is that we do not need to invoke other people's opinions to provide the necessary check. "What people say is only a special case of what objects do"[2], and it is enough that experience should be such that I can make regularity-generalisations about it. In making these generalisations of course I rely on my own memory; but so I do in interpreting what other people say. Regularity-generalisations make it possible for me to bring a complex collection of data to bear on a recollection, confirming or disconfirming it by showing its consonance or dissonance with other recollections or with present states of affairs. If I have observed a regular concomitance of A's and B's, my memory of an A at t and my memory of a B at t will confirm one another.[3]

Against Bennett it has been argued[4] that this does not give us an independent check on whether the event A occurred. For the occurrence of B, and of any other regularly concomitant or sequent events, must either provide a criterion for the truth of the statement that A occurred or else only a symptom of its occurrence -- the connection must either be conceptual or contingent. If it is a criterion, A and B are

1. Kant's Analytic, p.212.
2. Ibid., p.213.
3. Ibid., p.208.
4. By J. Hopkins, in discussion.

no longer independent; if it is a symptom, it cannot give us the necessary evidence about A, for we still have no independently checkable criteria for statements about A.

This objection will not do; it turns on a confusion about criteria. One can regard the relation between a statement S and the statement of its criteria as being logical, i.e. an entailment relation; in which case one can draw the neat distinction between criteria and symptoms which the objection presupposes. Alternatively one can regard as criteria whatever it is that constitute the best grounds there (logically) could be for asserting S: one could not understand S unless one knew that S was often true when its criteria, in this second sense, obtained. Here the only logical relation between S and the statement of its criteria is a vague probabilistic one: "Necessarily, if c, then there is a good chance that S".[1] The occurrence of B, or my memory of it, can count as a criterion (in this latter sense) for statements about A, even though A and B are independent events -- and it will do so if, as on Bennett's account, it is taken to be a necessary condition of referring to A that a rough concomitance between A-type

1. Some people conflate the two senses of "criterion", by speaking of a "loose logical connection" and clutching at the brow. "p entails q *ceteris paribus*" is unnecessarily perverse if all it means is "Necessarily, if p then probably q". But of course there are many problems here.

events and B-type events be assumed. The adversary of private languages is himself committed to an exactly similar move at this point, in explaining what the relation is between past-tensed statements and the memory-claims of everyone. For it is surely absurd to say that some set of memory-claims entails a past-tensed statement, or vice versa: the past is not a logical construction out of people's memories. An event does not cease to have occurred when everyone has forgotten about it (or even when all the other traces of it have gone as well)[1]; and the position is not saved by saying that though no-one does make the right memory-claims someone might, since for every possible person who does remember it I can find a million other possible people who do not.[2] So the opponent of private language will have to say that memory-claims are criteria for past-tensed statements in the second sense only. One could not have the concept of the past unless memory-claims were often correct, but it is perfectly compatible with this that everyone's memories should be wrong about a particular matter.

1. Though N. Malcolm apparently thinks it does. See Knowledge and Certainty (Englewood Cliffs N.J., c. 1963), pp.198f.
2. Cf. the mistake of the phenomenalist, who wants to say both that statements about material objects are equivalent to sets of statements about sense-data and that every actual person could suffer a simultaneous hallucination.

It is thus hard to see that there has to be any difference in evidential status between memory-claims and observed concomitants. It as as legitimate to check our memory-claims by reference to observed regularities as it is to check them by reference to the memory-claims of other people. In some circumstances it may be less reliable; but one could imagine others in which it would be more so.

V. But why is it even necessary to appeal to regularities? What the appeal provides, in effect, is simply a device for making some of my memories provide evidence for others; perhaps the device could be done away with. For I can corroborate or disconfirm my memory-claims about A by producing other (sincere) memory-claims, perhaps stronger ones, concerning the occurrence or non-occurrence of that very event, A.[1] What is the relevant difference between doing this and appealing to the memory-claims of other people? It may be asked why I should accept one rather than another of my own conflicting memory-claims: well, I may have more apparent memories pointing one way than the other, or those pointing in one direction may be phenomenologically stronger than those pointing in the other, and then I just do tend to accept the one with more or stronger support rather than the other. There is no more ultimate

1. Cf. Ayer, Proc. Arist. Soc. Supplement, 1954.

justification for this. But again the proponent of private languages is in no worse a position than his adversary: why should I accept what everyone else says and reject my own apparent memory when it conflicts with that? This could be answered, of course, if the statement that most people have apparent memories of A entailed that A occurred, but then it doesn't entail it -- any more than does the statement that my own strongest and most considered apparent memory on the subject is of A occurring. If I can rely only on my own apparent memories my memory-claims will be anything but watertight, but we have no watertight way of finding out about the past in any case, even relying on complex general laws and the memory-claims of millions of other people. It may be that we are in practice unable to have logically private languages because our memories are too fallible; and whether we in fact use such languages I shall not discuss; but it is only contingent that memory-claims are not more often accurate than ours are.

Wittgenstein would however reply to IV and V that they miss his point completely. For his attack on private languages is an attack on private acts of naming. It is not just that my memory of the rule "S names sensations of this sort" is problematic: the rule itself cannot even be set up privately. For what could it be to set up such a rule?

> It might be said: if you have given yourself a private definition of a word, then you must inwardly undertake to use the word in such-and-such a way. And how do you undertake that? Is it to be assumed that you invent the technique of using the word; or that you found it ready made?[1]

If on a particular occasion I decide to use the term S to name sensations like the one I am having then, it is still completely indeterminate what other sensations this covers. The only way of determining this is to see what sensations I use it to cover: this is what shows what I count as similar here, and gives a sense to "like this". And this is how we proceed in the public language. If after a certain time everyone were to use the word 'blue' for light in that range of the spectrum we call 'green', and vice versa, this would not show that they were wrong or even that their use of the words had altered - except in the trivial sense that unless a corresponding change had occurred in the world the extensions of the terms would have changed: grass would now be blue, not green. They might say that things had changed colour, but that the words 'blue' and 'green' had not altered in sense: grass is now the very same colour that sky used to be. Because a lot of people are involved

1. Phil. Inv. § 262; cf. §§ 257, 258, 263, 268.

any of them individually may misuse the words and be corrected by the others: but the group can't: if everyone says grass is blue, then that is how the word is used, and grass is blue. With a private language this inability on the part of the group to misuse words becomes the inability of the individual, since he is the only speaker of the language: and there is thus no distinction between his misusing a private word and using it correctly.

If everyone in a tribe agrees that grass and all other things that were green have changed colour and become blue, whereas it seems to us that they have not, we can say that the meanings of the words "green" and "blue" for these people is further specified by this usage. We should translate their colour-words into our language as "grue" and "bleen" rather than as "green" and "blue".[1] If, on the other hand, the members of the tribe call grass (etc.) green at t_{-1}, and at t_{-2}, as in the previous case, call it blue, but this time agree with us that nothing has changed colour (thus holding now not only that grass is blue but also that it always was blue), we cannot say that the new usage specifies the meanings further: rather, it gives the words new, and different, meanings. We can still say that they are not being misused, since for a word to be

1. N. Goodman, Fact, Fiction and Forecast (New York, 2nd edition c. 1965), p.74.

used correctly is just for it to be used as everyone(or perhaps the upper middle class) uses it, and everyone now uses it in the new way. But this does not prevent us, or them, from seeing that the meaning has changed; and indeed it is in practice a very familiar phenomenon that words change their meanings. Now with the private language the same sort of thing may happen: the speaker may use S at one time in one way, at another time in another, and may recognise that he has changed its meaning. But this is all that is necessary in order for him to have the idea of a mistake about the meaning of S. Perhaps, indeed, we can say that in one sense he must always use it correctly, since for S to be used correctly is just for it to be used as he uses it. But for one thing it is not clear that it would be right to say this, and for another even if it were he could still make and correct mistakes. It may not be right to say it, because the purpose of the speaker of a private language would presumably be principally to communicate with himself at later times by keeping a diary or records of some sort. Because the primary purpose of speakers of a public language is to communicate with one another this purpose is effected provided they all use words in the same way at the same time, and how the words may have been used the previous day or the previous century is irrelevant; this is why a word is used correctly if it is used as everyone now uses it. Even so, if I am reading a nineteenth-century book in which an argument is called

specious and I understand this to mean that the argument is invalid, I have misunderstood, for 'specious' then meant 'plausible'. Similarly, if the speaker of the private language is to understand his diary he must remember what his words meant at the time that he wrote it. So we may perhaps say that in a private language, if its purposes are to be effected, for a word to be used correctly it must be used not as the speaker now uses it but as he originally decided to use it. And even if we do not say this we must recognize that mistakes about the use of words are possible, through misremembering. For just as I misunderstand the old book if I do not know how the words in it were used when it was written, the speaker of the private language will misunderstand his diary if he does not remember how the word S was used, and so the purpose of the diary is defeated. Further, much of the point for him of calling a present private phenomenon an S must be to assimilate it to other, past cases, and this purpose too is defeated.

Wittgenstein would say that this is incoherent. For it supposes that one can make a comparison between the present private phenomenon and the class of phenomena one decided to call S-phenomena, a comparison which is independent of one's inclination to call the present phenomenon S as well. If I am to remember that it was decided that S should be applied to a certain set of cases,

then, in order to determine whether the present phenomenon falls within that set and hence ought to be called an S, I must remember the set in some other way than just as "the set of phenomena I decided to call S's".

> How do I recognize that this is red? - "I see that it is this; and then I know that that is what this is called". This? - What?! What kind of answer to this question makes sense?[1]

As Wittgenstein's pponent I seem to be committed to saying that the meaning of a word is a mental idea or image: for only some such inner picture could enable me to remember the set in the required way.[2] But Wittgenstein argues repeatedly and forcefully against the suggestion that the meaning of a word is a mental image; for the image itself requires interpretation, and we would not say of anyone who could use words correctly and had no corresponding images that he did not know what his words meant.

The question what one is remembering when one remembers what one's words mean or used to mean arises for the public language also. Wittgenstein holds that in this case there is no need to say "One remembers that the word was assigned to this mental image", and indeed that it would be useless to do so: rather, we have to rely on how other people use

1. Phil. Inv. § 380.
2. Phil. Inv. §§ 377ff.; 56, 276.

the word.[1] But here we must keep in mind the distinction between remembering how the word was used - so that one can understand ancient documents etc. - and knowing how it is used nowadays. If Wittgenstein is saying that in order to know how a word (in a public language) is (correctly) used nowadays we have to rely on how other people use it, he is surely right. But if he is saying that reliance on other people is necessary if it is to be possible to speak of remembering correctly how the word was used, he seems wrong. If this is the point that one's own memories do not provide good enough checks on one another, the argument is inadequate, as we saw; but I think it is not this point, and Wittgenstein does want to provide here an answer to the question what we are remembering when we remember what a word (in a public language) meant. But no adequate answer is given. If I appeal to other people on the question of what a word meant, all they can do is to tell me "the word W was then used to mean V", or perhaps "the word W was then used for this". Here what they are doing is reminding me of a correlation between words; for even if they use the demonstrative 'this' I can only understand what is being pointed at because I can understand the speakers' language and know the linguistic content. (Only so can I have reason to guess e.g. that an object's shape

1. Phil. Inv. § 381, 224, etc.

is being indicated and not its colour or the average atomic weight of the substances it is made of. And if what is indicated is a property, only so can I know from this instance anything about how to apply the word in other instances). So the answer to the question what we are remembering when we remember what a word meant seems to be, we are remembering certain correlations between words. But now what importance does the appeal to other people have, unless it is simply being used to make the point, discussed above, that one's own memories cannot confirm or disconfirm one another? For the speaker of a private language, too, can remember that he decided to use S to mean a T that is R, where R, S and T are all private words; there is no reason why he should be supposed to have to check the word against an image, any more than the speaker of a public language does. If what he remembers is a series of verbal definitions, it is possible that he should systematically misapply (i.e. use in a new way) all his words while maintaining the relationships set up between them: but this would show itself, for unless there were a simultaneous and corresponding change in the world the sentences which formerly expressed contingent truths would be likely now generally to express contingent falsehoods, and vice versa.

Wittgenstein might argue that there is still a problem.

The speaker of the private language can only have reason to suppose that he is using his words in a new way, though keeping intact their definitional interrelations, if he can have reason not only to rely on his memories about the relevant contingent truths and falsehoods, but also to reject the hypothesis that the apparent changes in truth-value are not due to a real change in the world. How can he decide between these three alternatives: that the world has altered so that what was contingently true is now false and vice versa; that his memory is misleading him about these contingent matters; and that he has come to apply the words to different cases despite preserving their definitional interrelations? Even when all seems well there is apparently no ground for confidence that the word is being applied in the same old way, for there may have been a change in the world accompanied by a corresponding change in the speaker's memories.

> Always get rid of the idea of the private object in this way: assume that it constantly changes, but that you do not notice the change because your memory constantly deceives you.[1]

But the same difficulty arises for speakers of a public language: for they also might preserve verbal definitions but apply words as they would not have done before. And

1. Phil. Inv. p.207e.

if it is possible for them to discover that a word is being used differently, this must be because they not only trust their memories of contingent matters but assume that the world has not radically changed. Again there are the three alternatives, and apparently nothing for the speaker of the language to choose between them: that the world has altered so that what was contingently true is now false and vice versa, that everyone's memory is wrong about the relevant matters of fact, and that they have come to apply the words to different cases though retaining verbal definitions. It is true that if the change happens simultaneously to everyone they will be able to communicate with one another as well as ever, though not to understand records, so that their primary linguistic purposes will be served; whereas for the speaker of a private language the loss of the ability to understand records is the defeat of his (presumably) primary linguistic purposes. But this is not enough of a difference to serve Wittgenstein's turn. The speaker of a public language also has to be able to rely on the assumption that words are still being applied in the same old ways (or if not that he knows what the difference is) if he is to understand records.

One can make two answers to this.

I. The first answer is that if he is to build up any body of working knowledge about the world the speaker must assume that it is not constantly changing. It is a

presupposition of our being able to arrive at any cross-time generalisations at all about our experience, and so a fortiori of having the idea of something which exists unperceived, that we should assume (when there is no evidence to the contrary) that words are used in such a way that simple generalisations can be formed about things. - And this is what their being used in the same way amounts to: not just that their definitional interrelations should be preserved, but that by and large they should preserve the same contingent truths. For when presented with the three alternatives above, one has to assume that the world has not radically altered, and that our memory is reliable: the first assumption being in any case a condition of cross-time generalisations, and the second involving no special difficulty - it is part of the concept of memory that it should be treated as reliable prima facie. Both assumptions are of course defeasible. But once made, and unless defeated by further memories about the same case or further evidence about other related cases, they leave only the third alternative, that the word is being used differently. So we do, necessarily, rely on the contingent uses of words in determining whether they are being used in the same way or not.

If for the speakers of a public language it is integral to the idea of using a word in the same way at different

times that it should be used so as to preserve contingent truths as much as possible; if, that is to say, defeasible assumptions have to be made that people's memories of matters of fact are trustworthy and that the world is not constantly changing; then one should not be surprised if the speaker of a private language has to make the same assumptions. And the fact that he has to make them in no way suggests that there is anything unintelligible about the idea of a private language. The only difference is that whereas one might manage to communicate with someone else without relying on these assumptions, because the past uses of symbols could be discounted as irrelevant, it is difficult to see how private words could be used in this way because it is hard to see what linguistic purposes the speaker of a private language could have which did not involve communicating with himself across time, even if the time be very short. Even if he uses words to work out a problem or clarify his thoughts he is still relying on their meanings not shifting throughout the argument. But this difference is not very great: for of course every public communication-system does rely on words not shifting in meaning at least over very short periods of time, and if it did not one could never have any reason for supposing that one's audience would understand one's signals aright. Indeed it seems not unreasonable to refuse to any intentional communication-system which does not depend on symbols' having,

and retaining at least for a short time, standard meanings, the honorific title of a language.

II. The second answer does not commit one to the position that the speaker of a language has to assume any regularity at all in the world. It is also intuitively more satisfactory, since it does not involve the assumption that one could not retain one's colour language just as it is if all the green things in the world were suddenly to become blue and all the blue things green.[1] By the terms of the first answer, one would be driven in such a case to say that no change had occurred in the world, and if one were to call grass now blue this could only be a case of change in the meaning of the word 'blue'. Many followers of Wittgenstein are inclined to say this in any case; the first answer can be used against them. But it does appear to be the case that however radical a change occurred in the world I could still have grounds for thinking that a change in the world had occurred and not just a change in my, or everyone's, use of words. For though all grass be now blue I can still remember it as it used to be, green, and this is separate from my memory that the sentence "Grass is green" used to express a true proposition. One can

1. And if the correlations between colour and wavelength were to break down, so that things continued despite the change to reflect light of the same wavelength as before.

express this by saying that one has a memory-image of expanses of green grass as well as the memory that they could be described by the use of the word 'green'; and one can recognize that if one came across grass like that now one would call it green - it would be (on this supposition) the same colour that sky now is. Wittgenstein ought not to object to speaking of a memory-image in this sense. Indeed I ought perhaps not to call it an image, but I know no better word: I do not want to suggest that the image requires <u>interpretation</u>, because then difficulties would arise over the method of interpretation. It comes already interpreted: it looks in my mind's eye just as green grass would look were I to see some. And I am not saying that the image is the <u>meaning</u> of the word: only that if there were grass looking like that I should call it green, and that to remember what a word means involves, at least in some cases, remembering what sort of thing it applies to - being able to appeal to an indefinitely large set of such images.

If this makes sense there is no need to rely on any sort of regularity in the world. Only memories need be relied on - one's memory-image and one's memory of what would have been said about the case. Neither sort of memory-claim is incorrigible: as we have seen, any memory-claim can be confirmed or disconfirmed by other memory-claims.

The first answer also required reliance on memory; this second one just extends that reliance to cover the memory-images of the situations in which the expression could be applied. I do not wish to get further involved in a discussion here of the legitimacy of talking in this way about memory-images; to those who find it still unintelligible the first answer still remains, and they cannot criticize that as intuitively unsatisfactory because they will be unable to accept as intelligible any more intuitively satisfactory way of describing the situation.

So the attack on private language is unsuccessful. But it did show the need, in an intentional communication system, for a distinction between what is the case and what seems to the speaker to be. And if one accepts the first answer and not the second, one will hold that it also showed that if speakers are to rely on symbols' retaining their meanings, even for a very short time, they must assume the world has not very radically altered. Even so, this does not entitle them to assume that it will not alter radically in the future: for it might change so utterly that no interpretation of the words could be found which would preserve a significant proportion both of the interdefinitions and of the contingent truths. Nor does it show immediately that their experience must have that particular sort of regularity which would enable them to

apply the idea of an unperceived existent - they might only be able to make generalisations which did not support counterfactuals. We must now see if we can get from here to some proof that one must have the idea of something which can exist unperceived. A private language can accommodate terms for unperceived existents, for private feelings may occur with the sort of regularity which induces us to form counterfactual-supporting regularity-generalisations about them; but it has not yet been shown that there could not be a language with no terms for unperceived existents.

Unperceived Existence.

The usual argument that to have any experience we must have the idea of something which exists unperceived is that otherwise there would be no room for the distinction between what is the case and what seems to be the case; and unless one could draw this distinction one could not know what it was to correct a mistake.[1] This involves a confusion similar to Kant's between two senses in which a judgement can be "objective", (1) as having a truth-value and (2) as being about items the existence of which is independent of anyone's awareness of them.[2] Any judgement objective in sense (2) will be objective in sense (1), but it is not so clear that the connection runs the other way as well. The consideration that one must have the idea of correcting a mistake in order to have a language does not, I think, even show that one must be able to express in language judgements which are objective in sense (1): the making of true or false utterances has to do with the communication, recording, considering, etc., of

1. Strawson's argument (in Bounds of Sense pp.99-109), that esse and percipi must be distinguishable because there must be room for a distinction between experience itself and the subject to which it can be ascribed, is also grounded on this claim and involves the same confusion.
2. "Aware", of course, in the sense in which animals can be aware of things and not in the sense in which it is a necessary condition of my being aware of something that I can describe it.

information, and there might be an intentional communication-system which was not used for these purposes but in which expressions still had standard meanings. A language with an imperative mood but no indicative would be an example of this. Here the meaning of an utterance could be learnt by observing what action satisfied the utterer and made him angry; a speaker could make a mistake by using the utterance in a non-standard way, and correct it by another which would better serve his purpose. But in such a case it would be reasonable to say that although he cannot express judgements the speaker must nevertheless make them, in order e.g. to know what his words meant. A language with no mood but the imperative could not express knowledge; and if one is to have that sort of experience for which the ability to use language is essential, language may be presumed to function partly to describe experience and so must contain judgements having truth-values, objective in sense (1). This is not a substantive point but a partial definition of the relevant sense of "experience".

It might be argued that all judgements are objective in sense (2) as well as sense (1), because truth-values exist independently of anyone's awareness of them. But we do not have to think of truth-values as entities. There might easily be, and no doubt are, languages in which there are sentences which can be used to make statements with truth-values, but no substantival expressions translatable as "the truth-value of S." And even if we do treat them as entities this would not show

that we had to treat any other sort of thing as capable of unperceived existence: people like Kant and Strawson, who think that one must have the idea of something which can exist unperceived, have in mind not truth-values[1] but things like material objects. In the case of a language in which although truth-values were not treated as items but in which judgements having truth-values could be made, one might say that users of it had the idea of something which could exist unperceived even though they could not make judgements about it; but since this something would merely be a truth-value it would not be an item of the really interesting sort. I shall henceforth ignore this possibility of treating truth-values as items.

Again, it might be argued, in a Wittgensteinian way, that to correct a mistake I need better evidence on the subject than I had before me when I made the mistake, and this must come from other people or at least from observing objects which behave regularly. This argument I have just discussed; but in this context it may be repeated that even if Bennett is right in saying that we have to be able to appeal to regularity-generalisations it does not follow that we must have the idea of something existing when we are not aware of it. The generalisations are needed only

1. Or numbers, for that matter.

to support or disconfirm memory-claims, and for this need only apply to actual cases: one thinks one remembers an A without a B but the generalisation suggests that this is wrong, either there was a B or there wasn't an A. It isn't necessary that the generalisations support counterfactuals. Even if they do this still does not give us objectivity in sense (2): "If there had been a pain there would have been a tickle shortly afterwards" makes sense even if pains and tickles cannot occur without someone being aware of them. The ability to make regularity-generalisations which support counterfactuals is a necessary, but not sufficient, condition for having the idea of a thing which can exist unperceived.

Nevertheless, I think it must be granted that in one sense it is transcendentally necessary to have the idea of something which can exist when one is not aware of it. Using a language involves forming intentions, and sticking to them for short periods of time at least: this is what following a rule amounts to. In order to do this one must have some sort of idea of oneself as persisting through time. And if I think of myself as a thing, a subject of experiences, I must be prepared to admit the logical possibility that the world contains other similar things, subjects of experiences. These other subjects would be capable of existing independently of my awareness of them, for I do not think of my own existence as dependent on

anyone else's awareness of me. So I must be prepared to admit the logical possibility that certain things should exist when not perceived by me.

This establishes less than one might have hoped. It does not show that I must actually apply to the world the notion of an unperceived existent; my recognition of the bare possibility that there be other minds does not guarantee that I should actually come across any. Nor does it even guarantee that I should have the idea of an unperceived existent, in one perfectly good sense - perhaps the most natural sense - of "have an idea": the logical possibility that there should be other minds may just never occur to me, and in a world in which I never come across any it would be quite likely not to. That I should have to grant the possibility as soon as I thought about it does not ensure that I ever do think about it.

It might be thought that we could get further than this. One way to try would be to go beyond the bare requirement that an utterance be made for some purpose and start instead from a more fully-fledged Gricean theory of meaning.[1] But this will not help. If A's goal in uttering p is to induce a belief in B by means of B's recognition that his utterance was intended to produce this

1. Cf. Grice, opp.citt.

belief, then the sort of correction of mistakes we shall find will still be A's discovering that 'p' was not the appropriate utterance for this purpose and using 'q' instead. And this remains true whether or not the content of the belief A was trying to induce had anything to do with items which can exist unperceived. It is of course true, but does not really help, that Grice's theory of meaning is an account of intersubjective communication, and one could not have the idea of another person to be communicated with unless one had the idea of something which could exist when one was not perceiving it oneself. Because of the failure of the argument against private languages, it has not been shown that languages must be useable for intersubjective communication. The Gricean account of meaning would of course require modification for a private language, but there is no reason to suppose it could not be given: intentions to induce belief in others might be replaced by intentions to induce belief in oneself at a later time, intentions to get oneself to do things, etc. (Incidentally, even a language which was used for intersubjective communication might not possess the means for expressing judgements about language-users - perhaps the languages of some animals are like this; though in this case the behaviour of the language-users would show that they had, in one sense, the idea of another being to be communicated with, and so of an

unperceived existent.[1]) Nor can we get anywhere by appealing to the fact that the words of a language have standard meanings: for unless the argument against private languages is valid there is no reason why semantic conventions should have to be public. There will be a difference between A's finding p unsatisfactory because it is not the appropiate utterance and his finding it unsatisfactory simply because it fails to subserve his purpose; in the former case his correction is based on his memory of what the semantic convention_s are; but we have found no reason to suppose that he cannot remember these without being able to appeal to other people.

There is a different line of argument that might be used to show that one must apply to the world the idea of items which can exist unperceived. This is based on the claim that there are no feelings, or items of any other sort, which are self-intimating. S is self-intimating if, and only if, "S occurs" entails "Someone is aware that S occurs": so if there is nothing that is self-intimating, there is nothing that cannot exist independent of people's awareness of it. Here, as always in discussing sense (2) of "objective", "aware" is to be taken in that minimal sense in which animals can be aware of things, and not in that in which "A is aware

1. Whether they can be called self-conscious is another matter. See below, pp. 143ff.

that p" entails that A is prepared to state that p. So if pain is self-intimating there is no such thing as unconscious pain: if there is a pain, someone is aware of it, though he may misdescribe it (come out with the wrong word) or even be unable to describe it at all because of a shortage of vocabulary. The claim that pain is self-intimating (in this sense) is not that if there is a pain someone must know that there is, in the sense that he is prepared to come out with some statement to that effect; and when people attack the thesis that pains, etc., are self-intimating it is generally this they are attacking. There is always room for verbal mistakes; the claim I am interested in amounts to the claim that there must always be room for factual mistakes as well.

The suggestion, which is due to Austin[1], is that there may always be features of an experience which we do not notice, or which we get wrong at the time. It is not just that we can misdescribe what we are aware of, though we can do that; the sort of mistake that can arise through failing to match the words with what one is aware of is purely verbal. For factual mistakes to be possible as well, the

1. This may well be unfair on Austin, for whom both the sorts of error involved seem concerned with matching the words to the experience as one was aware of it. But the view here ascribed to him is certainly one strand in what he wants to say, and I have heard it defended as Austin's view. See "Other Minds", in his Philosophical Papers (Oxford, 1961), pp.60f.; and Sense and Sensibilia (Oxford, 1962), pp.112f.

sensation or feeling must differ from one's awareness of it in such a way as to allow a mismatch of the two. Sensations and feelings must then be objects, not modes, of awareness. Now there might, as I have said, be grounds in the orderliness of inner experience for speaking of sensations or feelings existing when one was not aware of them, as one might speak of a headache persisting while one was asleep, but it has not been shown that it is transcendentally necessary that there should be such grounds. Austin appeals to evidence of a slightly different sort: that having told myself at the time, seriously and without misusing the words, that I was seeing something heliotrope, I may later simply by reflection on the nature of the experience come to recognise that it was really magenta. Or I might "take in" only certain features of a complex pattern - others I might be in no ordinary sense aware of at the time, yet later, again by reflection on the nature of the experience, I might come to recognise that it had them.

Now there are two possibilities here. One is that at the time I simply failed to describe my experience adequately even to myself; but then I could still have been aware, in the sense we are interested in, of having it, and my mistake is verbal after all - a failure to match the words to the content of my awareness. If this alternative is correct certain counterfactuals about my discriminatory responses will be true (I could have told where the "heliotrope" patch shaded into one which really was heliotrope but not where it

shaded into one which was magenta, for example). If, on the other hand, these counterfactuals are not true of me, it cannot be said that I was aware (at the time at least) of having the experience, and the sort of cases that Austin cites would provide grounds for introducing the idea of someone's having an experience without being aware of it. But as grounds they would be on a par with the regularity of experience, which can induce us to speak of pains continuing when unattended to; it has not been shown to be transcendentally necessary that such cases should occur (which is what would force us to introduce objectivity in sense (2)), nor even that they do occur. It may be true that we can remember features of experiences which we did not notice at the time, but it is not clear that this can't be treated as a case of inadequate description of experiences of which we were aware. - Even if it can't, and if it were transcendentally necessary that such cases should occur, the sort of entities we should then have to talk about would be capable of existing when no one was aware of them, but not, of course, of existing unfelt or unexperienced: there must be some sense in which I took in an impression of magenta even though I was not aware of it, and the pain which goes on in my head when I am asleep (if we allow that way of talking) is still my pain. So that even if the argument succeeded it would not establish a need to refer to items objective in the fullest sense, capable of existing without

being experienced by anyone.

One last and rather desperate attempt to establish something like this would be to say that those statements which seem not to express judgements in sense (2), i.e. self-ascriptions of feelings etc., do not express judgements in sense (1) either - they lack truth-value and so are not really statements at all, but merely Äusserungen. "I am in pain" is not a statement but a piece of pain-behaviour. - This view has little to recommend it; the strongest motive for holding it is the belief that "I am in pain" is incorrigible, and when we recognise that it is not - because I can describe to myself wrongly the experience that I have - its plausibility largely evaporates. "I am in pain" does constitute a sort of pain-behaviour, but that does not justify us in concluding that these words function merely as a learned response, like "Ouch!" They can be used quite coolly to communicate a report about one's inner state, and when so used they are not a sort of pretence or faking, as "Ouch!" would be if one said it to convey a belief and not because it was elicited by the stimulus. They satisfy Gricean conditions for meaningfulness.

We have failed, then, to find a way of deducing, from the consideration that language is a rule-governed, purposive activity using words with standard meanings, the conclusion that a language must involve judgements which are objective in sense (2), or that its speakers must at least apply to

their experience, even if they cannot express it in language, the idea of something which can exist unperceived (as animals signalling to one another perhaps do). We did manage to establish the very much weaker conclusion that they must have this idea, in the somewhat minimal sense that they must be prepared to admit, should they ever think about it, the logical possibility that there should be such things. But this is too weak to support the Strawsonian argument that we must make counterfactual-supporting regularity-generalisations about our experience; that depended on the claim that we must think of our experience as actually containing items capable of existing unperceived. It may be possible to establish these things in some other way; all I am saying is that the~~y~~ arguments we have considered are not sufficient.

A New Starting-Point: Self-Consciousness.

This lengthy discussion arose because I wanted to consider whether transcendental arguments ever succeeded in reaching non-trivial conclusions. We have found that transcendental arguments to the effect that such-and-such is a necessary condition of having a language, where language is considered simply as a rule-governed, purposive activity in which words are used in standard ways, are often rather disappointing. Can we perhaps, by building a little more into the notion of language cr experience from which we start,

arrive at more substantive results without already having built them into our enlarged notion of experience?

The natural enlargement to start with is to consider the language or experience of a self-conscious being. In doing this one is, perhaps, siding with Strawson and Bennett's Kant against the real Kant. The move needs to be clarified. We have just been examining the conditions for language, or for experience so far as he can describe it, of a being who can act purposively; and in a sense any such being might be called self-conscious, since he can carry out his purposes without getting confused about whether they are his purposes or not, just as in a sense he can be said to have the notion of time because his purposes take time to execute and (typically) he sticks to them throughout. But it has not been shown that such a being must be self-conscious in the fuller sense of having any linguistic means of referring to himself[1]; or indeed that he must have the idea of a time-order in the sense of being able to express in his language tensed judgements or judgements about dateable events. His awareness of himself and of time may consist simply in his ability to carry out his purposes, and accordingly to put himself right if he makes an unsuccessful attempt by trying again in the proper way. Still, we might agree with Strawson[2]

1. Cf. Strawson, Individuals, pp.81f.
2. Bounds of Sense, p.107.

that the "essential core of personal self-consciousness" is already present whenever one can make the distinction between experience itself and the subject of the experience, in the sense of being able to distinguish between what is the case and what just seems to be the case; and any being with a language must be able to do this.

The case of a being who could not only exhibit purposive behaviour in his use of symbols but also used these symbols to communicate with other beings, though still without any linguistic means of referring either to himself or to these other beings, would be a little more complex. He still could not tell us, or himself, that he was uttering a particular noise with the intention of inducing a certain propositional attitude in his hearer; but we, as scientists examining his behaviour, would find that it could only be explained or understood if we took it that this was what he was doing. He must think of himself as one among many, making use of the notion of another mind and not merely prepared to admit the logical possibility of there being such things if the question ever came up. Perhaps the signalling behaviour of some animals can be thought of in this way; but we should still hesitate to say that these animals were self-conscious; and this seems to be because they have no linguistic means of referring to themselves. Similarly, though they are aware of the passage of time, their "languages" may contain nothing analogous to dates and tenses.[1]

1. See Bennett on bees in Rationality (London, 1964).

When I speak of a language-user as self-conscious, without qualification, I shall mean that he has the linguistic means at his disposal to refer to himself as a continuing subject of experiences. Animals may have some idea of themselves as continuing subjects, and exhibit it in their purposive behaviour; but human beings have the inestimable advantage of being able to talk about themselves.

That I can talk about myself does not automatically mean that I have anything to say about other people; I may never have met any. I must recognise the logical possibility that there are minds resembling mine but not identical with it, for the recognition of this possibility is implicit in my concept of myself as an item - an item which is at least potentially one among others. It cannot be built into my idea of myself that it is logically impossible that there should exist other similar subjects of experiences. To refer to an item one must have some idea, however inchoate, of how to identify it and individuate it from other items: I need more than the demonstrative "this" if I am to refer to myself, for "this" could pick out a present experience; I must have some idea of the sort of thing I am - and for this I require the predicate "person", or at least "continuing subject", which as a predicate cannot have uniqueness of application logically built into it.[1] However, this does

1. Strawson, Individuals, p.99.

not guarantee that I know how to recognise other people if I come across them: I do not have to know what the criteria are for statements about other minds just in order to know that the world might contain people similar to me. One might live in the world of men, communing privately with oneself, admitting that there might be other subjects of experience but not realizing that there were any - either because one had formulated no criteria at all for recognising other people, or because the criteria one had formulated insisted on too close an analogy between the other and oneself. An adequate defence of this would take us too far afield here; but at least the view that our knowledge of other minds is arrived at by analogy becomes reasonable once one accepts the possibility of a private language.

If there are other people, it involves no contradiction to suppose that they go on existing when I am not perceiving them. For it is not a feature of my idea of myself that I can only exist so long as somebody else is perceiving me. My language need not contain the term "person" (or "continuing subject") just because it contains the term "I", but "person" could easily be introduced - it would make explicit the character of my reference to myself as <u>this</u> person. And as soon as it does contain the term "person", it contains a term which can apply to items capable of existing when not perceived by me, even though possibly I may not know how to

apply it to them. If in referring to myself I must be prepared to admit the possibility of items which can exist independently of me, I must also be prepared to admit the possibility of experience being sufficiently regular to warrant my making counterfactual-supporting regularity-generalisations.

But this, of course, really gets us no further than we got in the last section; it merely reproduces the argument at a different level. It was there pointed out that if one has any idea of oneself at all one must be prepared to recognise, should the question occur to one, that there might be other people; though this might be at a pre-linguistic level, one's awareness of oneself and recognition of others being expressed only in one's behaviour. We have now added that if one can talk about oneself one possesses, or could easily introduce, the linguistic means of talking about other people - at least for expressing the possibility that there should be other people, which we can perhaps do without having definite criteria for recognising them. It is still perfectly possible that there should in fact be no other people than me, so that it hasn't been shown that one must apply to experience one's idea of something capable of existing when one is not perceiving it;what is more, one still does not have to have this idea, in the sense that although one is bound to admit the possibility of such things if the question occurs to one, the question may

simply never occur to one. Consequently we have got no further with the justification of the claim that counter-factual-supporting regularity-generalisations must apply; it must be logically possible that they should, but then we knew that anyway.

I doubt whether our new starting-point is much more fertile on other issues. The self-conscious man must be able to refer to past events, since he thinks of himself as continuing through time, and he can, if he pleases to, refer to his own mistakes and recognise them as his. The unself-conscious language-user must of course know what it is to make mistakes, but he can't give linguistic expression to the idea: he does various things to bring about his goal and if they don't succeed, or if he remembers semantic rules which tell him he has done the wrong things, he does other things. And even if he can refer to past events he is still unable to express the idea of a mistake: for that one has to be able to distinguish between past actions of one's own and other past events. (If the past action, or utterance, was not his but somebody else's, that person's purposes might not have been the same as his, so that it might not have been a mistake at all). The self-conscious man can, in fact, if he feels inclined, have a stab at semantics, whereas the unself-conscious language-user cannot. But this doesn't really get us anywhere, since it is built into the idea of self-conscious experience that

one thinks of oneself as continuing through time.[1]

But whether or not an examination of the conditions for experience which is self-conscious (in the sense defined) yields interesting conclusions in other areas, it doesn't help establish that counterfactual-supporting regularity-generalisations must apply. This is a pity, for it would be an interesting conclusion to establish - it is not immediately obvious that they must. It is obvious that experience must exhibit some sort of order for a finite mind to get any sort of grasp on it at all; in a world in which I could see no similarities at all between any two events or things I should be completely lost. But I might perhaps be able to experience a world in which, although I could apply certain simple predicates over and over again, there were no regularities discernible in the flux. I could be little more than a spectator in such a world as this, for though I might be able directly to initiate certain changes in it these would have no predictable consequences; but such a life might not be quite impossible. And it is much more easy to imagine that

1. It would be interesting if it could be shown that such a language must contain tenses, i.e. that it could not get by with tenseless verbs, dates and relations like "earlier than" and "later than". This would take us too far afield here, so I shall only say that I know no successful attempt to make out a case for this. To show that tenses are dispensable it is not necessary to prove that for every tensed statement a logically equivalent tenseless statement can be found, as R.M. Gale seems to think. (The Language of Time (London, 1968), ch.4.)

I experience a world in which there are regularities, but in which I do not take any of my regularity-generalisations to support counterfactuals. We have at least shown that my ~~reason for this cannot be that I think the move beyond actual~~ reason for this cannot be that I think the move beyond actual instances to be in principle illegitimate, because thinking of myself as a person commits me to the possibility of there being items which exist when I am not aware of them. So it must be that the regularities in my experience are not sufficient, or are not of the right kind, to induce me to treat them as warranting counterfactuals. But this seems to be a possibility: it would be worth while to rule it out by showing that counterfactual-supporting regularity-generalisations must apply in any experience.

It is clear from the preceding discussion what amendment of our starting-point will enable us to get the desired conclusion out. Starting from self-conscious experience will not do; but if we build in publicity all will be well. It is a necessary condition of having an intersubjective communication-system that regularity-generalisations of the required type apply to one's experience, for one will have to find room in experience to apply the concept of another mind. And it is a necessary condition of having an inter-subjective communication-system in which one can (not only refer to oneself but) refer to one's interlocutor that one be able to express these generalisations, for only on the

basis of them can the interlocutor be supposed to continue to exist when one turns one's back and puts in one's earplugs. It is worth noticing that this is not simply because if one can apply the concept of another mind there must be occasions when one has to admit that something is continuing to exist although one isn't perceiving it: for I might, I suppose, communicate with other people who (as it happened) only existed so long as I perceived them; I might have a wife whom I never let out of my sight and who never let me out of hers, and there might be nobody else in the world. But I should then have to admit, not just the rather abstract possibility that there might be other people capable of existing independently, but that <u>this</u> other person (my wife) might continue to exist when I wasn't watching, though in fact I do always watch her. And this admission is a counterfactual; it is grounded on the regularities which lead me to call her a person.

So by asking about conditions for intersubjective communication, or experience of a world which is actually (not just possibly) public, we can get the conclusion we want. Unfortunately it now looks a lot more trivial than it would have been as a condition of having any purposive linguistic system at all. But perhaps its apparent triviality is partly due to our having become familiar with the arguments: perhaps it is one of those things which, obvious as soon as stated, yet require stating.

And the same goes (rather more strongly) for the conclusion reached earlier that any experience must take place in time, because it must have room for the correction of mistakes: this again is obvious when pointed out, but Kant denied it (not to speak of numerous theologians).

It might be reasonable to say that only a being who can refer to himself and to others has a language, or has experience, or has knowledge, in the senses we are interested in. This does not do too much violence to the ordinary concepts of language, experience, and knowledge; and it would enable us to present the counterfactual-supporting regularity generalisations as transcendentally necessary in any language, experience, or knowledge, simpliciter. A trivial circle arises if one treats, as I have done, the ability to refer to oneself and others as a linguistic ability: this circle is easily avoided - one could call a purposive communication system in which signs have standard meanings a symbolic system, instead of a language, and then describe this activity as symbolic, defining a language as a symbolic system which allows one to refer symbolically to oneself and to other people.

On the other hand, it might be felt that I have been unduly restrictive in my examination of particular arguments, since I have been concerned with the conditions for having a language or at least for having a purposive symbolic

system, and with the conditions for experience and knowledge only so far as these are taken to be the experience and knowledge of the language-user, or the user of symbols. But it has, I think, become clear what can be said about the knowledge and experience of a non-language-user, so far as he can be said to have these. And not very much can be said. If he is a purposive agent, even though none of his purposes include communicating, recording, rehearsing to himself, etc., then we can reach the same conclusions about his experience as we did for the user of a symbolic system who was less than fully self-conscious in the sense defined, i.e. who did not have the symbolic means of talking about himself; for the ability to make and correct mistakes is essential to any sort of purposive action. And if he is not a purposive agent, I doubt whether it is appropriate to say that he has knowledge or experience at all.

"Knowledge" and "experience" are not really distinct starting-points for transcendental arguments. There can be no knowledge without experience: even if one's whole life were spent doing mathematics one would still have to know what it was like to make a mistake. Nor experience without knowledge: if I have an experience I must, at least typically[1], know that I have it (though perhaps

1. The pain which continues unattended to, and Austin's experiences with magenta patches, would be exceptions; such cases have to be exceptional, since it is fully conscious experiences that provide the evidence for saying they occur.

under the wrong description). Unless, of course, "know" is used in such a way that I know that p only if I am aware that my awareness of p passes certain epistemological tests - and it is not this sense that people (including Kant) have in mind when they present transcendental arguments as about conditions of knowledge.

It may be asked why I have not considered the transcendental arguments which Kant uses in the Analytic of Principles. There are two reasons. In the first place, I have hitherto been considering those very general transcendental arguments which purport to set out necessary conditions for any knowledge or experience at all - in some reasonably acceptable general sense of "knowledge" and "experience". The only important argument of Kant's which purports to do this is the transcendental deduction of the categories[1], and I have discussed that. The arguments of the Analytic of Principles are concerned with a specific sort of knowledge or experience, that which beings with spatio-temporal intuition can have; Kant makes it quite clear that he regards this as a limitation of concern, which however retains its interest because our intuitions are spatio-temporal. The dividing line is ultimately arbitrary between giving content to the notions of knowledge and experience by building various

1. There is also the transcendental deduction of space and time, which he doesn't say much about. I shall examine it below, pp.118f.

things into them, as I have been doing, and limiting one's concern to knowledge/experience of a particular sort; but it is none the less natural to make some distinction, as Kant does, between conditions for any experience at all and conditions for experience of a restricted kind. In the second place, however, I not only have not considered these arguments, I shall not consider them in any detail: because I do not think they are among the most plausible of transcendental arguments. My concern is with Kant's method, not with his conclusions (so far as these can be separated); I do not think that a detailed study of the arguments of the Principles would throw any further light on the method of transcendental argument, and their manifold shortcomings have been well exposed by commentators.

Conclusion

It is difficult not to feel disappointed at the rather meagre results of the transcendental arguments discussed above, but this may be due to our having been led, mainly by Kant but also by many others, to expect results of a more exciting character - e.g. those about causality or the impossibility of a private language (symbolic system). It is not clear that these more exciting conclusions are false, it is just not clear how they are to be demonstrated. In most of the transcendental arguments that have been put forward there is

something valid, the conclusions are just generally too strong: thus we need some idea of a regularity, but not of a cause so far as has been shown. But I do not feel inclined to call the less exciting conclusions we have accepted entirely trivial. For it has sometimes been thought that a self-conscious being could refuse to admit the possibility of there being other minds; not everyone has recognized the need for regularity-generalisations which support counterfactuals, even in a public language; truths of logic have been called conventional, which has sometimes been taken to imply that there could be language-users who did not observe them; religious philosophers have often thought that God and other spirits could be outside time and still experience things.

Transcendental arguments are often disappointing, too, because we expect them to provide conclusions about what the world must be like, when they only tell us how the subject must see the world; which doesn't impose much limitation on the world itself. They frequently tell us only that the subject must have certain conceptual capacities, and not that experience must actually contain items on which these capacities can be exercised. To have the idea of myself as a subject of experiences I must admit that there may be other subjects, not that there are others; to have the idea of an object independent of me I must recognize the possibility of counterfactual-supporting regularities, but my experience may just not be regular enough for me to say I am aware of any objects. Kant, Strawson, and most Kantians

generally expect transcendental arguments to show something about what experience - the world as we know it - must be like; Bennett makes his expectation quite explicit.[1] Of course they do sometimes show things about what experience must be like; as Kant would say, they can be constitutive of experience. A speaker must not only have, he must also be able sometimes to apply, the notion of his own past action if he is ever to have the right to be sure about what he says; otherwise he might be aware that a distinction could be drawn between being right and seeming it, but be quite unable to draw it in practice, and so could never have any reason for believing that he actually was right and did not merely seem to himself to be so. But in a good many cases transcendental arguments only show that one must have certain concepts and not that one must be able to apply them; in those cases one might say that what was set up was the quite legitimate subject of transcendental psychology. It need not, of course, be considered a faculty psychology; nor need it create a world of appearances.[2]

The arguments we have considered are of the general sort, like the transcendental deduction of the categories: they set out conditions for the experience of any sort of being, or (as Kant would say) at least for any without

1. Kant's Analytic, p.101.
2. See above, pp. 56ff. I have profited considerably from discussions with N.J.H. Dent on the subject of this paragraph.

intellectual intuition.[1] One may wonder whether more exciting conclusions may be obtainable by transcendental arguments of the other sort, to do with conditions for (specifically) human experience. It would be useful if, for example, we could obtain from such an argument some justification for talking about material objects rather than sets of sense-data, other than that of economy. This would provide some sort of refutation of scepticism about the external world; and the character of the relation between sense-datum statements and material-object statements is obscure - it would be helpful if mysterious leaps like this could in general be dealth with by saying that statements of the latter category were required in order to get a language going.[2] Moreover it would give us some justification for calling material objects "basic particulars".[3] Any such argument would differ from the ones above, not formally, but in the more complete specification of the kind of knowledge or experience concerned. It is this sort of argument that Kant uses in the Analytic of Principles.

1. Though why Kant thinks intellectus intuitus has experience at all is not entirely clear. For it to "know" anything is for it to create it; there is nothing independent of it, and there is no room for it to make mistakes.
2. Cf. Anscombe, "On Brute Facts", Analysis, 1957-8.
3. Cf. Strawson, Individuals, ch.1.

E. Schematism

It is clear that Kant's chapter on the schematism of the pure concepts of the understanding has a pretty central place in KdrV; it has been taken by some recent German commentators to be the most important section of all. Unfortunately its importance is the only thing that is clear about it. Kant professes to be showing that an activity called schematism must be carried out by the imagination in order that the pure categories should have application to experience. In the deduction, imagination has already appeared as the faculty responsible for the synthesis of sense-manifolds.[1] Its role was to bring the manifold of intuition to the unity of apperception - it subsumes sense-experience under concepts by making it conscious. But it is unclear what further question remains to be answered by the chapter on the Schematism once it has been shown that the categories must necessarily apply throughout experience; it is unclear even what place it could have in Kant's architectonic, however artificial that may be.

The imagina tion is supposed to mediate between understanding and intuition, and Kant's argument for the need to schematise the categories is based on this. He says that the pure concept of e.g. causality shares nothing with any given intuition, so that to apply it to intuition

1. A 123f.; B 151.

we require some third thing,

> was einerseits mit der Kategorie, andererseits mit der Erscheinung in Gleichartigkeit stehen muss, und die Anwendung der ersteren auf die letzte möglich macht. (A 138/B 177.)

This turns out to be the transcendental schema, "ein Produkt der Einbildungskraft".[1] He has a perfectly intelligible point here, though it is not much illuminated by the example he gives:

> So hat der empirische Begriff eines Tellers mit dem reinen geometrischen eines Zirkels Gleichartigkeit, indem die Rundung, die in dem letzteren gedacht wird, sich im ersteren anschauen läßt. (A 137/B 176).[2]

The point must be simply that if a concept is to be applicable to sensible experience at all we must have some way of determining in what sort of experiential situation it is to be applied. The geometrical example seems to us a bad one, just because we are used to thinking of geometrical concepts without worrying about the empirical situations in which we could apply them; we have come to think of geometry as a branch of pure mathematics not directly concerned with practical applications. For Kant, of course, this was not the case.

The proposition that if a concept is to be used in ordering our experiential data there must be criteria for

1. A 140/B 179.
2. Reading, with Vaihinger, 'ersteren' for 'letzteren' and vice versa. Cf. also the letter to Herz of 21st Feb. 1772 (Ak. X: 124f.).

its application is, of course, straightforwardly analytic and obviously true.[1] The obscurity in Kant's account is partly due to his speaking as though this showed that there has to be a special faculty of imagination; for the claim about imagination should amount to no more than that trivial proposition. I doubt if Kant was really seriously misled about this; his excision of most references to imagination in the deduction in B suggests he had used it as a heuristic device which had turned out to cause nothing but misunderstanding. It has of course continued to do so, reaching a high point in the work of Heidegger, who sees the doctrine of the transcendental faculty of imagination, as supposedly set out in the chapter on Schematism, as the central theme of the Critique of Pure Reason.[2]

The Schematism according to Kant involves the notion of time. This, he says, is because it is only time that has something in common both with the pure concept and with experience - for every sense-manifold is temporal[3] and yet

1. Cf. G.J. Warnock, "Concepts and Schematism", Analysis, 1948-9; and G. Bird, Kant's Theory of Knowledge (London, 1962) p.63.
2. M. Heidegger, Kant und das Problem der Metaphysik. 2nd ed. Frankfurt a.M., 1951.
3. In this respect time differs from space, inner intuitions (of one's own mental states) being temporal but not spatial (A 34/B 50).

time is a priori like the categories. This prompts the question why the categories could not be said straight off to have something in common with experience, for we are supposed to have at least as good a proof that experience is ordered categorically as that it is ordered temporally. Still, leaving that aside: since Kant does not recognise that every percipient without intellectual intuition has to perceive the world as temporally ordered, he presumably holds that a being with another sort of intuition might schematize the categories in a different way. The deduction is supposed to have shown that it must be the same categories that he applies to his experience by thus schematizing differently.[1]

If the categories are activities of the mind in synthesis, they must have something to work on, to synthesise; but it is difficult to see how this could have anything in common with, or could be in any way homogeneous (gleichartig) with, the synthetic activity which is applied to it. To make the argument for schematism plausible we have to think of the categories not just as sorts of activity but as concepts. We have already seen that this interpretation makes less good sense of the Deduction (unless to have a pure concept of the understanding is simply to be able to carry out the relevant sort of synthesis).[2] But if the categories are concepts in the

1. B 145.
2. Sect. B above.

same way that our concepts of a horse or of iron are, except that they are a priori concepts, like mathematical ones, we can perhaps in some measure understand the argument that they must be schematized if they are to become applicable. The process of schematizing will be something like giving an interpretation, in experientially applicable terms, of an abstract logical system. The pure categories, in their unschematized form, may not lack sense any more than the well-formed formulae of an uninterpreted calculus lack sense. The comparison is not altogether adequate, though, for wheras it is relatively easy to see how the formulae of the system get their sense - through their role in it - it is not easy to see what the analogue of this is for Kant's categories. Unless we take the categories as in effect adumbrating a whole (Aristotelian) logical system; but in that case the system is already interpreted - we know that the 'not', the 'if....then', etc., are the same notions as we familiarly use in all our discourse, and then it has once again become hard to see what special problem there could be about applying categories to experience - why schematism (and schematism in time, at that) should be required. So if the categories are concepts, and not just sorts of synthetic ability, it is not easy to see quite what sense can be given to the idea of a pure, unschematized category. One could say "Substance is substance", "Substance is not

unity", etc., but here "substance" and "unity" can mean no more than 'A' and 'B that is not A'. This sort of thing would not amount to giving the categories sense through implicit definition.[1] It is not just <u>empirical</u> content they require schematism in order to gain, but any content at all. So if Kant wants to retain an intelligible distinction between pure and schematized categories he should restrict himself to the first sense, in which categories are taken to be sorts of mental activity in synthesis, and abandon the argument for the schematism of the categories.[2] He could retain the bulk of the chapter on the Principles, as showing the need for each category in turn; and I think, and shall argue, that a

1. Cf. A 147/B 186:

> In der Tat bleibt den reinen Verstandesbegriffen allerdings, auch nach Absonderung aller sinnlichen Bedingung, eine, aber nur logische Bedeutung der blossen Einheit der Vorstellungen, denen aber kein Gegenstand, mithin auch keine Bedeutung gegeben wird, die einen Begriff vom Objekt abgeben könnte. So würde z.B. Substanz, wenn man die sinnliche Bestimmung der Beharrlichkeit wegliesse, nichts weiter als ein Etwas bedeuten, das als Subjekt (ohne ein Prädikat von etwas anderem zu sein) gedacht werden kann.

   But the envisaged account of "Substance" treats it as a term in a logic or universal grammar which is not altogether uninterpreted but simply abstracts from content given in experience. And there is no <u>problem</u> about applying such notions to experience.

2. And thus we find him at A 147/B 187 referring to the pure categories as "nur Funktionen des Verstandes zu Begriffen" which "stellen aber keinen Gegenstand vor".

different and in any case more useful twist could be given to the idea of a schematism of the categories.

As we have seen, the general form of a transcendental argument is: If there is to be experience of sort s, then it must be the case that p; but, clearly there is experience of sort s; therefore, it must be the case that p. And if we are going to use such arguments to any purpose in philosophy a lot will depend on what we take as our starting-point, what s is to be. Ideally we should make our assumptions as small as possible, and start, as I think Descartes takes himself to be doing, with propositions the denial of which would involve the admission that one was talking nonsense; the truth of these propositions would then be necessary for any (intelligible) experience whatever.[1] Kant, as we have seen, regards the transcendental deduction of the categories as concerned with the conditions of experience for any being whose intuition is sensuous, whether spatio-temporal or not. And this is a not unreasonable starting-point, for although he would have to concede that it could intelligibly

1. This will be so if no possible denial of these propositions is intelligible. So this case is different from those considered by Strawson in part I of Individuals where the sceptical denials are unintelligible because they presuppose the "conceptual scheme" they explicitly attack. Strawson admits that someone with another "conceptual scheme" could make these denials without unintelligibility: in fact this is how Revisionary Metaphysics proceeds. (Individuals, pp.78, 109).

be denied (by intellectus intuitus) that such beings can have experience, this denial would not be very interesting, at any rate to us.

But though that is the starting-point of the Analytic it is clearly not that of the Aesthetic, where it is not left open whether space and time are to be the forms of sensible intuition. If it started from the same point as the Analytic the most the Aesthetic could hope to show would be that there must be some form or forms of sensible intuition and that these must play some part in giving us any knowledge of objects. And it seems to be an argument like this that he has in mind when he speaks of the Aesthetic as containing a transcendental deduction of "the concepts of space and time":[1]

> Wir haben oben an den Begriffen des Raumes und der Zeit mit leichter Mühe begreiflich machen können, wie diese als Erkenntnisse a priori sich gleichwohl auf Gegenstände notwendig beziehen müssen; und eine synthetische Erkenntnis derselben, unabhängig von aller Erfahrung, möglich machten. Denn da nur vermittelst solcher reinen Formen der Sinnlichkeit uns ein Gegenstand erscheinen, d.i. ein Objekt der empirischen Anschauung sein kann, so sind Raum und Zeit reine Anschauungen, welche die Bedingung der Möglichkeit der Gegenstände als Erscheinungen a priori enthalten, und die Synthesis in denselben hat objektive Gültigkeit. (A 89/B 121f.)

The actual argument seems to show that there must be "such pure forms of sensibility", and space and time are brought

1. A 87/B 119.

on as filling the bill[1] - but without any argument to indicate why they should be chosen to fill the bill. Clearly, Kant does argue in the Aesthetic that space and time do play this part, but the argument does not claim that nothing else could.

Nominally, the Aesthetic contains three arguments for the apriority of space and (in B) four for that of time. Joined to these are two arguments for each, space and time, to show that they are intuitions, not concepts, because they have parts, not instances. One of each set is called a "transcendental exposition", the others are part of the "metaphysical exposition" of the concepts; but in fact the third argument of the Metaphysical Exposition of time covers much the same ground as the Transcendental Exposition. The difference between these two sorts of exposition is supposed to be that the metaphysical one is the analysis of a concept in such a way as to show it to be given a priori[2], whereas the transcendental one is the analysis of a concept in such a way as to show it to be required if some particular sort of synthetic a priori knowledge is to be possible.[3] Transcendental expositions

1. Though where this argument appears (as an argument) in the Transcendental Aesthetic I do not know.
2. B 38.
3. B 40. See also below, pp. 239f.

are thus like transcendental deductions in that they exhibit something as required for knowledge - only in the case of transcendental expositions it is some quite specific sort of synthetic a priori knowledge that is concerned.

The first argument of the mataphysical exposition in each case is that space, or time, is a priori because the notion of space or time cannot be abstracted from experience, since in order to represent objects as spatio-temporally located we must already have some notion of space and of time. By itself this argument is not entirely convincing, for it doesn't really show that one could not abstract the idea of a spatial or temporal location from observing objects which we, after having performed the abstraction, would now describe as spatio-temporally located. The first argument appeared alone in the Dissertation of 1770[1]; and it may be that Kant supplements it in the Critique with a second argument because he felt it was inadequate. This new argument is to the effect that although we can think of space and time as empty of objects we can never represent to ourselves the possibility of there being no space or no time: so that space and time are necessary a priori representations underlying all our experience.

1. Ak. II.: 398f., 402.

I think it is clear that, at least in this second argument, Kant is openly starting from a psychological premise.[1] It has sometimes been thought he can't be intending this because it would impugn the apriority of his conclusion. But I think that would be confused. The conclusion is that space and time are themselves a priori, the forms imposed by human minds to order the manifold of sensations they receive; this conclusion is a philosophical one, its truth is supposedly a necessary consequence of the fact that human experience is of a certain sort - but this fact is contingent, or at any rate Kant appears to think it so. If one grants the premise that it is unthinkable that there should be no space or time, and unthinkable presumably because of the way human beings are constituted and not just because experience always has been spatio-temporal, then it seems comparatively reasonable to describe space and time as a priori conditions of intuition[2], though Kant so far as I know never really shows (though he often states) that time is an a priori condition of all our intuition.

To settle any doubt about the psychological premise

1. See also Ak. XXIII: 22f. (Refls. XVII and XVIII), which lend themselves to this interpretation.
2. See further below, pp. 230ff.

of the second argument, two things further can be said. Firstly, as we have noticed, Kant recognizes the possibility that there should be beings who did not perceive spatio-temporally, so he cannot conceive of the argument as showing this not to be possible. Secondly, if it is the case that coherent experience (logically) must be temporal, or spatial, or both, this is not obviously the case, and it is surprising that Kant should say nothing at all to suggest the logical impossibility of a percipient none of whose experience is spatial or temporal. We have a metaphysical exposition where, if that were so, we should expect a full transcendental deduction of (specifically) space and time, not just of "such pure forms of sensibility".

The other argument, which in B Kant calls the transcendental exposition of space or time, in no way alleviates the contingent character of the premise of the second. Here the argument starts from mathematics, as a body of synthetic a priori knowledge, and says that space and time must be a priori if mathematics is to be possible. Mathematics may be a priori, and hence apply necessarily to all our experience, because space and time are a priori, conditions which our minds impose upon experience; but this is consistent with its being logically possible that minds of some other sort should impose other, different forms on their experience, which would give rise to a

different set of synthetic a priori truths about their experience. We call space and time a priori forms of intuition because we can know in advance of experience that any experience we have must be spatio-temporal[1]; and this is supposed to follow from some fact about us, some property which we as human beings have but which may not be shared by all possible percipients. The fact seems to be simply that we can't imagine a non-spatiotemporal experience; and though this does not really seem, as it stands, good enough - for I may not be able to imagine a chiliagon and yet cannot tell a priori of all my experience that it will be chiliagon-free - there is nothing formally wrong with this sort of argument.

So the findings of the Analytic of Concepts are much more general than those of the Aesthetic. And if the transcendental deduction of the categories is not really designed to show more than that the mind must be active in the twelve ways in conceptualizing, we may well feel after it that "this long and tiresome work", as Kant himself called it, was pretty useless because its conclusions are far too abstract to be of interest except for their own sake. This is why Kant was right to bring on the idea of the schematism at this point. It is not that these

1. Cf. B2, B41.

activities of the mind "~~having~~ have nothing in common with" the deliverances of sensible intuition, but that we can't have more than the vaguest idea of what is going on unless we descend to a lower level of generality. And so in the section on the Principles, of which the chapter on schematism is a sort of introductory part, the argument returns to the level of the Aesthetic. Here again, it is not <u>all</u> knowledge involving sensible intuition that is the subject of the discussion but human knowledge, which - as has been postulated - involves spatial and above all temporal intuition. (For all our experience must be temporal, whereas only some - "outer" experience - need be spatial). This is not to say that the work of the deduction of the categories is in any way overset or made redundant; its conclusion is taken to stand, and what is asked is whether it can't be made less abstract by taking into account this restriction on the sort of knowledge we are concerned with.

If, taking the possibility of (sensible) knowledge or experience in general as our premise, as in the deduction of the categories, we can show by transcendental arguments that certain conditions must obtain; and if, taking the possibility of knowledge or experience of some more restricted sort as our starting-point, as Kant takes himself

to be doing in his work on the Principles[1], we can show by transcendental arguments that certain other conditions must obtain; what sort of correspondence may we expect to find between the two sets of conditions?

If we can give a transcendental deduction of a particular category we have thereby also shown the need for the corresponding schematized category. For if the categories are necessary for any experience at all they will certainly be necessary for the specific sort of experience we are now interested in - in this case, temporal experience. But the particular ways they have to apply in this sort of experience may be called their schemata[2]. So if the categories are kinds of synthetic

1. The relation between categories and principles is made clear at A 136/B 175:

> Diese transzendentale Doktrin der Urteilskraft wird nun zwei Hauptstücke enthalten: das erste, welches von der sinnlichen Bedingung handelt, unter welcher reine Verstandesbegriffe allein gebraucht werden können, d.i. von dem Schematismus des reinen Verstandes; das zweite aber von denen synthetischen Urteilen, welche aus reinen Verstandesbegriffen unter diesen Bedingungen a priori herfliessen, und allen übrigen Erkenntnissen a priori zum Grunde liegen, d.i. von den Grundsätzen des reinen Verstandes.

2. Thus A 140/B 179:

> Wir wollen diese formale und reine Bedingung der Sinnlichkeit, auf welche der Verstandesbegriff in seinem Gebrauch restringiert ist, das Schema dieses Verstandesbegriffs, und das Verfahren des Verstandes mit diesen Schemata den Schematismus des reinen Verstandes nennen.

Cf. also A 145/B 184; A 146/B 185f.

mental activity, their schemata will just be more specific kinds of synthetic mental activity: the original kinds, as they work for experience of the more fully specified kind. For instance, let us grant Kant is right in thinking that not all experience is necessarily temporal; and that the deduction of the categories establishes that one of the pure categories required for all experience is (the capacity to formulate) hypothetical judgements. Now if we add the premise that all (relevant) experience is temporal, naturally it follows still that we shall have to be able to formulate hypothetical judgements. But in view of the temporality of experience we shall have to say more than this, viz. that we shall have to be able to make hypotheticals about events: tensed, or at least dated, hypothetical judgements. And we can call the sort of mental activity now involved the schematized category. Of course the pure category is still at work; the sort of mental activity which constitutes the schematized category is a sub-species of the species of mental activity which constitutes the pure category. There are at least as many other sub-species of that species as there are intelligible alternative starting-points for the transcendental arguments. If, for example, there were no contradiction in supposing percipient beings who didn't perceive things temporally but rather in another order, which I shall call A-order,

the requirement of hypothetical judgements would still apply, but in the case of these beings it would come out as the requirement that hypothetical judgements about relations of things in the A-order be formulable, and this would constitute a different sub-species of the same species of mental activity - a different schematized category.

We have so far seen that there must be at least as many schematized categories as pure categories. But must there be only so many? Might it not be possible to schematize the same category in alternative ways given that experience is temporal or A-ordered? No: not if we consider the schematized category to be just the-category-as-applied-to-time, or to-A-order. And unless we do this the transcendental necessity for the schematised category will not have been shown. If, to take the same example, experience must be temporal and it must be possible to make hypothetical judgements about it, there seem, at least prima facie, to be two alternatives - the hypothetical judgements may be tensed, or they may be built of dated eternal sentences. Neither of these alternatives is transcendentally required. It may, of course, be possible to show that one of them derives its sense from the other and not vice versa, and if that were done then that alternative would be shown to be transcendentally required, and it would provide the schema of the category in time;

but then it would not be an alternative. Kant's fundamental error in the Analytic of Principles is to make his schematized categories more specific than his argument warrants, not realizing that at this level of specificity there are alternatives. The fact that we must make affirmative and negative judgements about things in time does not warrant the assertion that we have to make judgements about intensive magnitudes.

The fact that there is a one-one correspondence between the pure categories and their schemata in time (or in the A-order) in no way suggests that it may not be possible to show very much more by means of transcendental arguments starting from some more restricted sort of knowledge or experience than one can by starting from the idea of knowledge or experience in general (or for beings with sensible intuition). It may be possible to show much more about temporal or A-ordered experience than that it must be subject to the appropriately schematized categories. Kant recognizes this at A 149/B 188.

But schematizing the pure categories in time, as Kant does, should not really lead to very much more valuable conclusions than did the transcendental deduction öf the categories. As I have said, I shall not consider his arguments for each of the Principles here; they have been examined frequently, and their deficiencies have been made fairly clear. For my present purpose it is sufficient to

say that adding to the conclusions of the deduction the extra condition that experience must be temporal - or even spatio-temporal - does not really make much difference. It may show, for example, that we have to make hypotheticals about events, but not that these hypotheticals must be causal. They might express the non-causal regularity-generalisations considered above; or they might never relate events at different times, only simultaneous events. But even if they did have to be causal this would not show that <u>every</u> event must be the subject of the apodosis of such a condition$_K^{al}$ as Kant thinks. It may be for this reason that he brings on the detailed arguments for this conclusion that he gives in the Second Analogy; but these still involve the same mistake - supposing that hypotheticals about events in the time-order have to be causal. They involve other mistakes as well.

We can find a place for a sort of schematism even if we stop talking about categories and consider quite generally the sorts of things transcendental arguments can be used to establish. For it will generally be the case that if <u>x</u> is shown to be a necessary condition of all knowledge or experience of kind <u>K</u>, <u>x</u> will still be a necessary condition of knowledge or experience of kind <u>k'</u>, where <u>k'</u> is a species of <u>k</u>. And it may well be that <u>x</u> comes in differently under different conditions <u>k'</u>, so that in order to see just how <u>x</u> applies to the experience in

question we have to "schematize" x by taking it together with k' in some way, and thence working out that x', this specification of the more general x, will be the corresponding necessary condition of experience of kind k'. In a way x' will inherit the necessity of x.

As an illustration of this, let x be the requirement that one be able to talk about things which can exist unperceived, which we have seen to be a requirement for self-consciously public experience. Now it seems that for us the epistemologically basic unperceived existents are material objects; but it is not easy to show it to be a requirement for self-consciously public experience that one have the concept of a material object. I have suggested that one's experience might be of a sound- or flash-world[1], and I shall suggest that it is not enough to introduce space as an extra condition on experience, as one might experience a purely static spatial world in which items could be unperceived because obscured by the brightness of others.[2] But perhaps one can find some value for k' such that if one's knowledge or experience is to be of kind k' then one's concept of something which can exist

1. Above, pp. 92 ff.
2. Below, pp. 220 f., 224 ff.

unperceived must appear in the form of the concept of a material object. So that this concept will in some sense inherit the transcendental necessity of the more general concept of an unperceived existent.

And it might be possible to put this to an epistemological use which was far from trivial. Strawson regards transcendental arguments as a weapon against scepticism[1]; and Kant does so as well, for although the main object of his attack is ultimately dogmatism the first move is to subject the dogmatist to sceptical attack, as a result of which a critique of pure reason comes to be seen necessary, as showing which of reason's claims can, and which cannot, be rescued from scepticism[2]. The function of the sceptic is to call in question any assertions we may make which go beyond the evidence; transcendental philosophy answers him by showing that his doubt undermines itself, since it can itself make sense only if it is wrong: it is transcendentally necessary to make the assertion that goes beyond the evidence, and the necessary conditions of self-conscious knowledge or experience in general are a fortiori necessary conditions of the intelligibility of sceptical doubt.

1. Individuals, e.g. p.106.
2. A 738-769/B 766-797.

Kant's answer to Hume's doubts about causality is that unless every event has a cause no experience is possible; similarly, though I hope with more valid argument, we can reply to the sceptic who refuses to accept that anything could count as an unperceived existent that he is denying a condition of all experience and thereby of the conceptual scheme within which his denial is expressed[1]. The concept of an unperceived existent reaches outbbeyond all possible experience, and the concept of a material object inherits this feature from it, so that here again the move beyond the evidence requires justification, for sceptics can question its legitimacy. Indeed, this is a problem which arises wherever we have an expression the meaning of which goes beyond the criteria, as in the case of "pain" and many other psychological words. In such cases it sounds a bit weak just to say that the move beyond the evidence is justified on the grounds of simplicity of theory, for this suggests that at the cost of slight inconvenience we could cease to ascribe pains and drop material objects from our ontology; which seems untrue. It may not be a

1. He may not be convinced; for he may doubt even the simplest analytic truths, which we require for our transcendental arguments. Obviously if he sticks to his guns he can never be convinced of anything at all; but it is enough to show that this position is irrational - if he is right, his own doubts lack sense.

condition of having self-consciously public experience at all that one should have the concept of a material object; so that the sceptic can question the move from sense-data to material objects without entirely destroying his own position, since he may satisfy the general transcendentally necessary condition by finding unperceived existents of some other sort. But if we can find a value for k', such that it is a necessary condition of having experience of kind k' that one should have the concept of a material object, then at least the sceptic can be shown that if his doubt makes sense his own experience cannot be of kind k'.

Whether this is an effective reply to the sceptic will of course depend on the value of k'. This raises the question to what extent it is acceptable to limit the kind of knowledge or experience that is to provide the starting-point for a transcendental argument. How much "schematism" can we get away with? What, indeed, justifies our considering the requirement that experience should be - perhaps - spatial, or Kant's considering the requirement that experience should be temporal, as reasonable restrictions of the starting-condition? If conclusions reached on these bases are of use against sceptics, this must be because the sceptics are not inclined to dispute the restriction. And the arguments Kant calls transcendental expositions, which belong to his "analytic method" as

pursued in the Prolegomena, are of use simply because he can get people to accept, quite readily, the propositions from which he starts - e.g. that we have synthetic a priori geometrical knowledge. But he recognizes that this "analytic" way of proceeding is only of heuristic value; it still leaves plenty of room for sceptical attack, which proceeding by the "synthetic method", as in the Critique, is supposed to enable us to avoid. If one can "schematize" endlessly, restricting more and more one's initial definition of the knowledge or experience in the conditions of which one is interested, so long as one's interlocutor agrees that there is knowledge or experience of that sort, then any difference between the "synthetic method" that we have been mainly considering and the "analytic method" will disappear.

If our arguments are to be of much use against sceptics, we cannot allow them to start from premisses which no-one in fact disputes. Sceptics are on the whole possible people rather than actual ones; and the fact that there is nobody around who is inclined to call in question a particular leap beyond the evidence does not justify that leap. Although nowadays there is no general agreement about religious matters, there have been times at which there was; and it is clearly possible that a nuclear war might destroy everyone save a small group of Adventists,

who were at the top of a mountain waiting for the end of the world and so missed it. This does not mean that theistic claims would then be justified; nor would the legitimacy of these claims be affected if the Adventists proceeded to give birth to a collection of atheists and then died off.

To resist sceptical attack we should have premisses which no-one <u>can</u> dispute. Arguments about necessary conditions of there being any experience at all are successful in this, since the sceptic's doubt must either be wrong or unintelligible. If arguments about necessary conditions for experience of a restricted sort (<u>k'</u>) are to be successful, this cannot be because experience of some other kind is demonstrably impossible. But if for the sceptic's doubt to be intelligible he must admit that his experience is totally remote from the sort of experience available to human beings, then his doubt can no longer interest us.

Strawson, in <u>Individuals</u>[1], in a sense assigns a value to <u>k'</u>, for he holds that bodies, material objects, are basic particulars in <u>our conceptual scheme</u> - given our conceptual scheme we must be able to talk about material objects if we are to be able to talk about particulars at

1. Chapter 1.

all. And apparently he does not mean by "our conceptual scheme" simply the restricted starting-point for transcendental arguments that we, with Kant, have so far formulated (self-consciously public experience of a spatio-temporal world), but some starting-point more restricted still[1]. But he does not make clear what the further restriction is, and we know nothing more about our conceptual scheme than that it is the conceptual scheme used by speakers of English - and perhaps by speakers of some other languages as well. So, thus far, no analysis of its necessary conditions, so far as they differ from the necessary conditions of language in general, is possible except by examining the facts about English[2].

An examination of the conditions for knowing English is, certainly, an examination of the conditions for knowledge of a specific sort. But the limitation has now become so great that much of the interest of the argument has

1. Though this may be wrong. We saw that in Bounds of Sense he argues that experience of a spatial world presupposes that one have the idea of something in space which is persistent, movable, and capable of unperceived existence (see above, pp. 87ff.; below, p. 224); and the argument of Individuals, p.39, may be of the same sort - though on the next page he calls it "general and vague", and does not rely on it.
2. And that is how Strawson sometimes seems to proceed: cf. Individuals, pp.41ff.

disappeared, as has its value against scepticism. One could not know English unless one knew that English nouns generally form plurals in -s; but there is little temptation to glorify this with the title of a transcendental necessity. Nor would one, indeed, naturally account it part of "our conceptual scheme", because one thinks of this as something unavoidable in a way that knowing English is not - one could so easily start speaking French instead, and even forget one's English altogether. But then what is "our conceptual scheme"? Even if it is what all natural languages have in common - and I do not doubt that all other natural languages assign to material objects the same sort of position that English does - it still seems avoidable: one could not speak any natural language unless one possessed some capacity for making sounds, but one could become a Trappist and communicate by writing or sign-language.

To invoke "our conceptual scheme" is to name the problem and not to solve it; but it is to give it an appropriate name. "Conceptual scheme" suggests a pattern of concepts connected together by relations of logical and (more weakly) epistemological dependence; calling it "our conceptual scheme" suggests that we are unavoidably stuck with it, even though experience might not be impossible for beings of some sort with a different

conceptual scheme. And both these suggestions are appropriate here.

A conceptual scheme may be thought of as a set of empirical interpretations of the conditions for there being any experience. These transcendentally necessary conditions are rather general; they might be satisfied under quite different sorts of empirical circumstance. In some cases at least they can be applied in various incompatible ways: it is necessary to see some things as similar, though it doesn't matter which - but if we saw all events as similar we should be as stuck as if we saw none as similar. So, following Kant's suggestion that schematizing the categories provides empirical criteria for them, we can suppose that a conceptual scheme makes concrete the general transcendental conditions, by saying just what sorts of thing in actual or possible experience are to count as the instantiations of the general conditions, inheriting their necessity in the form of epistemological priority within the conceptual scheme. Thus, all sorts of items could be held to be capable (in some possible world) of unperceived existence, but our conceptual scheme chooses material bodies (and persons) as the basic ones, and requires any others to be introduced only afterwards and dependently. Similarly, it is transcendentally necessary that any percipient see

his experience as regular, but it doesn't matter where the regularities come; our conceptual scheme introduces a severe, though imprecise, limitation into where the regularities can come by limiting the range of what we can see as similar. The general transcendentally necessary conditions do not lack sense by themselves; there is no problem over their status, as there was over the status of Kant's pure categories. They can just be applied in various ways, and schematising them, fitting them into a conceptual scheme, gives us a way of applying them. They do not lead some sort of independent existence divorced from any criteria; it is just that only the general conditions, and not any particular schematisation of them, can be shown to be necessary for any experience whatever.

If we define a conceptual scheme in terms of the ways it specifies or schematizes the general transcendental conditions for there being any experience at all, it can never be a discovery that a particular conceptual scheme treats material objects as basic. For this is part of what makes it that particular conceptual scheme. It can, of course, be a discovery that <u>our</u> conceptual scheme does this, but it is an empirical discovery - it is a matter of finding which of the logically possible conceptual schemes we in fact use. It does not involve any analysis of the

notion "our conceptual scheme".

A conceptual scheme might be less general than ours is - the range of types of instance it allowed for each general transcendental condition might be narrower. It might, for example, allow only rectangular brown material objects to count as basic particulars, not other sorts of material object as well. Ways of referring to these other material objects could be introduced, but they would depend on the speaker's ability already to refer to rectangular brown things. Again, a conceptual scheme could be more general than ours is, allowing a broader range of empirical circumstances in which each of the general conditions could be cashed; for example, if events as well as material objects could be basic particulars, references to neither being dependent on references to the other. And perhaps a conceptual scheme might be less general than ours with respect to the empirical cashing of certain of the general conditions, more general than ours with respect to that of others.

There might also be conceptual schemes quite different from ours, where the schematism of the general transcendental conditions would not overlap with ours - if material objects and not events are basic particulars for us, a different conceptual scheme might treat events and not material objects as basic. And there might be conceptual

schemes which were in part different from ours, and overlapped in part; and in that case some of the general conditions would be cashed quite differently from the way we cash them, while others might be schematized either more generally, or less generally, or in exactly the same way as they are in our conceptual scheme.

Conceptual schemes differing from our own only in being more general offer no refuge for the sceptic who questions the moves our conceptual scheme requires us to make, like the move from sense-data to material objects. For his conceptual scheme will require them too. As for the sceptic who retreats to conceptual schemes of the other sorts, the answer to him comes down to this: that it is a very general fact about human beings that we use the conceptual scheme we do, and that no other conceptual scheme is an option for us. So that although we can understand the logical possibility of those other conceptual schemes from which his doubts can be stated, we can never treat these doubts as more than idle. This is essentially what Kant argues regarding space and time when he gives his metaphysical expositions of them. And we can add, perhaps, that the sceptic is recommending a change of conceptual scheme, while such changes, if they were possible at all, could only be made on pragmatic grounds - on the basis of considerations like theoretical

simplicity. But on grounds of theoretical simplicity, elegance, etc., our system seems preferable to any other I can think of. However, it is not clear that this consideration comes to anything more than the other one, because it is not clear that any other sense can be given to "seems simpler" here than "seems simpler to us": a being with a different conceptual scheme might be expected to think his own the simplest, and the difference between us would then be irresoluble.

It is sometimes thought that the sceptic's position must be idle in a very strong sense, in that the conceptual scheme in which his doubt gains sense must be totally unintelligible to us. If we ran into someone with such a conceptual scheme we could never understand what he said; we might not even be able to understand that he was trying to communicate. It would make things nice and neat if this were true; unfortunately, it is false. Of course, some conceptual schemes are so entirely foreign to us that we could not even begin to understand someone operating within them: what appeared simple to him might be ungraspably complex to us. But not all the logically possible conceptual schemes alternative to our own are so remote from ours: the sceptic can make his point by adopting a conceptual scheme much nearer home. Let us take again the sceptic who questions the legitimacy of speaking about material objects, and let us suppose that

for him it is not material objects that are the basic unperceived existents but patterns of events like flashes and bangs. If I came across someone - perhaps a Martian - with a conceptual scheme like this I should no doubt experience some initial difficulty in communicating with him, but communication would not be impossible. It might take me some time before it occurred to me that he was talking about flash-patterns, but when that hypothesis finally did strike me there is no obvious reason why I should not be able to translate his language into mine, sufficiently well for us to be able to talk about a good many things at least. The same goes for the case of someone who made just a few classifications which strike us as strange - colour-blind people provide no conceptual difficulties, and communication would never altogether break down with someone who thought grass was grue.

But the translations I effected could not remain faithful to the Martian stranger's conceptual scheme. If he says "F" when a certain flash-pattern occurs, and I therefore translate "F" as "Flash-pattern x is occurring", I am rendering what was an expression for a basic particular in his conceptual scheme by an expression for a non-basic particular in mine. And if I am to translate "x is grue", I render it "x is green up to time t and blue thereafter", thereby translating by a conjunctive colour-expression

what was originally a primitive colour-expression. So that if the stranger with the event-based conceptual scheme poses the sceptical doubt about material objects, I may be able to understand him; I may even understand that the question is not an absurd one for him to raise; but I cannot take his doubt seriously, because in order to understand it I have had to express it in my own terms, which do presuppose the falsity of that doubt.

That I have had to do this is due to the very fundamental fact that human beings do think in certain ways, alternatives to which are not open to them. They see material objects as the most obvious examples (together with persons) of things capable of existing unperceived; they introduce other particulars in ways which depend on their already being able to talk about material objects. They naturally see similarities in certain places, and dissimilarities in other places: "green" and "blue" are simple colour-predicates for them, and "grue" is not - it is naturally not as "well-entrenched"[1].

But even if alternative conceptual schemes were open to us, - even if we could come not only to understand what

1. Though human beings have no fixed way of dividing up the colour-continuum they all agree upon. Some Eskimos notoriously distinguish twenty-seven shades of white; but they do not co-classify (with respect to colour) things white before t and things black thereafter, which would be quite a different matter.

the Martian or the "grue"-user is saying but also to think like him - still nothing could justify us in taking seriously sceptical doubts about the moves our conceptual scheme requires us to make. The grounds for choosing between conceptual schemes can only be pragmatic[1]; for although some conceptual schemes can only be made to accommodate the facts by introducing considerable complexities, such extension will always be possible, so that no conceptual scheme can be rejected simply on the grounds that it doesn't fit the facts. The differences between conceptual schemes lie in that a statement which in one is pretty simple (e.g. in ours: "There is a table here") becomes exceedingly complex when translated into another (e.g. the Martian stranger's event-based language). And if the contest is to be fought on these grounds our conceptual scheme looks likely to defeat any proposed contender: it provides a simpler way of describing the world than any alternative it is easy to envisage. What happens in the world is more readily explained by reference to material objects than by reference to flash-patterns; the "grue"-user will require more complex laws

1. Cf. R. Carnap, "Empiricism, Semantics, and Ontology". Revue Internationale de Philosophie, 1950.

than we about the colour of grass, if we are right in our confident assumption that it will remain green after t.

Physics, now or at some future time, may show that the most efficient conceptual scheme for dealing with very small particles or with astronomically great distances might differ slightly from ours. But this would not establish that it would be simpler for us to adopt that conceptual system, for it is only rarely that we concern ourselves with fundamental particles and interstellar space. In order that physics should become simpler all our ordinary assertions about the everyday world would become more complicated, if we did adopt it; just as it makes elementary optics unnecessarily complicated to abandon the assumption that light-rays travel in Euclidean straight lines.

By giving up our conceptual scheme in exchange for one less general, for example, the one in which material objects which were not brown and rectangular did not count as basic particulars, we might be able to effect a reduction in our ontology. For we could then allow that the sceptic's doubt was justified in the case of any putative material objects which were not brown and rectangular. This, however, would not give us good pragmatic grounds for adopting such a scheme, because the reduction in the ontology would be more than compensated

for by increased complexity of laws. Wherever we now have one law dealing with the behaviour of material objects in particular circumstances, we should then need at least two quite different laws, one dealing with (brown, rectangular) material objects and one dealing with things of some other sort - sets of sense-data, perhaps. The obvious thing to do would be to avoid the multiplication of laws by disallowing the sceptic's doubt, and allowing the move from sense-data to material objects to be valid generally. But in that case the new conceptual scheme could no longer offer a reduction in our ontology; it would differ from our present one in that it counted only brown, rectangular material objects as basic particulars, but this is to its disadvantage. For it makes statements about other material objects more complex, since they will have to entail statements about the brown, rectangular things in terms of which alone references to other material objects can be introduced[1]. So the fact that an alternative

1. The introduction might work something like this. We observe that our basic particulars bear to one another spatial relations, and we specify our spatial matrix by reference to basic particulars. It won't matter which we count as origin, but nothing other than a basic particular can count as origin. We then find it convenient to reify our other sense-data, and do this by saying that at p_1 there is an F, at p_2 another F, at p_3 a G, etc. But since p_1, p_2, and p_3 are positions on the matrix which was set up by reference to basic particulars, and since no sense has been given to saying that there might be an F in some differently constituted spatial system, identifying references to F's and G's presuppose the ability to make identifying references to basic particulars and not vice versa.

conceptual scheme can reduce the number of basic items does not mean that it is to be preferred to ours on pragmatic grounds. And it is perhaps worth pointing out that no alternative conceptual scheme could cut down on the number of types of item that have to be considered basic, for that is dictated by the general transcendental conditions which different conceptual systems schematize differently.

To this, however, a partial reply remains open to the sceptic. He will still be open to the objection that alternative conceptual schemes, such as his, are not available to human beings. But he can argue that we are weighting things unjustifiably in our own favour if we claim that our conceptual scheme is preferable to his on pragmatic grounds. We should say only that it seems to us the simplest way of looking at things, and should recognise that to him some quite different system might seem simpler. For he may see similarities and differences where we see none; and if similarity and difference are in the eye of the beholder, simplicity must be too.

To say that I see various things as similar - perhaps, as all falling under the predicate "white" - is just to say that I classify them together in one respect. I do not make the classification on the basis of some an sich similarity which the things possess: if there were such

an sich similarities they would do no work, they would be irrelevant to the classification. For any apparently quite arbitrary set of things could be classified together by someone - they would then be similar for him. Someone might see no similarities at all between two billiard balls, and find it hard to distinguish one of them from the Taj Mahal; or see no similarity in colour between things green before t and green after t, and no difference between things green before t and things blue afterwards. If he did he would commit no logical error, and he might be able to get a coherent conceptual scheme going. And what factual error would he commit? If we say that he sees similarities where he shouldn't and differences where he shouldn't, what can this mean, except that he sees them where we don't?[1]

Following a rule, as in applying a concept to different instances, is a matter of what one does; and people might act quite differently. Or they might agree for a time and then diverge, as in the grue case or in Wittgenstein's favourite example of counting, where they agree about how to count up to 1000 but then disagree

1. Cf. Wittgenstein, Philosophical Invesigations; D.F. Pears, "Universals", Philosophical Quarterly, 1951.

as to whether going on 1001, 1002, 1003, or instead 1002, 1004, 2572, constitutes going on in the same way as before[1]. To say that it is a matter of what one does is not of course to say that one has any choice here; just that it is a contingent fact about us that we see similarities and differences in roughly the way we do. Perhaps I should not say "People might act differently"; it is such a basic fact about people that they all see similarities in roughly the same places that there is some inclination to build it into one's notion of a person. A creature who kept confusing the Taj Mahal with a billiard ball would not be one of us. But he might be the possessor of a coherent conceptual scheme, none the less.

Two white things can be called objectively similar in that everyone (in fact) co-classifies white objects in more or less the same way: the similarity is public. Even the predicates of a private language can be said to pick out objective similarities, if what is now meant is that it is possible to make and correct mistakes in classification. There is, as we saw above[2], a distinction

1. Much of Wittgenstein's interest, particularly in the Remarks on the Foundations of Mathematics, lies in the application of this to mathematics. But some people cling more desperately to Platonism in mathematics than to Platonism with respect to properties of objects, so - since they are inessential to my purpose - I shall put the mathematical applications to one side. Even if there is a right way and a wrong way to continue the number series, apart from what people do, numbers can only be applied to things if ordinary empirical predicates can.
2. pp. 116 - 131.

between misusing a word and giving it a new meaning: one can remember how one had decided to use the word, and recognize in the light of this that one has misused it. The fact that we can classify wrongly does not make it any the less true that the ultimate justification for saying that two things are both white is our propensity to co-classify them with respect to colour. What constitutes following any rule is ultimately a matter of practice - "This is how we do it" - but it is not impossible to follow rules wrongly. So the similarities we find in the world can be called objective; but this gives them no privilege over the quite different similarities the sceptic sees, for these may be objective too in just the same ways.

It might be said nevertheless that this is not enough to show that the sceptic's alternative conceptual scheme could provide him with as simple a way of looking at the world as ours does for us, or even a simpler way. There is no difficulty in supposing that for somebody "grue" might be as simple a predicate as "green" is for me, but perhaps this can only be the case when we are dealing with secondary qualities. An alternative conceptual scheme in which for instance event-patterns were basic particulars would have to cut up the world

much more differently from ours than this if it were to be equally simple: merely rearranging the ways in which secondary qualities are discriminated will not help. And it might be suggested that alternatives to our primary-quality discriminations are unthinkable. For things in the world can be classified according to (for example) their sizes and shapes, and it is not easy to imagine how there could be any genuine and equally simple alternatives to our classifications of size and shape. People might of course differ from us in discriminating them more, or less, finely than we do; but these will only be the sort of differences that can be settled by one party acquiring microscopes or more efficient measuring techniques. Could someone really fail to make the sort of discrimination we do between a small sugar-cube and a life-size sugar model of the Taj Mahal?

Someone who said this would be pointing to a genuine difference between primary and secondary qualities. Someone who did not discriminate more or less as we do with respect to size and shape could not be said to discriminate with respect to size or shape at all. Whereas someone who discriminated colours differently, as in the "grue" case, could still be said to be discriminating with respect to colour. But this does not show that the sceptic's conceptual

scheme is at a disadvantage as against our own, for unless we can give (as we have not done) a general transcendental deduction of primary-quality concepts, and thereby indeed of the concept of a material object, the sceptic can say that his alternative conceptual scheme is perfectly coherent though it dispenses altogether with such concepts as size and shape.

Our concepts of size and shape lie at the basis of our physics; without them we should not be able to explain many of the things that happen in the world. A ball of 3-inch diameter fits neatly through a ring with a hole in it of the same diameter; we can explain this readily, but without the notions of size and shape we should find it an entirely unpredictable event. And it is true that a being different from us who lacked these notions might also be quite unable to explain the passing of balls through rings. The world for him would be full of surprises, and there would be pragmatic grounds for saying our conceptual scheme was better than his. But on the other hand this might not be so. He might be able to explain all the phenomena - only not in terms of size and shape but by means of laws quite different from ours, which nevertheless gave the same results in all actual instances. His law statements might seem to us, so far as we could understand them at all, very complex, and

yet seem very simple to him just because he classifies things as similar in such a different way; and our laws might seem very complex to him. He might perhaps have one single, elementary generalisation which explained (1) the passage of blue 3-inch balls through 3-inch rings, (2) the behaviour of litmus paper in acid, (3) that green objects weighing between one and two tons fall when dropped, etc., etc. We might be unable to see anything in common among all and only what are for him the instances of a single general term, and be able to approximate understanding only by listing the individual instances (a task which in an infinite universe could prove infinite); he might have the same difficulty with the concepts of size and shape.

So it cannot really be objected to the sceptic that the conceptual scheme in which his doubts can arise is to be rejected on pragmatic grounds; for he might see the world very differently from us. On the other hand, this makes his doubt even more idle for us: if his conceptual scheme is not to be rejectable on pragmatic grounds, its differences from ours must reach deeply into the ways in which things are classified, so that to take the sceptic's doubt seriously our ways of thinking would have to be very different from what they are.

And thus ultimately we reject the sceptic's doubts - where these are not the doubts which general transcendental

arguments, about the necessary conditions of having any experience whatever, can show to be untenable, but are rather the sort of doubts which can sensibly be raised in other conceptual schemes, though not in our own - ultimately we reject these doubts by relying on the fundamental psychological fact that these alternative conceptual schemes are not live options for us. Since this is a factual claim, the sceptic would be within his rights if he were to dispute it. Human beings, he would say, could be taught to see things in quite different ways - though this might mean teaching them a new language entirely from scratch, since any rules for translating between the new language and the old ones might be infinite in their complexity. If they found the new way of looking at things just as good pragmatically as the old, they would be unable to decide between the sceptic and us. But if they found the new way better, the sceptic's doubts would be vindicated.

The issue between the sceptic and his opponent is thus an empirical one, in which no evidence seems to support the sceptic. He commits no logical error, but he makes an empirical assertion - that the sort of conceptual system within which his doubt can be stated, without its being refuted by the nature of that system,

is available to human beings. And this assertion he cannot back up. Alternative conceptual systems in which his doubts can be raised may be possible, but they are not available to us; so we cannot take these doubts seriously. This seems the best reply we <u>can</u> make to the sceptic who calls in question such moves beyond the evidence as that from sense-data to material objects. And I think it is an adequate reply.

I conclude that transcendental arguments from limited starting-points, arguments about conditions for knowledge or experience of a particular sort, can be of interest as arguments against possible sceptics. For by restricting the kind of knowledge or experience one can limit the class of possible conceptual schemes being examined, and our own may be a member of this limited class. Kant's restriction to spatio-temporal experience, in the Analytic of Principles, is of this kind. It is of greater interest than his examination of the conditions for geometry and arithmetic, in the Transcendental Expositions of the Concepts of Space and Time, or that of the conditions for pure natural science in the <u>Prolegomena</u>; simply because these arguments get to the same results after building more into their premisses. It is desirable to extrude the smallest possible quantity of neck.

Transcendental arguments from limited starting-points are of interest because (if they are valid) the sceptic can only avoid their conclusions and escape the charge that his own doubt is unintelligible by what is in effect a trick: namely, by holding that although we cannot make sense of his doubt nevertheless he can, because he possesses a conceptual scheme which enables him to express it as a sensible question. But unless this conceptual scheme is one that we could adopt, his doubt cannot interest us, except by reminding us that it may be logically possible for a being to think of the world quite differently from us; and this is not news. Whether we could adopt it is an empirical matter, but no evidence suggests we could and it seems an implausible thing to suppose. Kant's argument that it is implausible to suppose that we could adopt a conceptual scheme not based on space and time is his Metaphysical Exposition of these concepts. And these restricted transcendental arguments are of interest also because they exhibit how the general conditions of any experience work in human experience: they show how we schematize them. If arguments of this sort can be used, there are some grounds for hope that it may be possible to use valid transcendental arguments to reach more interesting conclusions than has hitherto been done.

II. Metaphysical Arguments

Kant did not regard his work in theoretical philosophy as completed with the Critique of Pure Reason; nor that in practical philosophy as completed with the Critique of Practical Reason. The Critiques clear the ground for what he regarded, or frequently claimed to regard, as the real task of the philosopher, metaphysics. His use of the term "metaphysics" fluctuates considerably. Sometimes he uses it in such a way as to include the work of the Critiques[1]; and we have already seen that he specially calls "metaphysical" two sorts of argument that occur in KdrV, the Metaphysical Exposition (of space and time) and the Metaphysical Deduction (of the categories). At other times he contrasts metaphysics with the work of the Critiques, the propädeutic to it.[2] The more restricted metaphysics of this latter usage can be divided into two parts, transcendent and immanent; the former being the bad metaphysics the illegitimacy of which he shows, the latter being that remaining part of metaphysics whose title he has assured and which can now proceed along the secure path of a science. The immanent metaphysics of nature is worked out in the Metaphysische Anfangsgründe der Naturwissenschaft[3], which was published in 1786 and written after the Grundlegung, shortly before the second edition of KdrV. He intended to produce a metaphysics of morals at about the same time[4],

1. Thus A 845/B 873; Ak. IV: 469f.
2. Thus A 841/B 869; Ak. IV: 279 (Prol. sect. 5).
3. Abbreviated MAdN. Kant did not regard this as completing the metaphysics of nature, only as providing the outlines and beginning it.
4. See L.W. Beck, Commentary on Kant's Critique of Practical Reason (Chicago, paperback edn. 1963), pp.11, 13.

but wrote the Critique of Practical Reason instead and did not finally publish the Metaphysik der Sitten until 1797.

Each of these works on immanent metaphysics consists prima facie of two parts. Metaphysics of morals can be divided into two parts, one examining the concept of right and the other the concept of virtue: the former being concerned with action which, regardless of motive, is in accordance with the moral law, and the latter with action which has as its motive respect for the moral law.[1] Metaphysics of nature appears to divide into two parts likewise, one concerned with physics and the other with psychology: for there are two sorts of objects of our senses, physical bodies and minds, and so we can attempt metaphysical analyses of the concepts of matter and mind with the hope that we should thereby assist physics and psychology by the discovery of a priori truths fundamental to them.[2] However no such analysis of the concept of mind is possible, Kant tells us. His reason is that mathematics cannot be applied to the phenomena of inner sense, only to those of outer sense. This matters, he says, because the task of a metaphysical analysis of the concept of mind would be to exhibit the conditions for the possibility of minds: but all that an analysis of the concept could show by itself would be that the concept contained no contradiction. In order to show that the object (a mind) is really possible the concept must be constructed in pure intuition - but to do that is to do mathematics.[3]

This is confused. He is saying that, because by means of analytic propositions involving the concept of mind (or

1. Ak. VI: 218 ff.
2. Ak. IV: 470.
3. Ibid.

matter) nothing is shown about the possibility of there being minds (or matter) except that no actual contradiction is involved, we must appeal to mathematics, which constructs concepts in intuition, as the only source of synthetic a priori knowledge on the question. But in KdrV there was another alternative: to establish synthetic a priori propositions by means of transcendental argument. He recognizes this here:

> Also mag zwar eine reine Philosophie der Natur überhaupt, d.i. diejenige, die nur das, was den Begriff einer Natur im allgemeinen ausmacht, untersucht, auch ohne Mathematik möglich sein,

but continues,

> aber eine reine Naturlehre über bestimmte Naturdinge (Körperlehre und Seelenlehre) ist nur vermittelst der Mathematik möglich, und, da in jeder Naturlehre nur so viel eigentliche Wissenschaft angetroffen wird, als sich darin Erkenntnis a priori befindet, so wird Naturlehre nur so viel eigentliche Wissenschaft enthalten, als Mathematik in ihr angewandt werden kann. (Ak. IV: 470).

But why? What is worrying Kant is that he thinks that it cannot be shown by any general transcendental argument that there must be matter or minds, or at any rate matter or minds of the relatively specific sorts which are the concern of physics and psychology.[1] I shall argue below that he does hold this view; it underlies his separation of immanent metaphysics from the transcendental or critical philosophy of KdrV. He expresses this point here by saying that without the help of mathematics it cannot be shown a priori how mind and matter "are possible" - i.e. that they exist.[2] This is reasonable enough: to show that would

1. Ibid. Obviously, for there to be experience at all there must be a mind of some sort.
2. The idea is that only if x actually exists can x be said to be completely possible, or possible in all respects: cf. G. Schneeberger, Kants Konzeption der Modalbegriffe (Basel, 1952), pp.6-11.

presumably be to demonstrate a priori that the existence of matter and minds was basic to our own conceptual scheme, and this we cannot do; though how applying mathematics could help with this is highly obscure. Kant is confusing this claim with another one, less reasonable: that without the help of mathematics we can have no synthetic a priori knowledge about matter and minds; and he makes the confusion because this claim can also be put by saying that without mathematics we cannot know a priori how they "are possible". But surely one can do a lot more than show these concepts to be free from inconsistency, without invoking mathematics. One could show it to be consistent with the general conditions of experience set up in KdrV that there should be matter and minds of the required sort, and so prove that at least no general transcendental deduction can be given of the inapplicability of these concepts. And one could examine the conditions for the sort of experience in which there can be knowledge of matter and minds of the required sort: one would be investigating the properties of a particular conceptual scheme, which happened to be our own. Kant normally regards non-mathematical arguments about the conditions for experience of the sort that ours happens to be as establishing synthetic a priori conclusions; were he not confused over the ambiguity of "are possible" he should allow that there can be synthetic a priori knowledge of a non-mathematical kind about matter and minds.[1] Indeed, most of what he tries to establish in MAdN has just this status, for he considers it synthetic a priori and, although it is supposed to show how the concept of matter lends itself to mathematical treatment, it does not itself provide that treatment. The muddle here is interesting, for it confirms

1. See further below, pp. 229ff.

what we might otherwise have reason to suspect, that he was not clear in his own mind what he was really doing in MAdN.

He has another and better reason for despising psychology: it is not an exact science, and he does not think it can be made into one.[1] He says the same - perhaps a little unfortunately - about chemistry.[2] In the section on the Paralogisms in A, as well as in the MAdN, he ascribes the lack of possible synthetic a priori knowledge about the mind to the absence of regularity among the phenomena of inner sense;[3] which does give him a reason for thinking that much less can be accomplished by examining the concept of a mind than by examining that of a material body.

The difference between a work in immanent metaphysics and the corresponding Critique is that the former involves the analysis of empirical concepts whereas the Critiques supposedly do not.[4] The concept of matter is said to be an empirical concept.[5] It is not entirely clear what is meant by calling a concept empirical rather than a priori or pure; but one thing seems to be that it cannot be shown a priori (by a transcendental argument) that there exists anything which falls under the concept.[6] However, metaphysics does not include the analysis of all empirical

1. Ak. IV: 471.
2. Ak. IV: 470f.
3. A 381f.; Ak. IV: 472.
4. A 845/B 873; Ak. IV: 295 (Prol. sect. 15); Ak. IV: 469f.; IV: 482; V: 181; XX: 285.
5. Ak. IV: 470; and cf. A 848/B 876 (quoted below, pp. 216f.). What is important about matter in MAdN is that it is moveable: and in KdrV we find Kant saying that the moveable in space is found only through experience (A 41/B 58, B 155n.), and that motus is an empirical concept (A 81/B 107). Cf. also Ak. IV: 482; Ak. XX: 272; and Ak. IV: 295 (Prol. sect. 15).
6. See below, pp. 213 ff.

concepts. The analysis of the concept of air, for instance, is said not to be a matter for metaphysics[1], and this is apparently because Kant thinks one can produce no synthetic a priori propositions about air (other than those one can produce about matter in general).

For he defines metaphysics as a system of synthetic a priori propositions setting out relations between concepts and not involving any intuition, pure or empirical.[2] It makes use of analytic propositions, of course, but they subserve the aim of discovering synthetic a priori truths. Metaphysics is thus different from mathematics, which also provides a system of synthetic a priori truths but does it by constructing concepts in pure intuition.[3]

Now, since properly metaphysical propositions are synthetic a priori, and since synthetic propositions can be shown to be true only by some reference to intuition as the "third thing" connecting subject and predicate[4], any valid metaphysics must rely on arguments which, though they advert to no particular intuition, consider the general conditions under which experience involving intuition is possible.[5] Perhaps Kant does not always keep this point clearly in sight[6], but this is the only way in which the synthetic a priori propositions of immanent metaphysics could be established. So the conclusions of MAdN, as well

1. Ak. IV: 273 (Prol. sect. 4 - properly sect. 2, as in P.G. Lucas' tr.).
2. B 18; A 837-42/B 865-70; A 844f./B 872f.; Ak. IV: 273f.
3. A 713ff./B 741ff.
4. A 155/B 194; A 721/B 749. This does not apply to practical propositions.
5. A 720ff./B 748ff.
6. E.g. at Ak. IV: 470.

as those of KdrV, must rest on arguments of a generally transcendental character, the difference being only that in the MAdN we have to deal, somehow, with an empirical concept, and that apparently without prejudice to the apriority of the conclusions. In order to understand this, we shall have to see what Kant means by saying that the concept of matter, with which the MAdN is concerned, is an empirical concept.[1]

At first sight, there are two things he could mean by calling it empirical. He might mean that we acquire the concept by abstraction from experience, as he thinks we acquire the concept of redness[2]; or he might mean that whether we acquire it by abstraction or not we still have to appeal to experience to find out whether it has any application. This second possibility subdivides into two. (a) It may be that despite the logical possibility that the concept should not have application it is transcendentally necessary that it should: it would be impossible for there to be any experience at all if it did not. Here the only need to appeal to experience would be in order to assure ourselves that there was any experience to appeal to. If the concept of matter is empirical only in this sense there is no difficulty as to why Kant should call certain synthetic propositions involving it a priori: they would have the same status as the conclusions of the Critique. On the other hand, the concept of matter will then have just the same status as the categories, and yet he calls

1. It may be that the empirical concept of human nature is intended to play the same part in the Metaphysik der Sitten (Ak. VI: 217). But I shall consider only the MAdN, partly because I am in general concerned more with the theoretical than with the practical works, but principally because (i) Kant is not very self-conscious about his method in the Metaphysik der Sitten and (ii) it was finished late, contemporary with some of the Opus Postumum, by which time Kant may have changed his mind about, or simply forgotten, what his procedure had been in the MAdN. Cf. Adickes, Kants Opus Postumum, (Berlin,

the latter "pure concepts of the understanding" and clearly thinks that immanent metaphysics has some extra empirical element, not present in the Critique. (b) the other way of taking it is to suppose that an empirical concept is one which (at least) cannot be shown by any transcendental argument to have to apply to experience. If the MAdN's concept of matter is empirical in this sense, experience in which it had no application would be perfectly possible; though as a matter of fact our experience is not like that. This is apparently the way Kant himself says the expression "empirical concept" is to be used, at A 220ff./B 267ff.; not that that proves this is also how he uses it in the MAdN.[1] Against this interpretation is the consideration that it seems, unless it is further qualified, to assimilate the MAdN too closely to the Analytic of Principles, since Kant thought that experience based on some form of intuition other than space or time was perfectly possible (though not available for human beings). And he never says that the procedure of the MAdN

1. Though at the same time he says that empirical concepts are derived from experience; and elsewhere in the Critique he certainly considers that to be their defining property. Thus A 727/B 755, and cf. A 50/B 74; also A 47/B 64, where he even adds that for this reason they can give rise to no synthetic a priori propositions. This is also his position in the Logik, sects. 3 and 4 (Ak. IX: 92f.). Given his belief that whatever is transcendentally necessary is imposed by the mind on the world, it is natural for him to think that the class of concepts not acquired by abstraction includes the class of concepts the objective validity of which can be shown transcendentally. It includes other concepts as well, presumably, since space and time are imposed by the mind on the world without being transcendentally necessary. I think this is the right way to interpret Kant: it is a necessary, but not sufficient, condition of a concept's being empirical that its objective validity can't be shown a priori. But the fact that this interpretation makes what he says about empirical concepts (fairly) consistent does not of itself guarantee that those who claim he used the expression differently in the MAdN are wrong. There are few terms, if any, which he always used in the same way; which is what makes him so difficult to understand.

is essentially the same as that of the Analytic of Principles. On the contrary, the fact that he includes the latter in KdrV and describes the former as a branch of a more specialized metaphysics suggests that he thought of them quite differently. Moreover, it is clear that he considers space and time to be a priori, and not empirical.

The commentators are in disagreement; and the question is an important one, for it is a matter of the status of specialized metaphysics (or rational physiology, as Kant sometimes calls it) in relation to the transcendental philosophy of KdrV. Adickes[1] argues that the existence of matter as the movable in space is brought in a posteriori at the start of MAdN; accordingly he criticizes Kant for supposing he could establish synthetic a priori propositions on the basis of this assumption. On the other hand Vuillemin[2] and Plaass[3] have recently argued in favour of something like alternative (a), holding that there could for Kant be no spatial experience without matter; so that the task of MAdN is to continue the argument of KdrV in order to reach some more specific conclusions, without adding any extra premisses.

According to Plaass, when Kant calls the concept of matter empirical he does indeed mean that in order to know that matter exists we must appeal to experience. But this is only because there might be no matter and consequently no outer experience. He thinks it necessarily true that there could be no spatial experience without movable matter, but Plaass does not think this constitutes an a priori proof of the objective validity of the concept of matter; a view he can only hold because he puts very strong conditions on what is to count as an a priori proof of the objective validity of a concept (i.e., an a priori proof that a concept is

1. E. Adickes, Kant als Naturforscher (Berlin, 1924), I 247-71.
2. J. Vuillemin, Physique et Métaphysique Kantiennes (Paris, 1955), ch.1.
3. P. Plaass, Kants Theorie der Naturwissenschaft (Göttingen, 1960), esp. chs. 4 and 5.

instantiated) - much stronger conditions than Kant's. He allows that the transcendental deduction of the categories would (if valid) constitute such a proof, but regards it as essential that it (unlike the rest of the Critique) applies to all experience, and not just experience of a restricted kind (spatio-temporal experience): and Kant did not think it can be shown that the concept of matter must apply even in non-spatio-temporal experience.[1] - But surely Kant holds that to show that a particular concept must have application if spatio-temporal experience is to be possible is to prove a priori the objective validity of that concept; otherwise he would have to admit that the arguments of the Analytic of Principles contain an empirical element. So presumably Kant would think that to show that unless the concept of matter had application there could be no spatial experience just would be to give an a priori proof of the objective validity of the concept of matter.

Vuillemin appeals to a puzzling passage in KdrV in which Kant distinguishes the mathematical principles of pure understanding from the dynamical ones on the grounds that the latter, being concerned not just with intuition but with the existence of appearances, are not unconditionally necessary, but only necessary "unter der Bedingung des empirischen Denkens in einer Erfahrung".[2] He compares this with Kant's attempt in KdrV to explain what he is going to do in MAdN:

> ... wir nehmen aus der Erfahrung nichts weiter, als was nötig ist, uns ein Objekt, teils des äusseren, teils des inneren Sinnes zu geben.

1. Plaass, op.cit., pp.58, 84ff., 98f.
2. A 160/B 199; but A 221f./B 268f. seems to go against this. Vuillemin, op.cit., p.23.

> Jenes geschieht durch den blossen Begriff Materie (undurchdringliche leblose Ausdehnung).... Übrigens müssten wir in der ganzen Metaphysik dieser Gegenstände, uns aller empirischen Prinzipien gänzlich enthalten, die über den Begriff noch irgendeine Erfahrung hinzusetzen möchten, um etwas über diese Gegenstände daraus zu urteilen. (A 848/B 876).

He thinks Kant's idea is that for the arguments for the dynamical principles, and the arguments of MAdN, to go through it must be postulated that we have sensible experience: that there is at least "une sorte de minimum sensibile donné".[1]

In the first place it is intensely difficult to see how anyone could think there was a greater need to have recourse to something given in the case of the dynamical principles than in the case of the mathematical ones. Each are supposed to apply to any possible spatio-temporal experience; the man with pure intuition of space and time but no empirical data could be just as well aware that if he did have any experience the dynamical principles would apply to it as he could that the mathematical principles would apply to it. In the second place, even if Kant did think there was some distinction between the two sorts of principles in this respect, Vuillemin's assimilation of the dynamical ones to the findings of the MAdN is still over-hasty. For Kant thinks that some extra empirical element, not required for the Critique, is needed for MAdN.[2] Vuillemin agrees there is a difference between the uses made of the given element in the two works: the notion has to be made more precise in the latter.[3] But he thinks this difference does not affect the status of the arguments, because like

1. Ibid., p.24.
2. A 848/B 876 etc. (See fn. 4, p.211)
3. Op.cit., pp.24, 39.

Plaass he considers it to be Kant's view that if there is to be spatio-temporal experience at all there must be experience of movable matter. Making the concept of matter more precise involves no new appeal to experience, since the properties assigned to matter in the definitions are properties which must characterize anything that can be an object of outer sense. Experience plays just the same role as in the Critique: it assures us that there are objects to be experienced.

So neither Vuillemin nor Plaass gives an adequate account of what Kant means by saying that the MAdN involves a special empirical element, the concept of matter. Their mistake was to adopt alternative (a) abcve: to suppose Kant held that any spatio-temporal experience must contain matter as defined. Had Kant held this he would surely have argued for it; but he does not, at least until the Opus Postumum - and from the fact that he held a view there one can't infer that he held it in the Critical period. Nor can one argue that Kant must have held this position because it is so obviously correct: it is not obviously correct. I shall now defend these assertions.

The four sections of MAdN are each introduced by a new definition of matter. For the Phoronomy, matter is the movable in space[1]; for the Dynamics, the movable so far as it fills a space[2]; for the Mechanics, the movable so far as it has, qua movable, moving force[3]; for the Phenomenology, the movable so far as it can be, qua movable, an object of possible experience.[4] Of the properties here

1. Ak. IV: 480.
2. Ak. IV: 496.
3. Ak. IV: 536.
4. Ak. IV: 554.

mentioned the Analytic of Principles seems, at least at first sight, to commit us only to *being in space* and *being an object of possible experience*. The key idea here is that matter is not only in space but movable in space: the definitions of the later sections all presuppose that of the Phoronomy, and it is accordingly that one on which I shall concentrate. For if Kant did not even think things in space must be movable he cannot have thought they must have properties which depend on their being movable.

Two arguments are sometimes cited to show that Kant held it to be transcendentally necessary that the objects of outer sense must be movable - either because it is claimed that these arguments are to be found in Kant, or because it is claimed that they are so obviously valid that Kant could not have failed to draw their conclusions, or both. The first of these is an argument Kant does not give, but it is extracted from what he does say in the Refutation of Idealism. So if it belongs to Kant at all there is no doubt that it belongs to his Critical period. The other, which is more explicit, is the argument from the *Opus Postumum*: it has been claimed that this is also to be found in MAdN.

If the argument of the Refutation of Idealism were valid, it would show that one must be aware of some permanent, objective item in space.[1] Let us grant that it is valid. The argument which some commentators extract from Kant goes on to point out a corollary: if one were aware of nothing in space that was *not* permanent one could still not have the idea of an objective time-series. If the external world were changeless, how could one place

1. B 274ff.

one's experiences in an objective time-order?[1] - But this does not get us to the required conclusion: for not all change is local change. The supposed aim is to show that spatial objects must be the sort of things that move, but this argument only supports the conclusion that they must be the sort of things that change. One can imagine a world in which there are coloured objects in space, including the bodies of various percipients; these objects change in colour, but never move. Perhaps the percipients even communicate by using subtle and complex colour changes. They may be able to act on one another and on things, though this action will not involve physical contact but causal correlations between colour-changes in the agent and colour-changes in that which is acted upon; so they can apply the concepts of action and change, but not that of motion. This may ultimately turn out to be incoherent, but I don't think the incoherence is immediately obvious and it is not revealed by the argument we are discussing.

There is no evidence that Kant wanted to use the above invalid argument.[2] But in the Opus Postumum he tries to give a transcendental deduction of the concept of aether. He argues that there must be moving forces everywhere in space, since otherwise we could never have unified

1. E.g. K. Hübner, Das transzendentale Subjekt als Teil der Natur (Diss. Kiel, 1951), p.33. In granting the argument of the Refutation of Idealism we have granted that an objective time-order must be inter-subjective, and not just one about which mistakes are possible.

2. It is true that Kant uses "matter" as the generic term for objects of outer sense at B 278 (for example); but given his distrust of technical terms there is no reason to suppose he conflates this usage with that of MAdN.

experience of things in space; for my experience of such a thing must be caused (in virtue of the Second Analogy), and it can be caused only by the action on me of some moving force.[1]

It is true that given the organs we in fact possess we could not perceive objects in space unless things moved, but it is not clear that there could not be a percipient whose sense-organs could be acted upon by something analogous to colour-changes, without this action being effected by particles in motion. Causal generalisations could be established by observing correlations between the behaviour of certain items and that of certain other items (sense-organs).[2] Kant's argument will only do if we suppose that the only way in which events can be caused is by the motion, or moving force, exerted by things.

It may be said that Kant also used this invalid argument in MAdN. For in the preface to that work he says,

> Die Grundbestimmung eines Etwas, das ein Gegenstand äusserer Sinne sein soll, musste Bewegung sein; denn dadurch allein können diese Sinne affiziert werden. Auf diese führt auch der Verstand alle übrige Prädikate der Materie, die zu ihrer Natur gehören, zurück, und so ist die Naturwissenschaft durchgängig eine entweder reine oder angewandte Bewegungslehre. (Ak. IV: 476).

One can also quote B 291f., and no doubt other passages in KdrV in which it seems to be assumed that we have to think

1. This argument appears in various forms. Cf. for example Ak. XXI: 216, XXI: 572ff., XXII: 535, XXII: 609-615.
2. This would involve action at a distance, but one would not expect the author of the Monadologia Physica and MAdN to cast doubt on the intelligibility of that. Adickes (Kants Opus Postumum, Berlin, 1920, pp.415ff.) claims that he does in the XIth Konvolut, at Ak. XXII: 425-38 and XXII: 514-38, but I think Kant is here only ruling out attraction over empty space because he thinks empty space can never be perceived - which is just the MAdN argument for a plenum (Ak. IV: 523-535).

of things in space as movable. But Kant often also speaks as though all intuition must be spatio-temporal, though this is certainly not his considered opinion. It is tempting and convenient to make a point by considering only the familiar case of the sort of knowledge we have, even though the point itself may be more general. Thus in B 291f. his real point is that if we are to see what an alteration is we must appeal to outer intuition. He says indeed that we must take movement in space as our example; but in default of any argument there is no reason to suppose he means more than that it is the example we human beings rely upon, and not that there could not be beings, with spatial intuition and no idea of motion, for whom some other sort of alteration in spatial objects would fulfil the same purpose. To <u>us</u> movement in space happens to be the most obvious example of alteration in spatial objects. And the same sort of consideration can be invoked in dealing with the MAdN passage: it is quite reasonable to read this as saying only what must be so for <u>us</u>, because of the way our sense-organs are constituted. Indeed, in support of this reading one can point out that Kant does not actually say an object of outer <u>sense</u> must be movable, but that an object of outer <u>senses</u> must[1], which suggests he is thinking of sense-organs. And if Kant <u>did</u> mean here to commit himself to the view that there could be no self-conscious being who experienced objects of outer sense without having the concept of movement, he is contradicting what he has just said about the empirical status of the concept of matter as the movable in space.

1. Cf. also Ak. IV: 543.

It is true that by the time he came to write the Opus Postumum he had come to think that the conclusions of the MAdN were as thoroughly transcendental as those of KdrV. It is for this reason that he is so sharply faced with the need for an Übergang from the supposedly quite abstract MAdN to the empirical propositions of ordinary physics. But at some stage he came to accept the bad argument we have been discussing, and it is not implausible to suppose that even before this he simply forgot what his position had been when he wrote MAdN - perhaps half-remembering the passage quoted above from Ak. IV: 476 he came to think of it in a way different from that he had originally intended. After all even in the Critical period - before he began to lose his grip - he forgot what he had meant by "transcendental analytic" in KdrV.[1]

Probably the main reason people think Kant held, in MAdN, that to have the idea of an object in space one must think of it as movable, is that they regard this proposition as obviously true. This is at best a shaky way of proceeding in the history of philosophy. But though it may be true, it is not easy to prove. Not that it is easy to disprove, either: that would be to show that a picture like the one I have given, of a world in which there is only qualitative and not local change, contains no latent inconsistency. In general, as we noticed above, it is hard to show that something is not a necessary condition for knowledge or experience of a certain sort.

I have already considered and rejected two arguments for the proposition, one allegedly from KdrV and one from the Opus Postumum. Another argument worth considering may be found in Strawson's Bounds of Sense.[2] Strawson says that

1. L.W. Beck, Commentary on Kant's Critique of Practical Reason (Chicago, Phoenix ed. 1963), pp.55f.
2. P.128.

any self-conscious being must experience things in a spatial order (or in an order importantly analogous to a spatial order), because to have the idea of something which can exist unperceived is to think of it as capable of <u>being somewhere else</u> than in one's perceptual field. I discussed this argument above[1]; but one might apply it here by saying that a world lacking motion would have nowhere that objects could <u>go</u> to when existent but not being perceived by me, and nowhere that I could go to to avoid perceiving static objects.

This difficulty can be surmounted by introducing into the picture of a static spatial world another feature: intensity: Items can now alter not only in (something like) colour but also in (something like) brightness. At one time there is visible a round yellow light-source with a black dot in the middle; gradually the yellow circle grows brighter, but the black dot does not, and before long the yellow has become so bright that the dot is no longer visible. Then the yellow becomes less bright again, and once more the dot is visible. - This experiment may of course be performed in our own world, by looking at a light-bulb end on and using a dimmer. - One can imagine, also, that at a certain stage some percipients can see the dot, others not. If this sort of phenomenon occurred pretty regularly, would it not be reasonable to think of the dot as continuing to exist all the time, even when it was not perceived? The grounds for saying this are apparently of the same sort as the grounds we in fact have for speaking of unperceived existents.

One might, however, object that this imagined world is not really <u>spatial</u>, and so is not a counterexample to the thesis that a spatial world must contain movable matter.

1. Pp. 95ff.

The percipients in the imagined world would think quite differently from us about distance, lacking operational definitions; they would not make the same use of position in reidentifying objects as we do. The notion of tracking an object along a spatio-temporal path would not be necessary for them, since in their world objects would have temporal paths, but single spatial positions. What is necessary for an arrangement of coexistents to count as spatial? Any parameter could be accounted a dimension, and "motion" could be represented as change from one position on it to another. And any change, apart from creation and annihilation, could be represented as change of position on some parameter. So we could say that in any world of which self-conscious beings could be aware there must be room for "motion", since there must be room for change (if only in one's mental states). If this conclusion is to be resisted, a difference must be found between spatial dimensions and other parameters.[1]

I think there might be a being who perceived a world as so arranged that we, if we could see it as well, should describe it as spatial, yet who lacked any idea of motion. This suggestion unpacks into two. In one case, the world can be our own familiar spatial world, except that all motion has stopped. We should then be dead, but is it logically impossible that there should be some other percipient who was not? He would be aware of a number of coexistent items which we, if we were still alive, would recognize to be spatially related; he would not think of space quite as we do, but there is a sense in which he could be said to perceive things as spatially related. In the other case, which would be more congenial to Kant, the percipient not only receives similar data to us, he also

1. I am indebted to Mr. W. Newton-Smith for the elaboration of this objection.

orders them similarly: he uses an order which has a great many of the properties of space - and not just abstract mathematical properties, but phenomenological ones. I don't want to press this second version of the suggestion; there are notorious difficulties about the intelligibility of phenomenological comparisons. (Not that these would have worried Kant - cf. A 353f. - and it is the second version rather than the first that would have seemed to him to provide the counterexample to the thesis that all spatial experience must be of movable items). If either version makes sense, that will do: we can distinguish the thesis that things must be subject to change (and so to "motion" in some parameter) from the thesis that spatial objects must be movable in space. And it has not been shown that we cannot assert the former while refusing to assert the latter.

It is perhaps arguable that in either version the static spatial world would have to be thought of as only two-dimensional, on the grounds that we could not distinguish small objects from far-away objects unless we could move. This may be true of us as our senses are at present arranged, but it is not obvious that a different sort of percipient might not be able to make the distinction directly - if it is true of us it appears to be an empirical truth. But even if the imagined percipient does have to think of his world as two-dimensional there will still be point in calling it spatial if there is reason to speak of the two remaining dimensions as spatial ones.

If this is right, Kant should not be presumed to have held that one must think of spatial things as movable just because one obviously must. And I suggest that there is no reason at all, either in Kant's works or outside them, to suppose that when he called the concept of matter empirical he did not mean at least to exclude the possibility

of proving that any self-conscious being with spatio-temporal experience must have the idea of movable matter.

There remains the question of whether he considers that we acquire the concept of movable matter by abstraction from experience, or whether experience comes in only to tell us that it has application to objects, the concept itself being produced somehow from the mind's own resources. On this he does not give us much guidance. But whichever interpretation we take the same difficulty arises: how can an argument have a synthetic a priori conclusion if it has an empirical premise? For the aim was to prove synthetic a priori propositions about the objects of experience. A priori propositions are supposed to be apodictically certain, necessary; but how can truths about experience be necessary if they depend on a proposition which is admittedly a posteriori and not necessary, that experience contains movable matter?

Adickes goes even further, and criticizes Kant for attempting the analysis of an empirical concept at all.[1] For empirical concepts cannot be analysed, as Kant recognizes at A 727f./B 755f. - This is wrong on both counts. If conceptual analysis is possible at all, empirical concepts can be analysed, and to analyse them is to produce analytic propositions. These analytic propositions are themselves, of course, a priori, independent of any experience; which is not surprising, for they are semantic propositions and do not characterize experience. Adickes is apparently confused on this rather obvious point. Kant is not. What he says at A 727f./B 755f. is that empirical concepts like those of gold and water cannot be defined:

1. Kant als Naturforscher I: 257ff.

> Denn, da wir an ihm nur einige Merkmale von einer gewissen Art Gegenstände der Sinne haben, so ist es niemals sicher, ob man unter dem Worte, der denselben Gegenstand bezeichnet, nicht einmal mehr, das andere Mal weniger Merkmale desselben denke.

His idea - unfortunately not much further developed - is that concepts like these are cluster-concepts: different, though overlapping, sets of criteria are used in applying the same concept on different occasions. Unlike some philosophers, he does not conclude that no analyses of these concepts can be given, but only that they will be incomplete and not very helpful. In this very passage he says that although an empirical concept cannot be defined it can be made explicit; and he says of the concept of water that one will not stop short at that, "was man bei dem Worte Wasser denkt", but proceed to experiments. In applying a cluster concept on any particular occasion I must use some particular set of criteria, so there will be analytic propositions about the relation between these criteria and the concept-as-I-use-it-on-this-occasion (was ich beindem Worte Wasser denke). No doubt they are not the most interesting things to say about the concept, but they can be said. And if some analytic propositions can be made even about the concept of water there is no reason to deny that very general scientific concepts, like that of matter as the movable in space, can be analysed. It is far from clear that such very general concepts should be treated as cluster concepts at all.

There remains the more serious objection to Kant, that we cannot have synthetic a priori knowledge about matter if the concept of matter is not a priori. We met a similar difficulty before, in considering the Schematism. Kant regards it as contingent, a fact about human beings, that our intuition is spatio-temporal: as with matter, it cannot be shown by a transcendental argument that the concepts of space and time must apply in every sort of (sensible) experience; yet he regards the conclusions of the Analytic of Principles as synthetic a priori. The two cases differ in one important

respect: the concepts of space and time are called a priori, so the problem there is not to see why Kant calls the conclusions reached through them a priori but rather to see why he calls the concepts themselves a priori. To understand this we must find some respect in which the concept of matter differs from the concept of space and time which might explain why it should be called empirical when they are considered not to be. It can only be a necessary, and not a sufficient, condition of a concept's being empirical that it should be impossible transcendentally to prove its objective validity. For a sufficient condition we shall have to turn to the genetic criterion: Kant's idea must be that whereas the concepts of space and time are not abstracted from experience but imposed by the mind the concept of matter is acquired by abstraction, not imposed. As I said above, he does often draw the distinction between empirical and a priori concepts in this way.[1] We can, of course, be sure that if a concept is empirical in the sense of being acquired by abstraction Kant would not have considered it susceptible of a transcendental deduction.[2] If this is not Kant's position it is hard to see why he should have called the concepts of space and time a priori and that of matter not, and hard to see what difference there can be between the procedure of the MAdN and that of the Analytic of Principles. But if it is his position, how can he suppose the synthetic propositions of MAdN to be a priori?

One source of difficulty lies in taking too strictly Kant's use of the term "a priori". When he introduces it he draws a distinction, which because of its obscurity is not usually made much of, between "a priori" knowledge and

1. Above, p.214n.
2. Though he is wrong to assume that if a concept's objective validity can be shown transcendentally it cannot be acquired by abstraction. The mistake is to suppose that there can be no features of the world which are transcendentally necessary without our being aware of them. See above, pp. 60ff. It may well be that the idea that any concepts can be acquired by abstraction is wrong, for other reasons (cf. Geach, Mental Acts(London,1957)sects.6-11); but this is not Kant's view.

"pure" knowledge.[1] Whatever is pure is a priori, but not vice versa. A priori knowledge which is not pure has "something empirical mixed with it", though being a priori it must of course be "entirely independent of all experience". The obscurity of this is greatly increased by the singularly inept choice of the proposition "Every alteration has a cause" as an example of a priori but impure knowledge.[2] We are here told that the concept of alteration "can only be derived from experience"; though elsewhere he explicitly calls it an a priori concept.[3] Setting the example aside, however, we can see that Kant is here saying that a synthetic proposition can be a priori despite involving an empirical concept.

Kant himself says very little about this. But something like his view can be defended, if we remember what his general account is of the synthetic a priori propositions of philosophy. The way to establish a synthetic a priori philosophical proposition is to show that it must be true if there is to be any experience at all, or at least any experience of a particular sort. If in specifying that particular sort we have to use concepts which are empirical in the sense of being acquired by abstraction, it would be reasonable to expect the conclusions to contain concepts empirical in the same sense; for if the kind of experience is not specifiable in a purely a priori manner this can only be (for Kant) because it is not the only kind available to us. He expresses something of this, though rather obscurely, by his distinction between the form and the content of experience. KdrV has to do with the form, MAdN with the content[4]; the form is a priori, the content "can

1. B.3. The distinction is not made in A, which suggests he may have become clearer on the matter about the time he wrote the MAdN. Cf. also Ak. VIII: 183f.
2. B 3.
3. A 82/B 108 (for example).
4. Ak. IV: 467; IV: 387 (Grundlegung).

never be given in any determinate fashion otherwise than empirically".[1] No alternative to spatio-temporal experience is available to us, because we impose the forms of space and time on the world; and we have seen that Kant infers that any conditions of the possibility of spatio-temporal experience will also have to be imposed by the mind. The argument he uses to show that space and time are a priori is that we cannot think them away; and when what is a priori is said to be universal and necessary, this is what is meant. It is only where alternative kinds of experience are available, so that none of them can be imposed by the mind in advance, that he can allow the necessary conditions of a particular kind to be other than mind-imposed. Our experience may be of a kind to which alternatives are available - not just logically possible, but available to us in a way any alternative to spatio-temporal experience would (if Kant is right) not be; and since any proposition can be called synthetic a priori if its truth is a condition of the possibility of experience of the sort to which our conceptual scheme[2] belongs, these conditions can be synthetic and a priori despite involving empirical concepts. - Here we can see Kant's conception of the synthetic a priori splitting apart: he generally confuses the class of synthetic propositions imposed by the mind on experience with the class of synthetic propositions which have to be true for experience of the same general kind as our own to be possible, and it is this that leads him into transcendental idealism[3]; but here, in the

1. A 723/B 751 (Kemp Smith tr.); cf. A 20/B 34.
2. See above, p.187.
3. See above, pp.60ff.

MAdN, we have a class of propositions a priori in the second way without being mind-imposed. It is a pity Kant did not reflect further on this possibility.

The difference between the transcendental arguments of the Critique and the metaphysical arguments of the MAdN thus seems to be simply that a transcendental argument sets forth conditions for experience being of K-kind, where 'K' is the sort of concept that the mind imposes (like the concept of space), while a metaphysical argument also sets forth conditions for experience being of K-kind, but 'K' is now the sort of concept that is acquired by abstraction (like the concept of movable matter). And metaphysical arguments, like transcendental ones, will proceed by analysis of the concept of K-experience, so that the nerve of any such argument will be an analytic proposition. If this is all that the difference amounts to, it seems reasonable to say that metaphysical arguments can also and in the same way establish synthetic propositions which are a priori[1], provided that by "a priori" one does not mean "mind-imposed", when an empirical concept of the right sort appears as a value for 'K'.

"Of the right sort" because, as we saw when discussing scepticism, not just any limitation on the sort of knowledge or experience, the necessary conditions of which are to be investigated, will do if the interest of the exercise is to be retained. Our experience must be of that sort, and for the sceptic to be properly confuted it should be a sufficiently fundamental characterisation of our conceptual scheme for any alternative to be closed to us. Or failing that - and if Kant is right other alternatives will be open - it must

1. Though to be consistent with the usage of B 3 he should not call them "pure", as he does at Ak. IV: 468f. Cf. Ak. IV: 295 (Prol. sect. 15).

be the simplest for us of all the available alternatives; for choices between conceptual schemes can only be made on pragmatic grounds. In that case, we must hope that different ways of looking at the world, which would make others among the alternatives seem simpler than our own, are not themselves available to us; for if they are, we really do have no redress against the sceptic who takes his stand on one of the alternative conceptual schemes, save to say that our dispute with him must be irresoluble.

The question what sorts of experience are open to human beings is not very tractable. Kant's argument that we can't avoid thinking in spatio-temporal terms is that no alternatives are imaginable[1], but ultimately the claim is an empirical one about whether we, or beings like us, could learn to deal with experiential data ordered in very strange ways. The concept of space is fundamental to our way of looking at the world; but the concept of movable matter seems almost as fundamental, and one might well want to maintain, against Kant, that we human beings could no more deal with a world containing no movable objects than we could deal with a world of Strawsonian auditory particulars. To imagine, as we did above, a spatial world of static objects requires the same sort of imaginative effort as to imagine a world which is not spatial at all, like Strawson's auditory one or a realm of purely private feelings. In such worldsparticulars could be identified and talked about, but the conditions under which this was done would have to be very different from those we are accustomed to. So perhaps it would be correct to say that here again we have a starting-point for anti-sceptical arguments which have just the same character as those of the Analytic of Principles, and that

1. A 23f./B 38, A 30f./B 46.

Kant is wrong to draw the distinction he does. On the other hand, it is also possible that Kant is right, and spatial experience is unavoidable for us but experience of movable matter not; or that alternatives to both are possible for human beings. This is a question of fact, but one which need not bother us since there is no reason to think alternative conceptual schemes simpler than our own are available to us; what is important is to notice where the question of fact comes in, for the confusion of questions of fact with questions resoluble a priori is prolific in philosophical misunderstanding.

What the MAdN provides is really a further schematism of the schematised categories - which is why it can be treated as starting from a further determination of our conceptual scheme, for we saw that conceptual schemes are to be identified (at least partly) by the ways they determine the general (determinable) conditions for there being any experience at all.[1] Just as the pure categories were schematised by considering their application in a spatio-temporal world, the schematised categories are further schematised by considering their application to a world of movable matter. Kant cannot express it in this way, of course, because of the confusion of his ideas about schematism. For, as we saw, he sees the imagination as bridging by means of schematism a gap between pure concepts and experience, which gap having been once closed by schematising the categories no further gap remains, to demand more schematism.

Inured as we are to finding Kant's transcendental arguments invalid, it comes as no surprise to find the attempted proofs in the MAdN entirely unconvincing. In part,

1. Above, p. 187.

perhaps, we should not look upon them as proofs but as amplifications of the definitions of matter which open each of the four sections of the work; Kant does not regard the MAdN as more than the groundwork for a metaphysics of nature, and metaphysics might achieve worthwhile results by analysing the concept of matter even if the MAdN does no more than present us with the concept. Kant certainly does intend his proofs to establish something - to establish Newtonian physics, in fact, amended by the abolition of corpuscles in favour of centres of force. But since he allows himself to precede his theorems by a large number of Erklärungen over and above the four definitions of matter, into which he can build as many facts as he likes under the pretext of defining empirical concepts, it would not be very remarkable if the theorems achieved their ends; especially since even where the Erklärungen fail him he is quite happy to present us with a barefaced postulate to help his proof along.[1] In fact they do not achieve their ends. The most interesting of the four parts is the Dynamik, in which the physical theory of the Monadologia Physica, according to which the world is built up out of centres of force, is redeveloped, with the difference that the centres of force are now thought of as mathematical points[2] instead of as little bodies of finitely small size. But even here the interest lies in the theory and not in the supposed proofs of it; it requires six Erklärungen[3], a postulate,

1. Ak. IV: 501.
2. So the notion of a centre of force can now only be a regulative idea.
3. Besides the initial one, which defines matter as the movable so far as it fills a space (Ak. IV: 496).

and quantities of invalid argument to get us from the obviously true consideration that bodies fill a space by keeping other things out of it, to such far from obviously true (and in fact in this case false) conclusions as that the repulsive force exerted by a point-particle varies inversely as the cube of the distance.[1] It is in the next section, the Mechanik, that Newton's three laws are deduced; the argument here depends explicitly on the Analogies of Experience, but even granting these it is hopelessly wrong. The law of inertia is proved as follows. Every event has a cause; matter, as the object of the outer senses, can only have external relations and can alter only through movement; so the cause of its movement cannot be internal to it, but must be something outside; so any body persists in its state of rest or motion unless forced out of it by some external cause. This is a pretty close paraphrase of his actual argument.[2]

It might be of more philosophical use not to analyse the concept of movable matter, as Kant professes to do, but to try to reach as a conclusion the result that experience must contain movable matter. And to show that experience must contain matter which is movable, fills space, and exerts moving force would be valuable, even though it is wrong to infer immediately ffom that that experience must be subject to the laws of Newtonian mechanics. But it is not easy to see from what more general, more obviously unavoidable, features of our conceptual scheme it can be deduced that we must make use of that concept of matter. If we cannot deduce

1. Ak. IV: 518ff.
2. Ak. IV: 543. For a detailed discussion of the arguments of MAdN see Adickes, Kant als Naturforscher I. 186-378.

it from anything it is reasonable enough to take it as a starting-point; but one would expect Kant to give for it something analogous to the metaphysical exposition of space and time - not in this case showing that experience without movable matter would be unimaginable, since that is not his view, but defending this conceptual scheme against sceptical alternatives by suggesting that it has pragmatic advantages. (Though one cannot do more than suggest until one gets some idea of what the genuine alternatives are). He is too unclear about what he is really doing to provide this.

At least we may conclude that Kant's method does not change when he turns from Criticism to Immanent Metaphysics. He still seeks to establish synthetic a priori conclusions by finding the analytically necessary conditions for knowledge or experience of a particular sort. But it is now a more specific sort than it was in the Critique, and must be specified by an empirical concept, one acquired by abstraction from experience; so that his conclusions, though still a priori inasmuch as they are necessary for our conceptual scheme, cannot be mind-imposed.

Metaphysical Expositions and Deductions

The metaphysical arguments I have been discussing are those of the Metaphysische Anfangsgründe der Naturwissenschaft. But we should not forget that there are also arguments in the Critique of Pure Reason which Kant calls metaphysical. At times, of course, he calls the whole thing metaphysics, for he sometimes uses this term in such a way as to include transcendental philosophy; and even when he is maintaining a distinction between metaphysics and the transcendental or critical philosophy of the three Critiques he considers as part of metaphysics those invalid arguments to illegitimate, transcendent conclusions which formed the substance of dogmatism. These I shall not consider, for he sets them up

only to destroy them; and the basis for this destruction is established by transcendental arguments, the character of which we have already discussed. But as well as these there are other arguments he does call valid and metaphysical: the metaphysical expositions of the concepts of space and time, and the metaphysical deduction of the categories.

Why these should be called metaphysical is not immediately clear. When in the Critique of Judgement he most explicitly contrasts metaphysical and transcendental principles, he says of the former that they "represent the condition a priori under which alone objects, the concept of which must be empirically given, can be further determined a priori"[1]; this describes what goes on in MAdN, but not in the metaphysical deduction or expositions. For he does not regard the concepts of (objects in) space and time or the categories as empirically given.

The terms "metaphysical exposition" and "metaphysical deduction" occur only in B; they are more used by commentators than by Kant. They are not part of his established terminology, as for example "transcendental deduction" is; so if the use of "metaphysical" in these contexts turns out to be a bit anomalous it does not greatly matter. "Metaphysical deduction" is in fact used only once, and that not in the course of the argument to which it refers but in the transcendental deduction; Kant says,

> In der metaphysischen Deduktion wurde der Ursprung der Kategorien a priori überhaupt durch ihre völlige Zusammentreffung mit den allgemeinen logischen Funktionen des Denkens dargetan.... (B 159).

1. Einleitung sect. V (Ak. V: 181).

If by "deduction" Kant means, as he says[1], "justification", why should that argument be called a deduction? It is important to the transcendental deduction that he should have established the connection between the pure concepts of the understanding and the logical functions of judgement; but establishing this does not constitute any sort of defence of the former, as the transcendental deduction purports to. What it does show is that the categories are not empirical concepts in the sense that they are not derived from experience, but provided a priori by the mind itself from its own resources.

But this is what a metaphysical exposition is also supposed to do:

> Ich verstehe aber unter Erörterung (expositio) die deutliche (wenn gleich nicht ausführliche) Vorstellung dessen, was zu einem Begriffe gehört; metaphysisch aber ist die Erörterung, wenn sie dasjenige enthält, was den Begriff, als a priori gegeben, darstellt. (B 38).

This makes it difficult to see where the difference between metaphysical expositions and metaphysical deductions can lie.[2] Perhaps Kant calls the metaphysical deduction of the categories a deduction because he regards it as justifying his table of categories, as against Aristotle's, by grounding it in the table of judgement-forms. Alternatively - and the alternatives need not be exclusive - he may so call it because it is so closely linked with the transcendental deduction. Or perhaps the answer lies in the emphasised "given a priori" in the definition just quoted - for Kant does hold the strange

1. A 84f./B 116f.
2. There may not be much difference: at Ak. VIII: 184 Kant says that "die Exposition der Kategorien" treats them "als bloss auf Objekte überhaupt angewandte logische Funktionen".

view that space and time (unlike the categories), though a priori, are somehow given.

But there is also a difference in the uses to which the two sorts of argument are put. As we saw in the last chapter, the metaphysical expositions are used to show that human experience has to be of a certain sort, as a matter of very general psychological fact, in order to justify Kant in going on to examine the necessary conditions of that sort of experience. The metaphysical deduction of the categories does not do this, for it is not supposed to be a mere psychological fact that there are only so many categories or logical functions of judgement.[1]

Why call these arguments metaphysical? Here again one can only guess. They are not transcendental; they investigate neither the necessary conditions for exprience, as transcendental deductions do, nor those for synthetic a priori knowledge of a particular kind assumed to be given, like transcendental expositions or the arguments of the Analytic of Principles. Yet they are of importance to the transcendental philosophy, and very likely Kant calls them metaphysical to indicate this, and to mark them off in this respect from other expositions, other pieces of conceptual analysis. Since he often uses "metaphysics" in such a way as to include in metaphysics the work of the Critiques, it would be natural for him to describe as metaphysical those conceptual analyses which belong to the latter but are not themselves transcendental arguments.

1. At Ak. XX: 272 he says that even the thought of an intellectus intuitus must be governed by the categories. And although he does not hold this in KdrV, he does hold that the experience of any other being must be.

Constitutive and Regulative Principles

The transcendental and metaphysical arguments I have considered are supposed to establish what Kant calls constitutive principles: synthetic *a priori* truths which impose conditions on the objects of experience, and which therefore help to constitute the nature of experience. But besides constitutive principles there are principles of another sort: regulative principles. Certain of the assertions of dogmatic metaphysics, which go beyond possible experience and so cannot be treated as true or false, are still of value as regulative principles. In the MAdN certain of the conclusions are regulative rather than constitutive; so presumably regulative principles can be established by metaphysical arguments. They are just as synthetic and *a priori* as constitutive principles, but are not strictly to be taken as stating propositions at all: rather as setting out rules for the understanding to follow.[1] The distinction is seen most clearly in the case of infinity: because no actual infinity can be given in experience, we cannot assert anything about it; but we can still make what use we need to of the notion of infinity, because we can conceive of a series such that the understanding is given the task of going on always to the next member.

The notion of a regulative principle gives rise to a lot of interesting questions. It enables Kant to construct what is in effect a formalist philosophy of mathematics (in his discussion of the Antinomies), of physics (in the MAdN), and of biology (in the *Critique of Judgement*), as well as to connect his speculative with his practical philosophy: for

1. A 508ff./B 536ff.; A 647/B 675.

principles of pure practical reason can only be regulative for pure speculative reason. However, I am concerned with Kant's method, and have interpreted this to mean: the sorts of argument he uses, and not the conclusions he uses them to reach; I shall therefore not concern myself with particular regulative principles further than to ask the question: how does Kant argue for or against them?

In mathematics and the sciences, one may feel, there is no need to argue for them except by showing how fruitful it is to adopt them. This was not Kant's view. He always argues for them, and his argument always is that reason cannot avoid adopting them: indeed, reason has an unfortunate tendency to treat them as though they were constitutive, and the best we can do is to be on our guard to recognize their regulative status. - This sounds like a psychological argument. There is always a psychological admixture, and frequently the argument is purely psychological. Thus when Kant argues in KdrV that because the categories belong to the understanding and space and time to intuition we have an inevitable tendency to regard the (pure) categories as capable of meaningful employment beyond the limits of possible experience, the assertion seems to be a psychological one: there would be no logical difficulty in conceiving a well-designed mind who felt no temptation at all to suppose the categories meaningful in their transcendent use. And if such a mind came nevertheless to use them beyond the limits of possible experience, but merely as regulative ideas, it would have to do so for some such reason as the scientific fruitfulness of doing this. - Admittedly Kant might contest the assertion that there is no logical difficulty in conceiving such a mind, on the grounds that although his argument was psychological it belonged to transcendental psychology, it being a necessary

condition of experience for intellectus non-intuitus that its faculties of understanding and sense should be separate. But granting that the faculties are separate, it seems a purely empirical matter that our faculty of understanding should not be content to remain within the bounds set for it by sense.

Arguments for regulative principles are not always as psychological as this - even if the psychology is transcendental. In the MAdN it is argued that perceptible matter is the product of the interaction of point-centres of force, and it is noticed that the idea of a point-centre of force can only be regulative, since such a thing could not be the object of any possible experience.[1] The argument for this is, apart from its conclusion, apparently the same as any other metaphysical argument: that this must be so if there is to be experience of matter (which is now being considered not merely as movable but as occupying space). It is certainly not merely a hypothesis which covers a lot of phenomena and suggests new laws; Kant does sometimes make hypotheses, but these hypotheses do not (or should not) involve regulative principles, at least so far as the hypotheses aare to be used to further speculative knowledge.

> Transzendentale Hypothesen des spekulativen Gebrauchs der Vernunft, und eine Freiheit, zu Ersetzung des Mangels an physischen Erklärungsgründen, sich allenfalls hyperphysischer zu bedienen, kann gar nicht gestattet werden....[2]

Why this cautiousness about hypotheses? It may be partly that Kant forgets that to employ the categories beyond the

1. Ak. IV: 505ff.
2. A 773/B 801; cf. K.d.U. sect. 90.

limits of possible experience is not necessarily to make an assertion about things in themselves: it may be for example to prescribe a certain procedure to the understanding in dealing with more and more spatio-temporal items, or finer and finer parts of spatio-temporal items - a regulative principle. But since, although regulative principles play a most important part in Kant's system, he never, so far as I know, puts one forward without arguing that it is in some way inevitable that reason should employ it, there is presumably more to it than oversight.

The problem with which Kant is faced is, how to formulate a scientific method so that theories shall be prevented from making claims which go beyond all possible experience to speculative metaphysics, while at the same time permitting them to go beyond experienceiin the way a comprehensive theory of natural phenomena must. And in Kant's time it was still not customary to draw a sharp distinction between science and metaphysics. The Leibnizians, who determined Kant's early approach to this question as to most others, sometimes presented speculative metaphysics as though it were a high-level scientific theory required by the experiential data; and Leibniz himself often gives this impression, particularly in his dynamically-orientated works, which were (mostly) available in Kant's time when his logical works were not. Kant guards himself against such a manoeuvre by making two moves.

The first is to introduce the notion of a regulative principle, which provides no knowledge but may for scientific purposes be treated as if it did. As soon as a theory goes beyond possible experience it enters the realm of regulative Ideas, and can make no further claim to truth-value. Hence if a speculative metaphysics could be established as a scientific theory its doctrines could not claim truth or falsity but only usefulness.

Kant apparently does not regard this as an adequate safeguard against the metaphysicians. They are still free to go on system-building provided they restrict their claims; we have the same jumble of systems as before, except that it is now recognized that their theoretical sentences are disguised imperatives of some sort. For (infinitely) many theories may fit the same empirical facts equally well and license the same predictions - Kant does not express the difficulty quite in this way, but it is what he has in mind. We get round it usually by appealing to some notion of a theory's simplicity, a notion in which there are well-known difficulties. Kant avoids it by making his second move. This requires the theory to be justified, not just by showing that it explains the facts, but by showing that the theory - the set of regulative principles - is somehow unavoidable. And this unavoidability can then be demonstrated, either by (transcendental?) psychology; or by showing, as with the point-centres of force, that their adoption offers the only way in which the phenomena (in this case, the existence of movable matter filling space) can be understood; or by a mixture of these methods.

But if the metaphysical argument used to justify talking about point-centres of force really is formally similar to a metaphysical argument used to establish some constitutive (i.e. non-regulative) principle, like perhaps the principle that all matter exerts attractive and repulsive force, why is the status of the conclusion so different in the two cases? The position is the same for transcendental arguments. For Kant considers that in synthesising my experiences into awareness of a world of objects I have to use the idea of a transcendental object, of something <u>entirely</u> independent of me - and which I could consequently never receive as a <u>Vorstellung</u>. And I must likewise make use of the idea of a transcendental subject in synthesising my experiences into a

single consciousness: i.e. I must think of myself as a unity, despite the fact that this unity could never be given in experience. So the ideas of a transcendental object and a transcendental subject are strictly regulative, despite the fact that I must both have them and make use of them in order to have any experience at all.

The answer to the question why the same sort of argument should sometimes yield constitutive principles, sometimes regulative ones, is of course basically that the conclusion can be a constitutive principle only if it lies itself within the limits of possible experience. If it goes beyond them, since it is not analytic not even transcendental arguments can assure us of its truth, much less metaphysical arguments. This is not because there is anything logically wrong with the notion of a metaphysical or transcendental argument yielding a true, and not merely regulative, principle which goes beyond the limits of possible experience; it is because the possibility is ruled out by what Strawson calls Kant's principle of significance, the thesis that the only synthetic propositions the truth-value of which we can assess concern matters which lie entirely within these limits. If it could really be shown, by analysing the concept of experience, that it was transcendentally necessary that there be things-in-themselves or the activity of synthesis, then it would be right to conclude that these things must exist, and not just that we must think they do. For, by the transcendental argument, unless these things exist there can be no experience, and our pretending they exist will not help in the slightest. But the principle of significance assures Kant (sometimes) that no such argument can be given.

He does not stick to this very rigorously. One can never actually be aware of synthesis going on, yet he never seems to face the consequence that the proposition that

synthesis occurs should be merely regulative. Indeed, perhaps all the propositions of Kant's transcendental psychology should really be regulative. At any rate, he was confused about the matter; his utterances about the status of the transcendental object and the transcendental subject are not usually plain or unambiguous. Nor, notoriously, does he invariably content himself with regarding the proposition that things-in-themselves exist and cause our sensations as merely regulative.[1]

He could have avoided this confusion had he distinguished clearly (1) transcendental or metaphysical arguments to the effect that p must be the case if experience (of the required sort) is to be possible, from (2) transcendental or metaphysical arguments to the effect that one must think that p - must assume it at least as regulative[2] - if experience (of the required sort) is to be possible, and both of these from (3) arguments to the effect that one must think: "p must be the case if experience (of the required sort) is to be possible"[2]. He seems at times undecided as to whether he wants to say (1) that it is transcendentally necessary that there should be things-in-themselves, (2) that it is transcendentally necessary that we should use the regulative idea of a transcendental object - i.e. that we must proceed as if there were things-in-themselves, or (3) that we must have the idea of a thing-in-itself, not because it is transcendentally necessary that we should, but because otherwise we should be unable to explain things - it provides the only explanation that our minds are capable of grasping (which may not stop it being a pretty bad explanation) of how there can be objective phenomena.

1. For lots of references, see Adickes, Kant und das Ding an Sich (Berlin, 1924).
2. The sophisticated philosopher does not have to believe it true, but he must act as though it were.

Arguments of type (1) are pretty straightforward: the ordinary metaphysical and transcendental arguments are of this type. Arguments of type (2) form a sub-class of those of type (1); they establish principles of the form "It is necessary that the percipient think that p". But where p is itself not a proposition susceptible of truth or falsity but rather a regulative principle, arguments of this type may be regarded as one way of giving the sort of vindication Kant requires of the unavoidability of holding that p: as with the ideas of the transcendental object and the transcendental subject. Arguments of type (3) can also be used to show the unavoidability of holding regulative principles, but they do it in a different way: they do not show them to be transcendentally (or metaphysically) necessary, only to be necessary assumptions for us to make if we are to understand phenomena. We think that the phenomena could only be as they are under these conditions; but we may be wrong. We may even be able to prove ourselves wrong, like Kant - he uses an argument of type (3) in the Critique of Teleological Judgement, saying that we must use the idea of finality in nature because we can see no other way in which the phenomena can be explained, while at the same time recognizing that KdrV has shown that they must be susceptible of a mechanistical explanation which does without the notion of finality.[1] But though we can show that our explanation is in fact not the right one, we are unable to do any better: we know what sort of explanation to look for, but we cannot actually produce it. So an argument of this type does not assert that we must take

1. K.d.U. sects. 64-5, and Einleitung sect. V. Kant does not think it possible that a mechanistical and a teleological explanation of the same phenomena could both be equally correct: K.d.U. sect. 78 (Ak. V: 411f.).

the truth of p to be a logically necessary condition for experience of any sort, only that we must find it the best explanation available to us of various facts. "Must", that is, as human beings; beings who saw things differently might be able to do better. No transcendental necessity comes in here at all, unless it can be shown to be transcendentally necessary that we should seek to explain these facts; it is just a feature of human beings that it is natural to them to look for explanations.

Whether the MAdN argument about point-centres of force is of type (2) or of type (3) is not entirely clear; but in neither case will its status now be particularly problematic. If it is of type (2), then it is a metaphysical argument of the same type as any other. If it is of type (3), what is being asserted is that we must think in terms of centres of force if we are to understand the behaviour of matter (though we might be able to experience matter without understanding it); and not because it provides the correct explanation, but because it provides the best explanation to which the human intellect can attain. Whether or not we, or Kant, want to extend the term "metaphysical argument" to cover arguments of this sort is a question of no great moment.

III. The Flight from Extension

I want to end by considering ~~a very~~ the most general criticism that may be brought against Kant's way of doing philosophy, whether he uses transcendental arguments, metaphysical arguments, or just ordinary conceptual analysis. We have seen the importance of the notion of analyticity for Kant: he argues by showing that something is an analytically necessary condition of there being any knowledge or experience (of some type).[1] And the very general criticism I want to consider is that the notion of analyticity does not make sense. I shall not discuss the different view that although it does make sense it is almost entirely useless in philosophy; what I have already said constitutes an argument that if it is intelligible it is of some use.

The attack on its intelligibility, which is led by Quine, proceeds by alleging the circularity of all attempts to define analyticity.[2] It is circular to define an analytic proposition as one whose contradictory is self-contradictory; to define it as a tautology involves the notion of synonymy, which has all the obscurity of analyticity and more as well; to define it as a proposition whose truth can be established by logic and definitions alone again involves the notion of synonymy, both in giving an account of what a definition is[3] and in making the point that the same substitution must be made for each occurrence of the same variable in the logical formula.[4] It will not do to say that an analytic proposition

1. On the analytic character of transcendental arguments see above, pp. 68-72.
2. W.V. Quine, "Two Dogmas of Empiricism", in From a Logical Point of View (Harvard, 2nd ed., 1961); and Word and Object (M.I.T., 1960). Cf. also N. Goodman, "On Likeness of Meaning"; B. Mates, "Synonymity"; and M. White, "The Analytic and the Synthetic"; all in L. Linsky (ed.), Semantics and the Philosophy of Language (Urbana, Ill., 1952).
3. Quine, "Two Dogmas", pp.24-7.
4. Strawson, "Propositions, Concepts, and Logical Truths", Philosophical Quarterly, 1957.

is one which is true in all possible worlds, since "possible" here means "logically possible", i.e. not involving a contradiction. Equally it will not do to make the approach through statement-synonymy[1] by saying that two statements are synonymous if they are verified and falsified by the same states of affairs under all possible conditions. For here "possible" again means "logically possible"; if it is logically possible, and yet not possible in any other sense, that A and B should occur separately, we don't want to have to say that the statements "A occurs" and "B occurs" are synonymous.[2]

Grice and Strawson reply to Quine that one need not be dismayed by the impossibility of defining analyticity without remaining inside a rather small circle of more or less ~~in~~ interdefinable terms.[3] Clearly this is not a defence that can be used quitegenerally; the terms of a metaphysical system may be interdefinable, but it does count against their intelligibility that one can never break outside the circle and relate them to other familiar expressions. The appropriateness of the refusal to be dismayed in the case of analyticity could be defended if it could be shown that the terms of its circle are indispensable: if, that is, they could be made the subject of a transcendental argument. It would not have to be shown that any language-user must have these ideas - clearly even the Sunday Times colour supplement leaves semantics a closed book to many undoubted language-users - but only that

1. Strictly we should talk of sentences and not of statements or propositions in giving a definition of synonymy, for the identity-conditions for statements and propositions involve synonymy.
2. This is what lies behind Quine's attack on this account: Word and Object, p.64.
3. Grice and Strawson, "In Defence of a Dogma", Philosophical Review, 1956.

these ideas (synonymy, analyticity, contradiction, etc.) must be employed in any account of how they use language. Transcendental arguments are analytic; but this need not import a circularity, for they need not make explicit use of the notion of analyticity: the point will be made against Quine if any account of language which does not use these terms can be shown to be impossible in some weaker sense than logically impossible. Moreover, it is conceded that the notions we are trying to defend cannot be reduced to terms acceptable to Quine; what matters is to provide a coherent defence and not to convince the extensionalist. Can this be done?

According to Quine it cannot, and he develops an account of language which does not involve any such notions.[1] In this account the meaning of an expression is a matter of the speaker's response to stimuli; semantics does not involve any reference to such nebulous entities as the intentions of speakers and linguistic rules (as we supposed above that it did[2]). There is thus no need for our ordinary idea of analyticity, though it is possible to make use of the tidied-up notions of stimulus-analyticity and stimulus-synonymy. A sentence is stimulus-analytic for a subject if, and only if, he would assent to it, if to anything[3], following any stimulation that might be imposed during the time under consideration. And two (general) terms F and G are stimulus-synonymous for a subject if, and only if, "All F's are G's" and "All G's are F's" are stimulus-analytic for him; which involves F and G having the same stimulus-meaning for him.[4]

1. Word and Object, esp. chs. I, II and VI.
2. Above, pp. 99ff.
3. He might be unable to produce any linguistic response to the stimulus.
4. Word and Object, pp.54f.

There are various difficulties with Quine's account of meaning. I shall not consider many of them. His account will do at best for a language which is extensional - roughly speaking, a language for which the meaning of a term can be represented as a function of the actual things to which it can be applied. Such a language has no room for intensional contexts - modal and indirect discourse contexts, for example, as they are ordinarily understood. Quine tries to provide an acceptably extensional way of handling these, whereby "is necessary" and "is believed-true by Tom" become predicates attachable to names of sentences[1]; but, as Geach and others point out[2], no such account of these contexts can be adequate if it treats the name of a sentence as nothing more than a designation of an inscription, and not as analysable in a manner which indicates its connection with what the inscription was used to mean - which in these cases would take us beyond the limits allowed to an extensional language. However, the inadequacies of what Quine has to say in its defence do not show that an extensional language is not perfectly possible; nor do they show that where natural languages allow us to depart from what is readily expressible in purely extensional terms they do not involve us in talking nonsense. There is a strong feeling, shared by many philosophers, that some such elimination of nonextensional expressions in favour of extensional ones must be possible, even though it means giving up talking about analyticity, and even though the details may be hard to work out. What lies behind this is partly the worry about what the

1. Word and Object ch. VI, and "Three Grades of Modal Involvement" in his Ways of Paradox (New York, 1966). He also suggests (Word and Object p.216) construing "believes-that-p" as a simple, unanalysable predicate attachable to names of individuals.

2. Geach, Mental Acts (London, 1957), sect. 18; cf. K. Reach, "The Name Relation and the Logical Antinomies", Journal of Symbolic Logic, 1938.

identity-conditions for intensional items could be, but more importantly a feeling that only what exists can be talked about: to postulate entities which do not exist, but only might (like logical possibilities, or things desired or intended), is at best ontologically untidy and at worse nonsense. If it could be shown that no extensional account of language will do it would be evident that these doubts are misplaced; it would not be clear just what the identity-conditions for intensional items are, but it would be clear that there must be such conditions, because it would have been shown that there must be intensional items.

It is true that it might be that although our language was extensional no adequate extensional account of semantics could be given. This was Wittgenstein's position in the Tractatus; we should then be totally unable to explain how it is that our sentences have the meanings they do, though none the less able to use them meaningfully. But to adopt this position one needs some very strong reason for claiming that our language really is extensional, despite both the impossibility of giving an extensional semantics and our strong belief that it is not impossible for us to work out adequate theories of meaning. Wittgenstein found this strong reason in Russell's logic; but that Russell's logic was adequate as the logic of ordinary discourse should never have been advanced as more than a hypothesis, and a proof that it relegated semantics to the realm of the mystical could well be regarded as counting against that hypothesis. Quine's strategy is different: he accepts extensionalism because he regards an extensional theory of meaning as both possible and correct. It cannot cope with non-extensional notions like analyticity, but that does not show that it fails to explain anything: it treats non-extensional notions as enshrining confusions which can be cleared up, and not as the indispensable core of any theory of meaning.

"Extensional"

The words "extensional" and "intensional" are used in various ways. I use "intensional" as synonymous with "non-extensional", and I use the terms to make three different, but parallel, distinctions. In one sense, a language or a logical system (and I shall use the term "system" to cover this disjunction) iscalled extensional if, and only if, it is truth-functional: i.e. if it employs no propositional operators other than truth-functional ones, and consequently no implication-relation stronger than material implication. In another and perhaps more popular sense a system is called extensional if, and only if, every occurrence of a name or definite description in it is, as Quine would say, "referentially transparent"[1]: any other name or definite description referring to the same thing can be substituted for it without affecting the truth-value of the whole sentence. And to preserve the balance between predicates and referring expressions we can add a third sense: calling a system extensional if, and only if, every occurrence of a predicative expression in it is such that any coextensive predicative expression can be substituted for it without affecting the truth-value of the whole sentence. Two expressions are called coextensive if and only if they each apply to exactly the same set of actually existing things - which we may assume, for simplicity's sake and to make extensionalism more plausible, to include things which actually have existed and actually will exist even though they do not exist at this moment, though some people would be hesitant particularly over future items. And an expression may be said to be taken in extension if, and only if, it is considered as applying only to items which actually exist.

The extensional/intensional contrast has been used to

1. Word and Object, pp.141 ff.

mark other related distinctions as well, but I shall not be concerned with them. In these other cases "intensional" is sometimes spelt with a t instead of an s, to show that it is intended to capture a different and more psychological idea: the attempt to formulate criteria for intentionality in this sense (or range of senses) is an attempt to mark off mental phenomena from others by pointing to the peculiarities of the objects of psychological verbs. The criterion most often formulated says that a non-compound declarative sentence is intentional if it uses a substantival expression in such a way that neither the sentence nor its contradictory implies that there is anything to which that expression applies[1]; clearly the circumstances which generate this sort of intentionality often produce a lack of referential transparency as well, and vice versa.

But let us leave this aside and return to the three sorts of extensionality distinguished above. Even these rather parallel definitions do not all mark off as extensional the same set of systems. It might be thought that they would; but so far from doing this, they apparently each draw the line in a different place. This has sometimes been denied, and I shall indicate my reasons for thinking it true, but I shall not here give the subject as much detail as it deserves; it is sufficient for my purpose in the present section just to point out that it is not obvious that the three definitions all give the same results.

1. This is Chisholm's first criterion, in "Sentences about Believing" (Proc. Arist. Soc. 1955-6) and in Perceiving (Cornell, 1957), p.170. It is, of course, inadequate by itself to mark off psychological verbs from others, being neither necessary (because of verbs like "knows") nor, as Chisholm thought, sufficient (because of cases like Urmson's "Blackcurrants require a generous supply of nitrogen"). Cf. the symposium between J.O. Urmson and L.J. Cohen, "Criteria of Intensionality", Proc. Arist. Soc. Supplement, 1968.

In order to avoid confusion, I shall call a system which is extensional in the first of the three senses truth-functional; one which is extensional in the second sense transparent; and one which is extensional in the third sense predicate-extensional. The only question I shall consider at any length at all concerning the relations between these is whether all transparent systems must be truth-functional, since Quine has argued that they must be and the argument looks rather suspect. But we can notice briefly that there is no obvious reason to expect that the converse should be true, i.e. that truth-functional systems should have to be transparent; "Nine is necessarily greater than seven" is not transparent, for nine is the number of the planets without the number of the planets being necessarily greater than seven, yet "Nine is necessarily greater than seven" contains no propositional connectives and one could build up a truth-functional system out of such propositions. Nor, it seems, do truth-functional systems have to be predicate-extensional; the two predicates "is the positive square root of 36" and "is necessarily the positive square root of 36" apply to precisely the same object, the number nine, and neither of them applies to anything else; yet they are not always exchangeable, for the same reason as before, for although the number of the planets is indeed the positive square root of 36 it is not necessarily so. And again one could build up a truth-functional system having simple propositions of this sort.

Transparent systems need not be predicate-extensional. An example of a system which was transparent but not predicate-extensional would be one in which modalities were construed differently from the way in which I have just been taking them, in that modal properties could be ascribed to things regardless of how those things were described. They would then ascribe essences, the modalities would be "de re"; as in "It is necessarily true of Socrates that he is a man", where

it remains just as necessarily true of him whether we call him Socrates or something quite different, and quantification into the modal context is allowed. Such systems are quite familiar; but two objections to them are brought by Quine[1] and others. One of these is that since these systems are transparent they must be truth-functional and therefore cannot be in any interesting sense modal; but I am going on to consider and reject the argument that any transparent system must be truth-functional; it is built into the semantics of these systems that expressions of the form "α is necessarily φ" entail expressions of the form "α is φ" but not vice versa, and so on, so that they are modal by Lukasiewicz's criteria.[2] The other objection, which is more cloudy, is that this position involves talking about essential properties, and there is no way of determining of an item (as opposed to an item under a description) which properties are essential to it. Now it is true that a lot of work still needs to be done on sortal concepts, but it does seem plausible to say that a sortal like "man" or "dog" denotes a property which is in some sense essential to men or dogs however they are described. Even if this is wrong as an account of the way we use sortals, what is at issue now is whether there might not be some system in which they were so used: and surely that is not hard to envisage. Such a system will not be predicate-extensional if there are values for F for which "is F" and "is necessarily F" are coextensive, for these predicates are not interchangeable in

1. Quine, "Reference and Modality", in From a Logical Point of View; Word and Object, sect.41.
2. As presented by Prior, in Time and Modality (Oxford, 1957), ch. 1, these are: CLpp, CpMp, EMpNLNp and ELpNMNp must all be laws; Mp and CMpp must be non-laws; Lp must be sometimes true and Mp sometimes false.

the one-way entailment statement "If x is necessarily F then x is F".[1] For example, it may in fact be true that all and only those who are men are necessarily-men, yet there might be a man who was so only accidentally (an incarnation of a Divinity, perhaps); and the system allows us to mark this possibility, as no predicate-extensional system could.

So it is not the case that all transparent systems must be predicate-extensional. The converse implication does, however, hold: all predicate-extensional systems are transparent. If there were a possibility parallel to the one above which predicate-extensional systems could allow, so as to avoid being transparent, it would be the possibility that the terms a and b, which in fact name the same thing, might not. But then "might in some possible world not be identical with b" is itself a predicate applying, among other things, to a; and the predicate "is identical with a and might in some possible world not be identical with b" applies uniquely to a. It is therefore presumably[2] coextensive with "is identical with b", and, since the system is predicate-extensional, interchangeable with it in all contexts. But b is identical with b, and so it should follow that in some possible world b might not be identical with b, which is absurd.

It appears intuitively that a system can be transparent

1. This statement is of course not truth-functional. If a system were both transparent and truth-functional I think it would have to be predicate-extensional as well.

2. One might object to this. On my view a is part of the extension of F if Fa is true and a actually exists. This has the consequence that an item, like the number nine, can be part both of the extension of an expression and of the extension of its logical complement, depending on how it is described: nine, so described, is the extension of "is necessarily the square of 3", but described as the number of the planets it is part of the extension of "is not necessarily the square of 3". So one might prefer to say that two predicates count as coextensive only if they each apply to the same (actual) things-under-the-same-descriptions, and not whenever they each apply to the same set of things however described. On this interpretation, I think, every predicate-extensional system would have to be transparent.

without being truth-functional, an example being provided by causal propositions as they are ordinarily understood. For we do not ordinarily consider the causal "if--then" to be material - and yet we feel that it doesn't matter under what description we consider the events and states of affairs mentioned in causal hypotheticals.[1] If the Battle of Waterloo is the same event as the battle at which Wellington defeated Napoleon, it has the same effects. Quine, however, provides an argument (derived from Frege) which is supposed to show that any system which is transparent must be truth-functional.[2] This I shall consider much more quickly than it deserves, since the question whether a transparent system must be truth-functional will not concern us in the rest of this chapter; but it goes like this.

Consider any compound sentence in a transparent system: perhaps, "If he strikes the match, the match lights". This can be represented as having the form

(1) $C(p)$

Since this belongs *ex hypothesi* to a system it must be possible to substitute logically equivalent expressions in it. Now it is a logical truth that the class[3] of things x such that both $x=x$ and p is true is the same class as the class of things x such that $x=x$, if and only if p is true. That is,

(2) $[(\hat{x})(x=x.p) = (\hat{x})(x=x)] \rightleftarrows p$

-- where '$\rightleftarrows$' is to indicate logical, not just material, equivalence. So we can substitute in (1), and get

(3) $C[(\hat{x})(x=x.p) = (\hat{x})(x=x)]$

1. Cf. Davidson, "Causal Relations", Journal of Philosophy 1967; and Anscombe, "Causality and Extensionality", Journal of Philosophy 1969.
2. From a Logical Point of View, p.159; Word and Object, p.198. See also Davidson, op.cit.
3. The argument can be stated in terms of definite descriptions if classes are found objectionable. Cf. Word and Object, p.198.

- or, in terms of the example, "If he strikes the match, then $[(\hat{x})(x=x \text{ and the match lights})=(\hat{x})(x=x)]$." Now take any proposition q which is materially equivalent to p (to "the match lights"). By the same argument that gave us (2), we have of course got

(4) $\quad [(\hat{x})(x=x.q) = (\hat{x})(x=x)] \rightleftarrows q$

But since p and q are materially equivalent, they must either both be true or both be false. If they are both true, then

(5) $\quad (\hat{x})(x=x.p) = (\hat{x})(x=x) = (\hat{x})(x=x.q)$

If they are both false, then

(6) $\quad (\hat{x})(x=x.p) = \Lambda = (\hat{x})(x=x.q)$

(where Λ is the null class). So in either case

(7) $\quad (\hat{x})(x=x.p) = (\hat{x})(x=x.q)$

And since, owing to the character of the system, (3) is transparent, we can now get from it

(8) $\quad C[(\hat{x})(x=x.q) = (\hat{x})(x=x)]$

Whence, by substituting logical equivalents as set out in (4), we get our conclusion:

(9) $\quad C(q)$

This shows, according to Quine, that transparent systems must be truth-functional.

The intuitively unsatisfactory thing about the argument is the use it makes of the expression "$(\hat{x})(x=x.p)$". There is something intuitively uncomfortable about the idea of a closed sentence coming within the scope of a quantifier; though a practised logician may not feel this oddity, any more than he may feel any strangeness about the paradoxes of strict implication. If the argument is to go through, "$(\hat{x})(x=x.p)$" must mean: that class of things which is $(\hat{x})(x=x)$, i.e. the universal class, if p is true, and which is the null class if p is false. And in general $(\hat{x})(Fx.p)$ is that class which is the class of F's if p and the null class if not-p. Such a class[1] is a new entity that we did

1. Or, in the class-free way of putting the argument, the referent of the corresponding definite description: $(\iota x)(Fx.p)$.

not talk about before; but it looks an easy one to introduce into a system in which $(\hat{\underline{x}})(\underline{Fx})$, $\underline{\Lambda}$, and $\underline{p}$ are already well understood. It can be introduced by the following equivalence, in which "$\phi(\ )$" stands for any propositional function taking a singular term as argument:

(10) $\quad \phi(\hat{x})(Fx.p) \rightleftarrows \{[p \supset \phi(\hat{x})(Fx)] .[\sim p \supset \phi(\Lambda)]\}$

In other words, any proposition in which the new class is mentioned is equivalent to a conjunction of two propositions which were expressible in the system before and which consequently are transparent. But since (10) introduces a new singular term, the transparency of the definiendum is not guaranteed by the transparency of the definiens.

If we translate (3) into the form which, by (10), is definitionally equivalent to it, we get

(11) $\quad C\{[p \supset ((\hat{x})(x=x)=(\hat{x})(x=x))] .[\sim p \supset ((\hat{x})(x=x)=\Lambda)]\}$

What follows $\underline{C}$ here is indeed logically equivalent to $\underline{p}$, so that the move from (1) to (3) is all right. But the move from (3) to (8) can now be seen to be valid only if the system is truth-functional, which is what was to be proved. On that assumption it is perfectly reasonable to move from (11) to

(12) $\quad C\{[q \supset ((\hat{x})(x=x)=(\hat{x})(x=x))] .[\sim q \supset ((\hat{x})(x=x)=\Lambda)]\}$

--given that $\underline{p}$ and $\underline{q}$ are materially equivalent. For a truth-functional system simply is one in which we can substitute for one another materially equivalent propositions in whatever context they occur, without affecting the truth-value of the whole context. Without that assumption, the move from (11) to (12) is entirely unlicensed: in other words "$(\hat{x})(x=x.p)$" is really opaque after all, it cannot be substituted for "$(\hat{x})(x=x.q)$" in all contexts, unless we assume that the system is truth-functional.

So the move necessary to introduce artificial classes of the form $(\hat{\underline{x}})(\underline{Fx}.\underline{p})$ is legitimate only if the system is a truth-functional one, since otherwise it would introduce opacity into a system which was supposed to be free of it.

And so Quine's argument does not go through, unless it begs the question.

Against this it may be said that expressions of that form may not have to be introduced into the system because it already contains them, if it is at all like a natural language. We have no very clear idea, outside the context of a formal system, of what a primitive and unanalysable predicate would be like; most of our ordinary predicates can be analysed, and in some cases the analysis might consist in showing the predicate really to be of the form (F( ).p). And it might be awkward to have a rule according to which no predicate could be of this form. This objection might be resisted; but for present purposes it is enough to point out that Quine's argument only works for a system which already contains, and so does not have to have introduced into it, predicates of this form[1] for all values of p. Such a system as that certainly is, if transparent, truth-functional; for in saying that a system is transparent one is saying that if it does contain expressions of this form then identically-designating expressions can be substituted for them. Thus in a system containing some expressions of the form $(\hat{x})(x=x.p)$, some materially equivalent propositions will be intersubstituable salva veritate in any context within the system; but it is not truth-functional unless all materially equivalent propositions are interchangeable; and since we cannot without begging the question introduce new expressions of the required form, we cannot guarantee that they will all be interchangeable unless the system already contains expressions of this form, not just for some, but for all values of p.

1. In fact, the argument only goes through if the value of F is such that "(x)(Fx)" is a necessary truth (as, for example, "(x)(x=x)" is): otherwise there would be no logical equivalence between "$(\hat{x})(Fx.p)=(\hat{x})(Fx)$" and "p".

So I think Quine's argument does not really show that all transparent systems must be truth-functional. To avoid truth-functionality they will have to have some greater restriction on the admissibility of predicates[1] than Quine has, but exactly what restriction they adopt does not matter, provided only that some limitation is put on the formation of predicates of the conjunctive F( ).p form. The same point can again be made if we turn to the question whether all predicate-extensional systems must be truth-functional; for here again an argument, parallel to Quine's, can be drawn up to show that they must, but this argument also turns, in the same way, on the free availability within the system of predicates of the F( ).p form, which as we have seen can in general be freely introduced only into truth-functional systems.[2]

A slightly different, more Fregean, way of trying to reach this same conclusion, that predicate-extensional systems must be truth-functional, would be to say that "p" is logically equivalent to "Truth is characterized by p". So we can go from "Cp" to "C[Truth is characterized by p]" (where C( ) is any propositional context, as before). If "p" and "q" are materially equivalent, the predicates "is characterized by p" and "is characterized by q" are coextensive, since each applies only to truth. So in a predicate-extensional system we can move to "C[Truth is characterized by q]" and thence to "Cq".

1. Simply restricting class-abstraction won't do, because of the possibility of stating the argument in terms of definite descriptions - using e.g. "the x which is an even prime and p" instead of "$(\hat{x})(x=x.p)$".

2. The argument would, of course, proceed by claiming an equivalence between "p" and (for example) "$[(x=x.p)=(x=x)]$", and pointing out that "$__=__.p$" and "$__=__.q$" are coextensive if p and q are materially equivalent.

Once more the argument works if the system already contains as primitive the rather artificial diction used, but if this has to be introduced the question is again begged, in just the same manner. The predicate-extensionalist who wishes to avoid truth-functionalism will refuse to let us treat truth as an actual entity characterized by propositions.

This section has been essentially a digression, since all I really wanted to do was to make the point that there are at least three things that can be meant by calling a system extensional, and that these three things do not obviously, at least, come down to the same. I have suggested that for a system to be extensional in one of these ways is not necessarily for it to be extensional in either of the others, except that predicate-extensional systems do have to be transparent; but it does not matter, for the main purposes of this chapter, if what I have said on this score is wrong. Quine's account of meaning yields a language which is transparent and predicate-extensional, since the meaning of any expression (whether a singular or a general term) has to be given by the set of stimulations that would elicit it; and since he considers such languages to be truth-functional he does not think he can allow non-truth-functional operators in his purified system. But in what follows I shall not be considering Quine's account of meaning particularly; it is interesting principally because it illustrates the difficulty of giving a causal account of meaning in any detail. I shall be concerned with a more general question. Taking any of the three senses of "extensional", can a theory of linguistic meaning be adequate if it can only give an account of extensional systems, and in particular if it is itself expressed in extensional terms (so that these systems can contain their own semantics)?

Some would say that the answer was "Obviously not"; but that is too simple. My main aim, of course, is to try to defend analyticity: "It is analytic that...." generates opacity, predicate-intensionality, and non-truth-functionality.

Extensional Causality

If the idea of one event's causing another cannot be expressed in purely extensional terms, this does not immediately show that there could not be percipients who spoke a purely extensional language; for we did not succeed in providing a transcendental deduction of the idea of causality (as opposed to that of regularity). But it would certainly render extremely implausible the suggestion that only those parts of our ordinary discourse that can be translated into purely extensional form are intelligible. This suggestion does indeed look implausible enough when we think that only notions like analyticity are ultimately going to be ruled out, but after all even the most thoroughgoing intensionalist would admit that the notion of analyticity is not as clear as he would like it to be. If causality has to be dropped as well the implausibility becomes staggering. Even more importantly, one thing, perhaps the main thing, in favour of the extensionalist approach was that it offered a causal account of meaning, which looked scientifically and ontologically neat. This neatness is lost if the semantics of an extensional language have to be given in non-extensional terms.

But although clearly our ordinary causal statements are non-extensional, since at least they are not truth-functional, I think a language extensional even in all three ways can contain statements sufficiently like them to do much of the same work. Regularities could be observed and used to ground statements to the effect that events of one type are

regularly followed by events of another type - perhaps we might call them strictly Human causal statements. They must be truth-functional, transparent, and predicate-extensional; nevertheless they would be of some use in discovering order in the world - if we could use only extensional discourse we might still have something we could call science.

The main hurdle is truth-functionality. It is comparatively natural to think of our ordinary causal statements as transparent, and even as predicate-extensional, for we think of causal statements as setting out a relation between events or between things in the world, regardless of how we may choose to describe them. Recently, convinced by Quine's argument that transparent systems are truth-functional, some have concluded that causal statements are not transparent (or presumably predicate-extensional), a view which can be supported by saying that causal statements offer explanations, require one to see the phenomena in a special way.[1] Singular causal statements are backed up by general ones, and so should describe events in the way that brings them under the generalisation; even general causal statements should be put in a form which most readily enables them to be subsumed under more general ones. Clearly, there are advantages in this way of talking, but the dispute over which is <u>the</u> correct way of taking ordinary causal statements is a barren one. For certain purposes, when it is the pattern of the explanation that we want to stress, the opaque or predicate-intensional form is better; for others, when we want to underline the connectedness of the phenomena, transparency or predicate-extensionality or both are preferable.

1. Cf. Anscombe, "Causality and Extensionality", <u>Journal of Philosophy</u>, 1969. This position is argued on its own merits, without regard to the Quine argument, by N.R. Hanson, <u>Patterns of Discovery</u>, Cambridge 1958.

But before we turn to consider truth-functionality, we should look at a difficulty which would arise for anyone who insisted that all causal statements are transparent.[1] Causal generalisations are of the form "A-type events cause B-type events", and their primary function is, to enable us to say given an A-type event that a B-type event will occur; or perhaps, that it will be more probable, for the observed regularity may not have been a 100% one. But suppose, now, that 96% of A-type events are followed by B-type events and 32% of C-type events are followed by B-type events; what happens if a particular event turns out to belong to class A as well as to class C? Thus it might be that milk's being kept above 30°C for 24 hours was followed by its going sour with a probability of 96%, whereas milk's being kept above 15°C for 24 hours was followed by its going sour with a probability of only 32%. Yet any particular case of milk's being kept above 30°C for 24 hours is also a case of its being kept above 15°C for 24 hours. Given such a particular case, we are faced with a contradiction: the probability of the milk's going sour has to be both 96% and 32% (both above and below 50%).

One could escape this by saying that the probability is relevant to the description of the antecedent, but that is to abandon transparency. It seems that this notion of probability cannot be used by the speakers of a transparent language. Clearly there is nothing to stop them observing that 96% of A's are followed by B's and only 32% of C's are; but they may not infer from this that if this is an A then there is a better chance that it will be followed by a B than if it is a C.

Where the observed correlation is a 100% one we can permit the inference about a particular case, for it will not be possible to assign different probabilities to the

1. This difficulty was pointed out to me by Mr. J.D. Kenyon.

same event. This would so far still be true if we took some other figure than 100%, say 8%, so long as we stuck to it - only permitted inferences about particular cases where the antecedent event has been followed by the consequent in 8% of the cases, and in no other circumstances. It is more useful to use the 100% rule than the 8% one, but it would be better still if we could insist not on 100% but rather on 98% or more, not allowing the difference between a 98% observed regularity and a 99% observed regularity to give rise to different predictions in different cases. It is convenient to be able to make causal generalisations even where the observed concomitance has not been quite complete. But to avoid running in to the same difficulty as before we must be careful not to allow the contradictories of statements of our chosen form to give rise to inferences about particular cases. Otherwise, where a particular event is both an A and a C, and all (or, over 98% of) A's are followed by B's and not all (or, not over 98% of) C's are followed by B's, we could infer both that this event will be followed (almost certainly) by a B and that it is not the case that it will be followed (almost certainly) by a B. So from the facts that not all C's are followed by B's and that this is a C we can allow no conclusion to be drawn. It will still of course be possible to argue from a particular instance to the contradictory of a universal causal statement: if this A is not followed by a B, not all A's are followed by B's. And we can still argue that a case of A's being followed by a B confirms the hypothesis that all A's are B's.

But there is still a difficulty. Suppose all A's have been observed to be followed by B's and all C's by non-B's. Now for the first time we find an A which is also a C; this licenses the predictions that a B will follow and also that no B will follow. One must reply that this shows either that it was wrong to say all A's cause B's, or wrong to say

all C's cause non-B's. But what about the case in which we observed (only) nearly all A's to be followed by B's, and nearly all C's to be followed by non-B's? Since we cannot avoid the incompatibility of the predictions "Probably a B will occur" and "Probably a non-B will occur" by making them relative to the way we describe the antecedent event, it looks as though it will not do to say that the regularities from which we can permit ourselves to make inferences about particular cases are ever less than 100%.

It seems then that with a purely transparent or predicate-extensional language we can observe regularities and use these observations to predict or explain a particular case, provided we do not allow these particular predictions or explanations to assign different degrees of probability to the consequent, or indeed any degree less than 100%; at any rate on the sort of account of probability I have been working with, viz. in which the probability of H on E is a function of the relative frequency with which the E-H connection has occurred in the past. This leaves us with a limited notion of causality, and one might wonder whether it would be of much use in constructing a science. The difficulty is that few of the regularities we observe are likely to be 100% ones - there are so many intervening factors which, to start with, we can't foresee. So apparently we shall have practical difficulties in setting up a science operating with this transparent notion of causality. However, there is a possible way out of this difficulty at least. For we can insist on the "100%" formula, but where only (say) 50% of A's have been observed to be followed by B's we can permit ourselves to make the hypothesis that there is some further discoverable factor, or factors, Q, such that the combination of A and Q is followed by B in all cases. It is important that Q must be a discoverable factor, or at any rate that some restriction be placed on Q, for

there will be in principle some condition (even though the specification of it may be infinitely complex) for any A and any succeeding B such that B will occur if A occurs together with that condition.

It is when we add the requirement of truth-functionality that the extensionalist's causal statements really do come to look very different from our own. We regard our causal generalisations as supporting counter-factuals; there is no place for these in a truth-functional account, since a false proposition materially implies any proposition. And we think of them as having some explanatory content which is lost if any proposition with the same truth-value will do just as well as antecedent or consequent in the hypothetical. If it is true that I strike a match and it lights, and true also that there is no life on the moon, we are inclined to regard my striking the match as a better explanation of its lighting than the lifelessness of the moon, and also to regard the fact explained by my striking the match to be its lighting and not anything to do with the moon. Such selectiveness is not, it seems, open to the truth-functionalist.

Two lines of defence are open to him. The first and stronger one[1] is to say that his conception of causality is no different from ours: ours is at heart truth-functional too, and the apparent departures from truth-functionality are to be accounted for by means of the Gricean idea of conversational conditions. The second and (in the logical sense) weaker one is to concede the differences between his conception and ours but to argue that he still has left an idea sufficiently like ours to do a large number of the same jobs.

1. Cf. for example D.F. Pears, "Hypotheticals", Analysis, 1950; and Grice's lectures at Oxford in Trinity Term 1965.

The proponent of the first line can say that the oddity of "If there is no life on the moon the match lights" is due not to this being false but to its being an entirely pointless thing to say. It provides the information that either there is life on the moon or the match lights, but this is not the sort of information anybody wants and it is not even a very good way of expressing it. Things are different with "If I strike the match it lights"; this is of interest just because it is an instance of a general (truth-functional) hypothetical, "Whenever anyone strikes a match it lights". And indeed we do feel that it is some such general hypothetical which is responsible for the explanatory character of "It lit because I struck it". Counterfactuals can be approached in the same way: it is indeed <u>true</u> that if I were standing on my head just now the moon would be made of green cheese, but it is a pointless thing to say, whereas that if I were standing on my head blood would be flowing to my head is not only true but (more) pointful, since it instances a well-confirmed generalisation about people who are standing on their heads.

The proponent of the first line can make out a powerful case. But it is sufficient for my purposes here if the second line of defence can be made to work. One can start to do this by drawing a distinction between saying "If <u>p</u> then <u>q</u>" just because it is true and putting it forward as an explanation of <u>q</u>, by saying that an explanation of <u>a</u>'s being <u>G</u> has to be (roughly) of the form

(1) $Fa.\ (x)(Fx \supset Gx)$[1].

This sort of thing is perfectly expressible in a purely

1. This could of course be complicated to take account of cases in which it is a predicate variable that is universally quantified over and cases in which <u>p</u> and <u>q</u> are more complicated than the simple <u>Fa</u> form.

truth-functional system, so we can allow that people using such a system can give explanations not altogether unlike the typical sort of causal explanations we ourselves give. It is not so clear that they themselves can make this point, for since (1) is truth-functional they will happily substitute any proposition of the same truth-value for Fa wherever it occurs; and we would not want to say that

(2) Hc. (x)(Fx ⊃ Gx)

provided an explanation of Ga just because Hc has the same truth-value as Fa. They might be able to express the distinction between explanations and merely true hypotheticals by using quotation-marks in Quine's sort of way, i.e. so that ""p""[1] is regarded as naming the sentence "p" and not as containing an occurrence of the sentence "p"; then they could say

(3) "Fa. (x)(Fx ⊃ Gx)" explains a's being G, whereas "Hc. (x)(Fx ⊃ Gx)" does not.

For they would not then regard (3) as containing any actual occurrence of Fa or Hc, but only names of these, just as if it had explicitly referred to (1) and (2) above under those labels, without spelling them out. So that the postulated material equivalence of Fa and Hc licenses no substitutions: they are no more constituents of (3) than the word "bee" is a constituent of the word "beetle". Unfortunately, however, this way of construing quotation-marks will not do generally if the system is to contain its own semantics, as has already

1. Points of this sort are always difficult to state since in talking about a proposition one has to name it and in talking about its name one has to name that name, whence the need for double quotation-marks. What is essential in what follows is that in the sentence

   ""p"" names "p"

   the second p is named.

been mentioned; for if ""p"" is not to be regarded as containing an occurrence of "p", but just as an arbitrary name for "p" - in which case it might just as well be written "Z" - it will not be possible to explain to someone what ""p"" (or "Z") means unless he already knows what it means. We can say it names "p", or that it names Z, but then we have ""p"" ("Z") itself in our explanans - we wanted to get to the proposition itself but we have still only got the name of the proposition.[1] It could be suggested however that although the assignation to them of arbitrary names is not the only way of talking about sentences or propositions, nevertheless it is a possible way - in other words, that side-by-side with ordinary, analysable quotation, in which ""p"" contains an occurrence of "p" as well as mentioning it, there is also to be found unanalysable, Quine-type quotation, in which ""p"", like "Z", no more contains "p" for any logical purposes than "concatenation" contains "cat". If this were so our truth-functionalist could still assert (3) so long as his quotation-marks were of the second, Quinian type. But not even this will do if the truth-functional system is also transparent. For then since the Quinian arbitrary name ""Fa"" names the very same sentence ("Fa") as the ordinary, analysable quotation ""Fa"" does, the latter can be substituted for the former, and then - since we now have a genuine occurrence of "Fa", even though it is in quotation marks - any truth-functional equivalent of "Fa" can be substituted for it , as before.

But although the truth-functionalist cannot state the distinction between an ordinary hypothetical and an explanatory

1. K. Reach, "The Name Relation and the Logical Antinomies", Journal of Symbolic Logic, 1938.

one, he may still make use of it in practice. When we call him a truth-functionalist we mean that he accepts that substitution of truth-functional equivalents in any contexts preserves the truth-value of the whole, but not necessarily that he finds the new proposition just as useful or interesting as the old.[1]

At the very least, he has a notion of implication which lies at the core of any more complex notion of implication that we may have: it cannot be the case both that the antecedent is true and that the consequent is false. And he can make universally quantified implication statements about events, and apply these generalisations to particular cases. It is true that if his language is to be transparent and predicate-extensional, as well as truth-functional, the application to particular cases can only be allowed if the generalisation is a 100% one (or at least can be made into one by bringing in an hypothesis about a discoverable but as yet undiscovered extra condition). This may not be enough to warrant us in saying that he has "the idea of causality"; but it does seem enough to justify the assertion that he has a conception of a way events can be related which is like our idea of causality in the way that it helps him to understand and manipulate his world. And there is at any rate no prima facie reason why among the phenomena he can come to understand in this way should not be his own utterances of meaningful expressions.

So it will not do simply to say that extensionalism is wrong because even extensionalism cannot dispense with the non-extensional idea of cause in its theory of meaning. The stripped-down, purely extensional notion of causality retains what is essential to it, and is not obviously

1. Though of course if his interest in them differs he will be unable to state that it does.

inadequate for semantics. Naturally, a lot more would have to be said to show that it <u>is</u> adequate fof anything recognizably like our scientific theories; but it seems reasonable to give to the extensionalist the benefit of whatever doubt there may be here.

Human Behaviour: Causation and Teleology

Can an account of human behaviour be given in purely extensional terms? And in particular can an account of language-use be so given? I shall assume that the extensionalist can make use of ordinary causal explanations, even though the last section suggests that, at least if he is a truth-functionalist, he can only make use of stripped-down analogues of them; for it is simpler to talk in this way, and the translation into the stripped-down form could always be effected. But the most usual explanations we give of human behaviour, including language-use, are apparently not of the ordinary causal form. Actions are explained by reference to the beliefs and desires of the agent; statements about beliefs and desires are neither transparent, truth-functional, nor predicate-extensional. And in explaining the action by reference to what the agent wants we bring on the idea of a goal, which may be the action itself or may be something seen as lying beyond it: and since the goal may never be attained, and if attained will only be so after or at the same time as the performance of the action, it looks as though no amount of tinkering will enable us to treat it, in a manner acceptable to the extensionalist, as an ordinary efficient cause of the action.

Three courses are open to the extensionalist. One is to say that no intelligible account of meaning can be given;

a rather desperate course, since it leaves it literally unintelligible how people can communicate. The second course is almost equally heroic: it is to say that the explanation of human behaviour in terms of beliefs and desires (which I shall in future call purposive explanation) is never the correct one. The true explanation is always in straightforward efficient-causal terms. This means that at the moment we understand very little about our own behaviour. This position, though so extreme that perhaps no one would wish to accept it, is I think invulnerable to philosophical attack. It is highly implausible, but involves no contradiction; nor can it be refuted by a transcendental argument.[1] Certainly, if having a language necessarily involves having intentions, we cannot have a language unless we have intentions; but the acceptability of the premise, as offering a partial definition of language, depends on the assumption that we do behave purposively. If we do not, then of course we don't have a language in <u>that</u> sense, but we may have some other sort of signalling system, the properties of which will remain to be investigated. In my discussion of transcendental and metaphysical arguments, above, I did of course assume that our linguistic behaviour was purposive; but this was a reasonable assumption then, since

1. Malcolm tries to refute it in this way: "The Conceivability of Mechanism", <u>Philosophical Review</u> 1968, esp. pp. 67-71.

the whole enterprise assumed the viability of the analytic/synthetic distinction and so the falsity of extensionalism.

The third course open to the extensionalist is to say that purposive explanations can be reduced to extensional terms. And this is by far the most interesting and plausible course for him to take, and I shall investigate it in what follows. I shall argue that the fundamental question here turns out to be an empirical one; so that the case for the intelligibility of notions like analyticity ultimately rests on an empirical issue.

At the outset I should make it clear that the question I am raising is not that of whether desires <u>cause</u> actions or rather are related to them in some other way. That question turns on whether one can separate sufficiently the criteria for the occurrence of a desire from the criteria for the occurrence of the action flowing from it. Pears has offered strong reasons for thinking that one can[1]; and certainly, since desires can be caused, this would enable us to explain an action by citing a causal chain. This does not help the extensionalist, however, because he is still unable to express the concept of desire: to say that A wants <u>x</u> is already to

1. D.F. Pears, "Predicting and Deciding", <u>Proceedings of the British Academy</u> 1964; "Desires as Causes of Actions", <u>Royal Institute of Philosophy Lectures</u> 1966-7; "Are Reasons for Actions Causes?" in A. Stroll (ed.), <u>Epistemology</u> (New York, c. 1967).

say something intensional. The question I am concerned with is rather whether one could replace explanations which use the concept of desire with efficient-causal explanations of the sort which dispense with it (in favour, perhaps, of the concept of a neurophysiological state).

The extreme positions on this question are occupied by Ernest Nagel and Denis Noble on the one hand, and Charles Taylor on the other.[1] Nagel and Noble both hold that if a piece of behaviour can be explained purposively it follows logically that it can be explained in a non-teleological way as well. Taylor, on the contrary, holds that if a piece of behaviour can be explained teleologically it follows logically that it cannot be explained non-teleologically as well[2], and he also holds the converse of this: so that teleological explanations compete with non-teleological ones for the same ground. (He takes purposive explanations to form a subset of teleological explanations). Both these extreme positions are wrong.

The argument of Nagel and Noble is essentially this.

1. Nagel, The Structure of Science (London, 1961), ch.12; Taylor, The Explanation of Behaviour (London, 1964), esp. ch.1; Noble, "Charles Taylor on Teleological Explanation", Analysis 1966-7; Taylor, "Teleological Explanation", Analysis 1966-7; Noble, "The Conceptualist View of Teleology", Analysis 1967-8.
2. Cases in which this can apparently be done, as for example with thermostats, are treated as cases in which the teleological explanation is not properly applicable but only used as a sort of rough shorthand - on the grounds that here the non-teleological explanation is the correct one (gives all the right predictions), so that the teleological one must be strictly incorrect.

A teleological (or purposive) explanation says that a certain event, a piece of behaviour of a system, occurred because the state of the system and its environment were such that the behaviour was required if the system was to obtain its goal; or at least, that they were such that the behaviour was, or seemed, conducive to its obtaining the goal. And this may be backed by a covering generalisation to the effect that the system always, under such-and-such circumstances, behaves in a way conducive to its obtaining this goal. But they point out that since the initial condition must always be some event or state of affairs there is no need to specify it in terms of the goal - i.e. to specify it as "the state of the system and of its environment being such that behaviour B is, or seems, conducive to the system's goal". Instead, we can just describe the observable features of the system at the time. There is no real necessity to describe a rat's condition at t as one of hunger for cheese - thus indicating that eating cheese would be conducive to its goal; we can provide instead a physiological description of the rat at t. This again will provide us with the antecedent of an hypothetical: "If the rat is in such-and-such a state, then it will produce behaviour B". This again may be regarded as backed by a covering generalisation to the effect that rats in this physiological state behave in this way. But now we are explaining B by reference simply to its

efficient causes, without any mention of goals; for the specification of the antecedent is in physiological terms, and not in terms of any goal to be reached by means of B.

This argument was supposed to show that whenever a teleological explanation can be given for any B, it follows logically that a non-teleological explanation can also be given. All it actually shows is that any particular piece of behaviour was preceded by the system's (and its environment's) being in some physical or physiological state specifiable without reference to its goals: a fairly unremarkable point. What has not been shown but merely stated, and what would need to be shown, is that some generalisation is true to the effect that whenever the system (and environment) is in a physical or physiological state of a particular recognizable type, B-type behaviour follows.[1]

According to the teleological account, B-type behaviour

1. Possibly it may not be the case that all explanations must be backed by generalisations, but I can see no feature at all in the consideration, that B was (on this one occasion) preceded by a certain specific condition, which would justify us in saying that this consideration gave us any explanation of the occurrence of B, unless it is that this specific concomitance of the physical condition and B is to be regarded as an instance of some such generalization.

will occur whenever the state (S) of the system and its environment (E) are such that B is or seems conducive to the goal (G). According to a non-teleological account, B-type behaviour will occur whenever some intrinsically-described state of affairs, C, occurs. To show that the teleological account entails the causal account it would be necessary to show some connexion between the two antecedents. But this has not been shown. All that has been shown is that, whenever S and E are such that B is or seems conducive to G, some physical (intrinsic) description of the system and environment can also be given without mentioning G - but not necessarily the same one each time: on one occasion it may be C_1, on another C_2, on another C_3, and so on, and C_1, C_2 and C_3 may not have anything in common which could figure in the antecedent of a generalisation of the form "Whenever C occurs, B-type behaviour follows". Thus, it is logically necessary that if the rat is hungry for cheese and considers B the best way of getting it, there be some intrinsic physical description true of him. But it does not follow that there must be some such physical description true of him whenever he is hungry for cheese, and considers B the best way of getting it.

Of course, it may be that there is some intrinsic description true of him whenever S and E are such that B is or seems conducive to G. But if this is so the fact

that it is so is contingent: a matter to be established empirically. If we could show that it was in fact so, we should have established (prima facie) a case in which both a teleological and a non-teleological account can be given of the same event. Taylor's objection to the view that there are (ultima facie) any such cases will be considered in due course.

There is an objection to what I have said so far. Nagel and Noble can argue that to show that some intrinsic description must apply to the system and environment just is enough to show that one can get going a non-teleological generalization of the form "Whenever C, B", and so effect the reduction of teleological to non-teleological explanation. I argued that it was not sufficient because the physical state of the system might be different on different occasions when S and E were such that B is conducive to G: on one occasion it might be C_1, on another C_2,... on another C_n. But why not in this case take the disjunction $C_1 \vee C_2 \vee \ldots \vee C_n$ and present this as the antecedent condition of our non-teleological hypothetical? If we are to insure ourselves against the sort of counterexample in which C_1 or C_{23} occurs when S and E are not such that B is required for G, we shall have to specify our C's pretty exactly; and we may, in a particular given case, have to go so far as to bring in singular terms, references to dates and places, etc., to do this. For it may be that C_1 is the state of the rat at t_1

and p_1, when S and E are also such that B is required for G; but also, for all that has been shown to follow logically, the rat may be in C_1 at ~~many~~ many other places and times, unless C_1 is so specified as to bring in date and place; and on none of these other occasions on which it is in C_1 may S and E be such that B is or seems to the rat required for G - and on none of them may B follow. So no causal law of the form "Whenever $C_1 \vee \ldots$, then B" would be true: rather we would have to put it in the form "Whenever C_1 at p_1 at t_1, $\vee \ldots$, then B".

This is a suspicious-looking candidate for the status of a causal law. If we have to specify all our disjuncts in this way, what we shall have will be a list of all the physical states of the system and environment which ever precede actual occurrences of B. It is not clear why simply shoving this list into the antecedent clause of an hypothetical should be supposed to provide us with any explanation of B. Yet this is all that the existence of a teleological explanation of B logically guarantees that we shall be able to do in the way of giving a non-teleological explanation of it.

- Still, perhaps this is not by itself an entirely satisfactory reply. What I have just said rests on our unwillingness to call an hypothetical of this form explanatory; and the notion of what is or is not explanatory is pretty hazy. And there is a more serious objection to it. For if it is in

some sense a contingent matter that I do or do not see two states of affairs as similar, there might be a being who just did see as similar all the antecedent conditions C_1...., etc., which ever do occur, even though I (and perhaps all other human beings) am unable to find anything in common amongst them except by bringing in reference to the goal. This being would then be able to give a causal explanation of B which was not vacuous in the way our hypothetical involving the complete disjunction was; his generalisation can even support counterfactuals. His "# causes B", where '#' is his term for the common feature shared by members of the disjunction and by these only, plays the same role as our "C causes B" where we can detect C as the feature common to C_1, C_2, C_n, etc., the various antecedents of B-type events[1]. It has (for him) the same sort of abbreviatory and assimilating, and explanatory, power.

If we are prepared to count "# causes B" as a causal explanation of B, just because there might be some being who saw as similar, and as different from everything else, all the different antecedents of B-type events[1], then it will be analytic that every event has a cause (unless there is a first event). For one can, if omniscient, in principle take any event-type and draw up a list of all the states of

1. Or even a subset of them.

affairs which immediately precede the events of that type; and then there might be some being who could see all and only these states of affairs as intrinsically similar, even if we can't. This suggests strongly that one should put some restriction on what is to count as a causal explanation, in order to rule out this sort of thing; because we do not normally consider "Every event (except the first) has a cause" to be analytic.

But it is not clear how this is to be done; and it is sufficient to draw in our horns a bit, and recognise that the sort of explanation we can give of a phenomenon depends partly on our being human beings, beings who just do see similarities in certain cases and dissimilarities in others. The sort of explanations my imaginary beings would give, using '#' and the like, would be of no use to us at all. To defend extensionalism on the grounds that such beings are possible would be like advocating an alternative conceptual scheme because there might be beings sufficiently unlike us to find it preferable on pragmatic grounds. So instead of saying that the proposition that B can be teleologically explained does not entail the proposition that B can be non-teleologically explained - which would be the straightforward denial of the thesis of Nagel and Noble - we can at least say that the proposition that a teleological explanation which is of use to human beings can be found does not entail

that a non-teleological explanation which is of use to human beings can be found.

But Nagel and Noble can bring on another argument. This is that teleological explanation is vacuous, like "Opium causes sleep because it has a virtus dormitiva" when this is taken as a tautology. Any cause-effect statement, on this view, can be treated as a teleological one in which the effect appears as the goal: for to describe something as a goal adds nothing to the explanation. Against this, there are two things to be said. One is, that the opium example is vacuous because it is assumed that there is no way, in principle, of verifying that opium has a virtus dormitiva except by seeing that it causes sleep. But fairly clearly (though I shall say more about this shortly) it is possible to find out what someone, or even some animal, wants without seeing him actually attain it. One must accept this even if one holds that the connection between the criteria for ascribing desires and those for saying actions have occurred are too close to permit one to say that desires cause actions. The other, and related, thing to be said is that the behaviour we explain teleologically differs from the behaviour of simple machines in that it exhibits what has been called plasticity: if one course of action doesn't enable it to attain its goal it tries another, and so on, until finally the goal is reached (or some other factor intervenes). About this, again, I shall say more later.

A rather different argument might still be used to try and show that every teleological explanation can be put into a form which extensionalists might find tractable. For on Taylor's account, which I have been using, the basic form of a teleological explanation is: the situation and environment of the system are such that it must perform B, the behaviour to be explained, in order to achieve the goal G. The suggestion is, that any explanation of this sort can be treated as a backwards-causal explanation: the subsequent achievement of the goal causes the performance of B when the situation and environment are suitable. Whatever else may be doubtful about backwards-causal explanation it is no less reducible to extensional form than forwards-causal explanation, provided the range of items we allow as "actual", and hence as values for extensional variables, includes those which will be actual, and provided we consider our propositions, for purposes of truth-functional substitution, as tenseless. The important thing about Taylor's teleological explanation is that no way of characterizing the different sorts of situation and environment which lead to B can be found, other than that in them B is required for G; but this can be put just as well by saying that in them G is sufficient for B, which is just the backwards-causal account.

But the reversal will not work for all the cases allowed

to be teleological by Taylor's account: at best it will do only for those in which the goal is actually achieved. This indicates the centrality of the sort of case in which the system only tries unsuccessfully to achieve its end. A consequence is that, if backwards-causal explanation is acceptable, the further distinction which Taylor draws between purposive and non-purposive explanation ought to be given up; and accordingly I shall give it up. For he calls teleological behaviour purposive if and only if it involves the desires and beliefs of the agent[1], but if we have grounds to speak of an agent as trying we shall eo ipso have grounds for ascribing it desires and beliefs. Of course, we can continue to speak of the other cases - where there are no desires and beliefs - as teleological if we like, but they will be teleological only secondarily. If backwards-causal explanation is legitimate, they can be explained in that way and in extensional terms; so they do not provide any insuperable obstacle for the extensionalist; any more than do thermostats, which we know how to explain efficient-causally even though we can describe them (perhaps rather loosely) in teleological terms as well. But purposive behaviour we do not, typically, know how to deal with in efficient-causal terms.

1. Taylor, Explanation of Behaviour, ch.3.

Properly teleological (purposive) explanation can of course be used in certain cases where the goal is achieved. If a rat makes several unsuccessful attempts to reach the cheese, and then finally reaches it, clearly it would be unreasonable to withhold the title of "purposive behaviour" from the successful attempt. To construe all unsuccessful attempts in a quite different way from successful attempts would be to complicate things ridiculously. When we call the successful attempt purposive, we subsume it under the same generalisation as the unsuccessful attempt; it is extrinsic to the explanation of it, of that piece of the animal's behaviour, that it was in fact successful.

The criteria for describing as purposive the behaviour of a system are behaviouristic, and so apply to animals just as readily as to human beings - at least there is no conceptual bar here. We can speak of animals as trying to achieve goals, in just the same circumstances in which we can say the same about people; there is nothing derivative, no anthropomorphism, in the application of such language to dogs and bacilli. It is true that a derivative element enters in practice; for we are inclined to ascribe, especially to the larger animals, motives and intentions very similar to our own and consequently to think of their behaviour as purposive when it resembles our own even very superficially in similar circumstances, without looking hard to see whether this way of explaining what they do is really appropriate.

Animals can't declare in advance what their goals are; men often don't. In such cases, if we have no inductive reasons for supposing that the subject's goal in a certain set of circumstances will be G, we shall be unable to make any antecedent prediction of how he will behave, if his behaviour is to be purposive. But this does not make purposive explanation vacuous. In many cases we do have inductive evidence what his goal will be, having observed similar subjects in similar circumstances. And where the explanation does not help us to predict his behaviour because such inductive evidence is lacking, it still has a classificatory role to play. When an event is explained by some generalisation it is classified as similar to a whole lot of other events: with purposive explanation this classificatory function is more important than the predictive one, because of the difficulties of knowing in advance what goals an animal, or someone other than oneself, has, how strongly he desires them, and how he sees his situation. And even where we were not able to predict how he would act we can still, after the action is completed, have grounds for ascribing to him a goal different from anything actually achieved in the action.

Let us imagine a rat running through a maze to get a bit of cheese. It runs straight towards the cheese, finds its way blocked, doubles back, takes the next most obvious-looking

route, and so on. Let us suppose,moreover, that it has been starved for some time previously. Now we know that rats which have been starved want cheese; and if we are looking for a cause for the whole complex of its trying-behaviour, we have it here. But this does not make the rat's behaviour any the less purposive: desires may be caused (and presumably they generally are) without the behaviour that flows from them necessarily requiring non-teleological explanation. If what we want to explain is not the rat's trying to get the cheese, but simply its starting off by running in a straight line towards the cheese, the cause of the desire is no longer all we are looking for: we explain the running by reference to the desire, and that explanation is purposive. Of course, if every time we put a hungry rat into a similar maze it always moves along in a straight path, and if thwarted goes off at a tangent as a clockwork rat might, then the behaviour doesn't require purposive explanation at all. But if it displays any sort of intelligence or persistency in pursuing an end despite obstacles, the position will be different. This is not to say that in any case the particular manoeuvres it makes need be unpredictable for a purposive explanation to apply, but they will not be (completely) predictable on the basis of non-teleological laws alone. We may have a generalisation to the effect that every rat of intelligence n does φ in this sort of circumstance, and have reason to classify our

rat as of intelligence n; but both the generalisation and the classification are made by observing purposive behaviour.

The upshot of this is that the attempts to show that, whenever one can give a teleological explanation, it logically must be appropriate also to give a non-teleological explanation, have failed - unless we are prepared to count as explanations those which might be given by beings who classify things as similar in ways entirely different to ours, and which to us would be useless because infinitely complicated.

The fact that (useful) teleological explanations do not in general entail (useful) non-teleological explanations does indeed show that the two sorts of explanation are different, i.e. not logically equivalent. It would be over-hasty to conclude from this, as Taylor does, that the two sorts of explanation are rivals: that if an event can be explained in the one way it can't be explained in the other. Taylor does not of course deny that the behaviour of a thermostat can, at a certain level, be explained in both ways, but here he says that the non-teleological explanation is more basic: if we can give the non-teleological account we shall be able to construct the "teleological" one, and know under what conditions it will apply, but not vice versa.[1] But at the most basic level of explanation no event can be

1. Explanation of Behaviour, pp.18f.

accounted for both teleologically and non-teleologically. And this is a fairly natural thing to think, if explanation at this basic level is thought of as setting forth conditions which are necessary as well as sufficient for the occurrence of the event explained. The same event cannot have two quite different sets of necessary and sufficient conditions. Taylor's mistake, of course, is that teleological and non-teleological explanations have not been shown to be as different as all that. They are not logically equivalent, but they may still be as one might say causally equivalent: it may as a matter of fact be the case that the state of affairs described in the antecedent of one generalisation obtains whenever and only when the state of affairs described in the antecedent of the other obtains. And in that case each antecedent can perfectly well give a set of conditions both sufficient and necessary for the occurrence of events of the type described in the consequents: for the occurrence of events of the type described in the one antecedent is sufficient and necessary for the occurrence of events of the type described in the other. And there is another possibility, depending on the fact that a particular event can be described in more than one way, and so can be subsumed under more than one generalisation. Here there is no "causal equivalence" between the generalisations, but

they can still both (or all, if there are more than two) give necessary and sufficient conditions for the occurrence of the same event. For it may be that all and only A's cause B's, and also that all and only F's cause G's; and the event to be explained may be at once a B and a G. Of course, given that this B is a G, and given also the generalisations, it follows that the A which preceded it could not have occurred unless an F had occurred, and conversely; but nothing follows about any relatedness of F's and A's in general (except that the occurrence of an F is not incompatible with the occurrence of an A).

At least *prima facie*, then, the behaviour that we explain teleologically might turn out also to be susceptible of non-teleological explanation in either of these two ways. And if it does turn out to be susceptible of non-teleological explanation, and if this non-teleological explanation can be expressed in purely extensional terms, we shall not be able to argue against the extensionalist that he is unable to give any account of the behaviour we call purposive, and so unable to explain how words and sentences are used. And he will be able to say that our purposive explanations are unnecessary, so that it will be no great loss if he jettisons them.

Let us take first the case in which events turn out to be susceptible of both sorts of explanation in the *second* way:

that is to say, not because of any equivalence between the explanatory generalisations, but because of the fact that the same event may be describable in two ways and so fall under two, quite different, generalisations. The first thing that strikes one is that it will be a coincidence of truly staggering dimensions if <u>every</u> event we can explain teleologically is also susceptible of non-teleological explanation in this way. For the teleological and non-teleological accounts will, unless by accident, classify events quite differently. It is hard to think of any ground that one might have, apart from blind faith, for supposing that all these non-teleological explanations are discoverable. But suppose they are. The extensionalist cannot remain content with this, as we might: he must go on to say that the new non-teleological explanations are right, and the old purposive ones wrong, for if he admitted that the purposive ones were right as well he would be admitting the inadequacy of extensionalism. And if they are wrong they must be entirely wrong, for the non-teleological explanations in no way correspond to them: so they cannot even be dimly perceived half-truths. So here we have got back to the position I spoke of before as an impregnable but uncomfortable refuge for the extensionalist: despite appearances, ordinary explanations of human behaviour are <u>completely</u> wrong. Apart from reiterating the implausibility

of this, I don't think any more can be said.

The more promising approach for the extensionalist is to admit that ordinary purposive explanations are not completely wrong. He cannot admit they are right either, because they are not extensional, but they need not be entirely wrong if each purposive explanation can be replaced by a single non-teleological explanation (or at worst by a small set of them) which classifies phenomena together in the same way and has exactly the same consequences. Is this possible?

The essential feature of any purposive behaviour is its plasticity, "persistence towards the goal under varying conditions".[1] It is this that makes us employ purposive explanation in a particular case, as in the rat example above. The behaviour to be explained may be accompanied by other behaviour, in such a way that we can make it all comprehensible by seeing it as a series of attempts on a goal. If it is not, to explain it purposively is still to recognise that it would be accompanied by such other behaviour, if for example the goal had not been already achieved, or if the system had not turned its activities towards the attainment of some other goal. Purposive explanation says that a thing is done because it seems to the agent conducive to a goal, and so it licenses the inference that anything else (within limits) which seemed

1. R.B. Braithwaite, Scientific Explanation (Cambridge, 1953), p.329.

equally conducive to the same goal would have been done instead had the antecedent circumstances been right. And it also licenses the inference that if the first attempt is not successful the agent will try again, unless he is prevented, changes his mind, etc.

Non-teleological explanation which covered the same ground would have to make use of the idea of negative feed-back: treating the non-attainment of the goal by a previous action[1] as itself a causal factor in the production of a further action. "The non-attainment of the goal" is not the right way to describe it since the aim is to eschew the idea of a goal. But if, say, the goal is that the rat should eat the cheese, then the fact that the rat is <u>not</u> eating the cheese is a perfectly good fact, describable without using any non-extensional expressions, and it is perfectly capable of being a causal factor affecting the future behaviour of the rat. So if $\underline{\alpha}$, $\underline{\beta}$, $\underline{\gamma}$, etc. are actions of the rat; $\underline{C}$ the initiating condition - the neurophysiological state, perhaps, in which he is whenever and only when his paramount desire is to eat cheese and doing $\underline{\alpha}$ seems the simplest way of going about it, doing $\underline{\beta}$ the next simplest, and so on; and if $\underline{H}$ is the eating of cheese, then we have something like the following causal schema:

1. "Action" is of course used neutrally, not as suggesting that it really <u>ought</u> to be explained purposively.

$$C \rightarrow [\alpha.((\alpha.\sim H)\rightarrow\beta).((\beta.\sim H)\rightarrow\gamma).((\gamma.\sim H)\rightarrow\delta),\ldots]$$

This may be read either as providing an explanation for a particular piece of behaviour (say δ) by taking the variables as standing for particular actions and events, or as the corresponding generalisation, by taking them as standing for action- and event-types.

One thing wrong with the schema as it stands is that it is open-ended. Unless we can make it actually state what comes after δ, the only way we can have any clue is by implicit appeal to the teleological idea that the rat will do whatever seems to it conducive to its getting the cheese. Now of course there is only a limited number of actions it will perform with this end in view. A rat's inventiveness is limited - so is a man's - and other things, from change of mind to death, may stop him from pursuing the goal even before his inventiveness is exhausted. But the purposive explanation supports endless counterfactuals about what he would have done if he hadn't died, or changed his mind, or failed to think of it. We cannot make the non-teleological account do this, without making it infinitely long[1]

1. If our extensionalist is a truth-functionalist, he will strictly be unable to regard his explanations as supporting counterfactuals, since he must regard all hypotheticals with false antecedents as true. But he may regard the counterfactual instances of his explanatory generalisations as not just true but interesting.

and if it were infinitely long it would not be of any use to us, any more than the explanations provided by imaginary beings who see similarities where we see none are of any use to us. Death, change of mind, and so on, pose a less serious problem than lack of inventiveness, for we can accommodate them by listing in our α, β, γ series all the things it would have occurred to the agent to do had no interfering factors cropped up. But we cannot cope in a similar way with what he would have done <u>if</u> he had only thought of it. He <u>could</u> have thought that any of a literal infinity of things was conducive to obtaining his goal. There is no reason to include only those things which <u>would</u> have been conducive to obtaining the goal: we have already gone beyond these by including attempts which he actually performs and which fail. What we have to include is all those things which he <u>might think</u> conducive to his aim, and there is no end to them. It is true that if I saw a rat methodically stroking its whiskers with a lump of cheese immediately in front of it I should not say that its overwhelming present desire was to eat the cheese, but this is only because I have no reason whatever to suppose that by stroking his whiskers the rat is invoking the gods to make sure that the cheese is not the bait of a rat-trap. Although this hypothesis would be a wild one, it seems <u>possible</u> that the rat might think along these lines, in which case the whisker-stroking would indeed seem to him conducive to the

goal (dead rats can't eat cheese).[1]

This difficulty may look comparatively unimportant. After all, we have got a possibly quite long list of things for the rat to do just by drawing on what he will think of if his previous attempts fail; perhaps it doesn't matter very much if the non-teleological explanation can't go on to tell us about what he would have done after that, if he had had more ideas. But the trouble is not as limited as that. The action α, which he performed first, was performed first because he thought it the simplest and most effective way of achieving the goal, according to the purposive account. But of course he might not have thought this. If his beliefs had been a little different he might have thought ψ the simplest and most effective way of achieving it. And in that case, the purposive account indicates, he would have started with ω. The non-teleological account tells us nothing of the sort: it can only (as it stands) cope with the case in which α is the first thing he thinks of, β the second, and so on. Of course, we might succeed in finding another antecedent state, intrinsically describable as C',

1. It might be said that we know rats don't have such thoughts because there is no way in which they could express them. The example may then be put using human beings instead of rats. If the point were well taken, it might be grounds for thinking that the sort of purposive explanation which can be applied to rats is altogether simpler than that which can be applied to humans.

such that

$$C' \rightarrow [\omega.((\omega.\sim H)\rightarrow \psi).((\psi .\sim H)\rightarrow X)...]$$

and another, C", such that

$$C'' \rightarrow [\alpha.((\alpha.\sim H)\rightarrow \omega).((\omega.\sim H)\rightarrow \varepsilon)...]$$

and so on indefinitely. Now however we no longer have a single non-teleological explanation, but an indefinitely large set, corresponding to the one teleological explanation, "α was done because it seemed to the agent in the circumstances to be the best way of attaining the goal". And since there is no limit to the things that might be thought conducive to a goal, presumably the set will in fact be infinite. Of course, infinitely many members of the set will set out causal sequences that never occur, because no one, or no rat, ever does think the actions they involve conducive to that goal. So if we were trying to set up non-teleological laws from observed data we should never set these up as laws. Still, if we want to provide a non-teleological law which covers exactly the same ground as a teleological one, it will have to take the form of an infinite conjunction of these sequences which never do occur and of those which do, like (perhaps) those which start from C, C', C", etc. Yet infinitely complex laws are of no use whatever.

This difficulty can be in part surmounted if science is kinder still to the extensionalist, and allows the discovery of some factor common to all the antecedent conditions C, C',

C", etc. whenever G is the goal; or, in non-teleological terms, whenever it is ~H that is the causal factor that triggers off the later actions in the causal sequence. In discussing Noble and Nagel, we saw that there is no logical necessity that there should be some factor among the antecedent conditions which remains constant whenever G is the goal, but that doesn't show it is logically impossible that there might turn out to be one. Let us, then, suppose that the intrinsically-described antecedent conditions of all the G-directed behaviour we examine contain the factor K together with some other factor L which varies according to the nature of the series of actions which the agent performs in the attempt to achieve the goal. We break up C^n into $K+L^n$, etc.: what we are effectively supposing is that we have found some extensionally describable - perhaps neurophysiological - correlate (K) for the desire, and another (L) for the various relevant beliefs, instead of simply finding an extensionally describable correlate (C) for the desire-belief complex. This enables us to formulate a generalised non-teleological law about the behaviour of systems with K as antecedent condition:

$$(L)(\exists x)(\exists y)(\exists z)\ldots[(K+L)\rightarrow(x.((x.\sim H)\rightarrow y).((y.\sim H)\rightarrow z)\ldots)]$$

x, y, z of course range over actions.

This is much more promising as a non-teleological replacement for a teleological explanation; but we still have the

'and so on' problem. This, however, is no longer very serious. Since the present formula only contains variables ranging over actions, and not names of particular actions, we do not have to rely on the teleological account to know what comes next. Nor do we in stating any particular instance of the generalisation, because the substitution-instances for x, y, z, etc., and their number and order, will be determinate for any particular value of L, and there is no problem of the thing running on to infinity unless L is the intrinsic description of a set of beliefs that could belong only to God. The difficulty which remains is that in one case the agent may only be able to think of one thing to do, in another of ten, in another of fifty. He may even be able to think of nothing to do but still be in K: because he would do something if he could think of anything. So the length of the formula should differ for different values of L. What one wants to do, and can't, is to build in some condition to the effect that in a case where the agent can think of ten things to do then ten terms should be allowed to appear as members of the series x, y, z, etc., where he can think of fifty then fifty should, and so on. But the difficulty is not serious:

> Where K together with any L is the antecedent state of the system, if that antecedent state leads to two or more successive actions, these actions are related in this way: the conjunction of the occurrence of

the first with the non-occurrence of H leads to the second (provided the antecedent state still obtains and no interfering factors are present), and so on.

The 'and so on' here is harmless: we know how to go on. It is true that a single attempt to attain G may be divisible into several actions which are not related quite as separate attempts are. But they do fit this account, because if a single attempt can be divided into two actions, say opening a door and going through, then if by any chance it happened contrary to expectation that the goal (reaching cheese) was attained simply by the first of these then there would be no need to perform the second.

Since the concomitance of teleological and non-teleological explanations would have to be established empirically, it would always be logically possible that it should break down. The purposive version licenses the inference that any other actions which seem to the agent conducive to G will (in the absence of interfering factors) be performed. The non-teleological version licenses the inference that any other L which comes to be conjoined with K will (in the absence of interfering factors) lead to its own series of actions related in the specified way. However many cases we examine, even if we find each different L concomitant with a different set of beliefs about what is conducive to G, and conversely,

it may still turn out that there are cases in which an L occurs without its usually associated set of beliefs, cases in which a set of beliefs occurs without its usually associated L; and even cases of sets of beliefs to which no L corresponds at all, and L's to which no sets of beliefs correspond. If such breakdowns of concomitance occur, the non-teleological account will license one prediction and the purposive account another. But this, of course, one would expect, for they are after all different explanations.

So it is not possible to produce a non-teleological explanation which permits exactly the same inferences as a given teleological explanation, since they are different explanations - they are not logically equivalent. The extensionalist is therefore committed to saying that our ordinary explanations of human behaviour are wrong, and not merely in that they use intensional language to say what could and should be said extensionally. On the other hand, so long as the concomitances do hold the two accounts correspond pretty closely; and if in fact the concomitances never do break down then the difference between the accounts will lie solely in their supporting different counterfactuals[1] about what would happen if the concomitances did break down. In that case there will be no empirical way of deciding

1. See above, p. 300n.

between the two explanations, so neither the extensionalist nor the non-extensionalist will be able to make his case by appeal to experience; and at least it will not be possible to argue that extensionalism cannot explain the facts of human behaviour (including, therefore, language-use).

I said above that the adequacy of extensionalism could be defended by saying that human behaviour could be explained by non-teleological explanations which differed radically from teleological ones in that they cut up the phenomena quite differently. It is an empirical question whether these explanations can be found, but as I said I find it implausible to suppose they can, since it would involve a massive coincidence. The extensionalist, being unable to treat the teleological explanations as equally valid, would then have to hold that our ordinary purposive accounts of human behaviour are completely wrong. And that they should be completely wrong is, I suggested, very implausible indeed. But extensionalism would have maximum plausibility if non-teleological explanations corresponded so closely to our teleological ones that there was a one-one correlation between them. It does seem possible for the correspondence to be very close: so close as to enable all the same predictions to be made about actual events, though of course it breaks down over certain counterfactual possibilities. Whether such a correspondence can in fact be set up is again

an empirical question. It does not involve the same massive coincidence needed in the other case; many people would still consider it highly unlikely that the required concomitances should be discoverable, but in view of the sort of advances biology and neurophysiology have made the possibility must be taken seriously. Extensionalists, of course, are notorious for the trust they put in science. If the correspondence can be set up, the extensionalist is at least not committed to dismissing ordinary purposive explanations as <u>completely</u> wrong, for they will still be right in the important respect that they yield the right predictions. He will presumably hold that when I say "The rat did it because he wanted the cheese" I am somehow banking on the fact that some explanation of the non-teleological negative feedback type is true, a fact of which I have only "<u>une perception obscure et confuse</u>".

This still seems to fall short of being an entirely plausible thing to say. But in calling it implausible we cannot base our allegiance to our purposive explanations on the fact that they work. If the facts turn out as the extensionalist hopes, then however well-confirmed a particular purposive explanation may be, the corresponding non-teleological explanation will be exactly as well confirmed. Just as one might ultimately have to defend one's preference for "All grass is green" as against "All grass is grue" not by appeal to evidence but by saying that we are the sort of people for whom "green" is a primitive predicate and "grue" defined,

so here one could say that we are the sort of people to whom purposive explanations are natural. But humanum est errare: extensionalism may be right, even though it makes us say things that don't come naturally.

It seems therefore that the viability of extensionalism depends on an empirical question, whether the behaviour we call purposive can be satisfactorily explained in extensional terms. If the answer to this question is affirmative, the extensionalist's case looks a strong one. His opponent might defend purposive explanation by calling it just as valid as its non-teleological counterpart; but it is difficult to see how the argument could now be settled, except by bringing on considerations of conceptual economy; and certainly if one does this the victory lies with the extensionalist. It is not as though the extensionalist were proposing a conceptual scheme alternative to our own; he is proposing an unusual view about what the structure of our conceptual scheme really is.

The intensionalist might argue that even if human behaviour were entirely explicable without invoking teleology languages could still be intensional, in any or all of the three ways. This would be wrong; it is not only necessary, but sufficient, for the truth of the thesis that human languages cannot accommodate intensional expressions, that human language-use should be susceptible of non-teleological explanation. If language-use can be explained in non-

teleological terms, how could one acquire the notion of a possibility? Let us imagine that the linguistic behaviour of the speakers of French is explicable extensionally, while that of the speakers of English is not. Then for any particular value of p there is a discoverable set (call it S) of causal conditions which produce utterances of the sentence "Il est possible que p". Now if we want to find out the meaning of French sentences and translate them into English, ought we not to translate "Il est possible que p" by "S obtains", which is extensional in English, instead of by our non-extensional, modal sentence "It is possible that p"[1]? "Il est possible que p" is uttered whenever and only when "S obtains" is true: the two sentences have the same assertibility and confirmability conditions, and given the efficient-causal account of meaning there is nothing that can be meant by saying they mean different things. So a language cannot contain expressions like "It is possible that p", taken as we normally take it, if the semantics of that language are non-teleological; we normally take it as neither truth-functional, transparent, nor predicate-extensional. But even if it were non-extensional in only one of these three respects the argument would clearly go through just as well.

1. "p" here is of course the English translation of the French "p".

If "I recognize the Morning Star" and "I recognize the Evening Star" are uttered in different circumstances, one of them will be equivalent to "S' obtains" and the other to "S'' obtains"; the appearance of opacity arises only because we mistakenly take "the Morning Star" and "the Evening Star" as isolable referring expressions.[1] And if it were objected to this that "I know that p is possible" or "I know that I recognize the Morning Star" is quite different from "I know that S (or S') obtains", these sentences could of course themselves be handled in the same way.

The same point, by the way, can be made about future-tensed sentences, unless one allows the utterance of "It will be the case that p" to be in part caused by the (future) event of p's occurrence. For then again, if S be the causal conditions of the utterance, "S obtains" will have the same meaning. Perhaps the extensionalist would be well advised to accept backwards-causation.

If this is true, it is true whether or not the non-teleological account of language-use really can be expressed

1. Thus although, as Quine says (Word and Object, p.50), "Indian nickel" and "buffalo nickel" might have different stimulus-meanings for a speaker who did not know that all Indian nickels are buffalo nickels and vice versa, one should not infer that because he does not use these terms interchangeably his language is not a transparent system. For it is not clear that he can be said to be referring to the coins rather than to the stimulations, which are admittedly different. As Quine himself says, "Where other languages than our own are involved, coextensiveness of terms is not a manifestly clearer notion than synonymy or translation itself". (Word and Object, p.54).

extensionally. If, contrary to what I have suggested above, it is not possible to preserve in a truth-functional language the core of the idea of causality, and if nevertheless it be true that the linguistic behaviour of a certain group can be explained causally and without invoking ideas of purpose, then the language of these people must still be transparent, predicate-extensional, and truth-functional (and so cannot contain its own, truth-functional, semantics), for just the same reason. So non-teleological explicability is <u>sufficient</u> to make a language extensional in all three senses, whether or not it can contain its semantics; but it is only <u>necessary</u> if we postulate that the language can explain itself, for it seems possible that although language-use required purposive explanation, intensional in all three ways, speakers might choose only to use the extensional parts of their language. This makes it clear that any extensional language which contains its own semantics must be extensional in all three ways; for it must be explicable non-teleologically, i.e. in terms of efficient causes (or something weaker), and so cannot contain any intensional expressions.

The upshot of all this is, that the remaining issue between extensionalism and intensionalism is the empirical question of how linguistic behaviour is to be explained. It is not possible to show <u>a priori</u> that behaviour explicable purposively must also be explicable in the non-teleological

way, in which the antecedent conditions are specified intrinsically, without reference to goals. It is equally impossible to show a priori that behaviour cannot be explained in both ways. More would have to be done before one could be quite certain that extensionalism contains no incoherence; but the main objection to it has been its inability to cope with human behaviour. It seems quite possible logically that this difficulty should be overcome; though one may be excused for not thinking it very probable that it will be. But if it can, if for the explanation of human language-use no reference to ~~the~~ purpose is necessary, then it cannot be possible for human beings to acquire a non-extensional language; those features of our languages that we think not to be extensional must be just muddled, useful only so far as they correspond to and can be replaced by things that can be said extensionally.

Conclusion: Analyticity

We were considering extensionalism in order to find out what to make of the extensionalists' dismissal of the analytic/synthetic distinction as unintelligible. We have reached the conclusion - admittedly not a very satisfying conclusion - that the assessment of extensionalism ultimately comes down to a question of empirical fact. It follows that

the validity of their rejection of analyticity likewise depends on the same question of fact. It sounds paradoxical to make a question of meaningfulness depend on a question of fact, but really it is not: if we consider a race of beings different from ourselves, there is nothing surprising about its being an empirical question whether they can understand non-extensional ideas.

If extensionalism doesn't work, its objection to analyticity goes. Obviously we cannot conclude immediately from this that the analytic/synthetic distinction is intelligible. A system can be non-truth-functional, opaque, and predicate-intensional without containing any concept of analyticity; since causal statements can be taken as non-extensional in all these respects, an example would be provided by a system containing causal but not logical modalities.

However, it would be easy to introduce logical modalities like "It is analytic that...." into a system like this; the system already contains the materials for the introduction, in a way that an extensional system does not. For if extensionalism doesn't work, and if our system is to be capable of expressing explanations of behaviour, it must be capable of expressing the notion that X desires G and believes B conducive to G, for this is what a purposive explanation consists in. So speakers of this language must sometimes find it reasonable to ascribe a particular belief

to a speaker; and since purposive explanations are to be used to explain his behaviour in uttering sounds, it must sometimes be reasonable to ascribe to X the belief that p even without going on his utterance of the sound "p". This in turn means that they must be making use of criteria for the identity of beliefs which are different from criteria for the identity of utterance-types; so there is nothing to prevent two different utterance-types being expressions of the same belief. Nor is there anything to prevent two different utterance-types being standardly used to produce the same belief in the hearer. These things may never in fact happen; in our own language genuinely synonymous sentences are very rare. What matters is that they could easily be made to happen, and speakers would already have the machinery to deal with the situation if they did. It is but a short and easy step from this to their actually having a word for "synonym" in their language. Then having acquired one term in the circle of expressions outlawed by Quine they can easily introduce others, defining analyticity in one or other of the ways canvassed on p.$^{250}_{A}$ - perhaps calling a statement analytic if and only if it is synonymous with a statement of the form "If p and q then p".[1]

Quine himself grants that in order to translate one

1. Cf. Word and Object, p.65. Perhaps this definition contains what is worthwhile in Kant's (A7/B10).

language into another, and indeed in order to effect translations from one part of the same language to another or to understand the speech of someone else, we have to make "analytical hypotheses". An analytical hypothesis correlates two sentences and says that one is to be treated as a translation of the other; it is, in effect, an hypothesis that the two sentences are synonymous. And not merely stimulus-synonymous, sharing the same stimulus$^{u}_{\lambda}$-conditions; for he does not consider it necessary to invoke analytical hypotheses when translating observation sentences, which are those sentences the stimulus-meaning of which does not vary (greatly) from person to person.[1] Analytical hypotheses correlate sentences which are <u>not</u> stimulus-synonymous, and allow us to translate one by the other. This notion of translatability, which gives analytical hypotheses their content, must therefore belong firmly with synonymy and analyticity in the circle of intensional semantical terms; so that in saying, as he does[2], that the making of analytical hypotheses is indispensable to us, Quine is admitting our practical need to use intensional terms in understanding one another.[3] But he insists that analytical hypotheses

1. <u>Word and Object</u>, pp.43, 68.
2. <u>Ibid.</u>, p.75.
3. Cf. <u>ibid.</u> section 45.

are always and only hypotheses; alternative sets of them are always equally possible, and there is no such thing as one particular set's being right.

> The point is not that we cannot be sure whether the analytical hypothesis is right, but that there is not even...an objective matter to be right or wrong about... Rival systems of analytical hypotheses can conform to all speech dispositions within each of the languages concerned and yet dictate, in countless cases, utterly disparate translations; not mere mutual paraphrases, but translations each of which would be excluded by the other system of translation.[1]

For this reason, he says, it would be wrong to think that the idea of analytical hypotheses gives a sense to the elusive idea of synonymy.

> We could claim this only if no two conflicting sets of analytical hypotheses could be tied for first place on all theoretically accessible evidence.[2]

This reminds one of Kant on regulative ideas. Quine is telling us that although we can never actually be right in asserting that two expressions are synonymous (unless we just mean that they are stimulus-synonymous) we must nevertheless all the time, in our attempts to understand one another, act as if we could be. We can never have speculative knowledge about synonyms, but the finding of

1. Ibid., p.73.
2. Ibid., p.75.

them is "set as a task for the understanding". In such circumstances it is not obvious that Quine is right to dismiss the ideas of synonymy and analyticity as unintelligible; they must have some sort of sense, if this use is to be made of them, and the verificationist doctrine that p must lack sense if there are no circumstances under which its truth-value can be definitely determined requires a lot of discussion. It is no argument against it, perhaps it is even an argument for it, that it would make Kant's regulative principles strictly senseless as well, except when interpreted as pure imperatives; but it is not clear that the unverifiable hypotheses that the universe is symmetrical, or constantly doubling in size[1], are really without content. Nor is it clear that one ought to adopt a positivistic attitude to scientific theories. These questions must await a developed theory of meaning, for it is only in the light of such a theory that the claims of verificationism can finally be assessed.

If, however, the analytical hypotheses we employ are really bits of muddled intensionalism, useful only in virtue of the fact that in giving purposive explanations we are manifesting an indistinct awareness of corresponding, and

1. Assuming that the laws of nature change in whatever ways may be required to preserve the unverifiability.

true, non-teleological explanations; then Quine ought to hope that, even if we cannot do without them, one day our more rational descendants may be able to - although that may mean their abandonment of the idea of sentence-by-sentence translation of one language, or idiolect, into another.

I do not doubt that there is some truth in Quine's thesis of the indeterminacy of translation. How much, it is hard to say. The answer depends, to some extent, on settling the question about the principle of verifiability, since on a positivistic account two apparently different theories which have exactly similar empirical consequences will not differ in meaning[1], and so could be mapped onto one another, or onto a single equivalent theory in a different language, by a variety of alternative sets of analytical hypotheses. But, whatever the answer is here, there is no reason to believe that Quine is right to say that it can never be settled that two expressions are synonymous (except where only stimulus-synonymy is involved). Indeed, if he were right, and if it is also true that language-use cannot adequately be explained non-teleologically, the outlook would be serious, for it would then be impossible

1. Euclidean geometry with one set of physical correspondence-rules, and Riemannian geometry with another, would be such a pair of theories. Cf. L. Sklar, "The Conventionality of Geometry", American Philosophical Quarterly Monograph 3: Studies in the Philosophy of Science (Oxford, 1969).

intelligibly to give an intensional account of meaning and we should have to admit that no explanation of human language-use could be given at all - though we could agree with him that it is useful to act as if it could.

What precisely the criteria for analyticity and synonymy are is a difficult question which I shall not go into. It, too, cannot be finally answered except by giving a full theory of meaning. To say, for example, that (affirmative) analytic judgements are those in which the connection of the predicate with the subject is thought through identity[1], or to give any of the other traditional definitions, is to provide no more than a vague approximation. But it is reasonable to be confident that it is not in principle impossible to set out appropriate criteria. We are quite capable of telling, in certain circumstances, that two people have the same desire; and quite capable of telling, in certain circumstances; that they both share a particular belief; so there should be no difficulty in principle in telling that two expressions are used for the same linguistic purpose.

It is true, of course, as the believer in thorough translational indeterminacy will not hesitate to point out, that beliefs and desires are always manifested together. So

1. A 7/B 10.

in ascribing a belief to someone on the basis of his behaviour[1] I must make some assumptions about his desires, and in ascribing desires I must make assumptions about his beliefs. But this does not condemn me to continual uncertainty. The rat in the example considered above[2], stroking its whiskers with a lump of cheese before it, may be very hungry but performing a ritual believed to be an essential prelude to safe eating; the gardener mowing the lawn may be acting on the belief that pushing a mowing-machine for twenty minutes will cause him to arrive on the moon. And with sufficient imaginativeness (one might require quite a lot) the rest of their behaviour could be construed in such a way as to fit in with this. For normally one would expect such desires and beliefs to manifest themselves more straightforwardly in other behaviour of the gardener and the rat; but they might have so many strange desires and beliefs that it was rarely right to give their behaviour the interpretation we should consider most natural. We can never have sufficient evidence to rule out the logical possibility that these interpretations are correct[3], and

1. Even if the behaviour is the utterance of a linguistic report. For these are made for some purpose, and with some beliefs about hearers and meanings of words. But we should not appeal to linguistic reports anyway, since we are raising questions about how these can be understood.

2. P.301.

3. Some would object to this (especially, perhaps, in the case of the rat: cf. above, p.302 n.1.). If they are right, that suits my argument even better: there is no room at all for indeterminacy of translation in such cases.

to this extent alternative analytical hypotheses are indeed always possible; but this should not disturb us, any more than it should disturb us that we can never have sufficient evidence to rule out the logical possibility that my typewriter should occasionally sprout wings when nobody is looking at it. The logical possibility of error[1] dogs all our beliefs, but this should not condemn us to total scepticism. And under certain conditions ascriptions of desires and beliefs to others can be as certain as any claims can be which go beyond what one is immediately aware of.

Consequently, I think it is right to conclude that the question of the intelligibility of the analytic/synthetic distinction, and hence of the legitimacy of the whole way of doing philosophy which depends on that distinction, can only be settled empirically. Though, seeing how much would be lost if extensionalism were vindicated by the facts, it is perhaps reasonable to express a faith that they will favour intensions. This is a lame defence of analyticity, but I think it is all that one can do. Of course, my argument in this chapter has itself made use of non-extensional concepts, but this is fair enough: it is only by tentatively assuming that such concepts do make sense that we can find an issue to discuss at all. My argument has

1. Even if only of verbal mistakes: cf. above, pp. 138ff.

in fact been transcendental in character, though more like a transcendental exposition than a transcendental deduction. For it has sought to show what the conditions are for intensional and extensional languages, and has not tried to show intensional notions to be necessary for anything that could be called language, or experience, at all; indeed, if it is correct, this could not be shown a priori, unless it were made true by definition (as it was in the first two chapters, and as is for many purposes convenient). Like Kant's transcendental expositions this argument is of little value against the really pig-headed sceptic, here the extensionalist who continues to deny that there can be intensional languages. We can never force him from his position, for, as we saw, even if it were clear that human behaviour could only be explained intensionally he always could hold that therefore it could not be explained at all. Moreover, whatever happens we can never attain complete certainty that it is not possible to explain human behaviour extensionally: it may always be the case that there is a discoverable non-teleological explanation, which we have just failed to find. A normal transcendental argument tries to convict the sceptic of rendering his own doubt unintelligible; here the intelligibility of the sceptic's position is not in question whereas that of analyticity is.

It would be a fantastic distortion of Kant to suggest

that my argument in this chapter, or any substantial part of it, is to be found in his works. He never, so far as I know, considered the concept of analyticity to require any defence, or further explication than he gave of it in the Critique. He would not have agreed that it is an empirical question whether human behaviour is susceptible of both kinds of explanation: he considered it certain that it is; for he thought it transcendentally necessary that every event in the world of appearances should have a cause, and yet as involved in the concept of a rational being that its actions should be free; though since by "free" he means not just "purposively explicable" but also "not determined by empirical laws of cause and effect" his position has difficulties mine avoids. Since he held that rational beings must be free presumably he would have accepted that freedom is a necessary condition of having a language capable of expressing logical modalities.

I have tried to explain and to defend as best I can Kant's method of doing philosophy, by transcendental and metaphysical arguments and by conceptual analysis in general. It seems unsatisfactory to leave the matter here, for at each stage the defence has been less complete than one might have hoped. Transcendental arguments are of some use, but much less than is often supposed; even less general arguments delineating the features of particular conceptual schemes

tend to be disappointing; the concept of analyticity is condemned to the perpetual threat of redundancy as science advances. On the former issues the position could be improved by the discovery of new and better transcendental and metaphysical arguments. But on the question of analyticity I cannot see how a defence can go further. Philosophy, as conceived in a Kantian way, can defend itself only up to a certain point. The remaining question, odd though it may seem, can only be dealt with empirically.

B I B L I O G R A P H Y

E. Adickes, Kant als Naturforscher. 2 vols. Berlin, 1924.

J. Bennett, Kant's Analytic. Cambridge, 1966.

R.B. Braithwaite, Scientific Explanation. Cambridge, 1953.

*G. Buchdahl, "The Conception of Lawlikeness in Kant's Philosophy of Science". Forthcoming, in the proceedings of the Third International Kant Congress.

R. Carnap, "Empiricism, Semantics and Ontology". Revue Internationale de Philosophie, 1950.

*K. Cramer, "Zur systematischen Differenz von Apriorität und Reinheit in Kants Lehre von den synthetischen Urteilen a priori". In Subjektivität und Metaphysik: Festschrift für Wolfgang Cramer. Frankfurt a. M., ?1970.

*K. Cramer, "Non-Pure Synthetic A Priori Judgements in the Critique of Pure Reason". Forthcoming, in the proceedings of the Third International Kant Congress.

D. Davidson, "Causal Relations". Journal of Philosophy, 1967.

D. Davidson and J. Hintikka (eds.), Words and Objections: Essays on the Work of W.V. Quine. Dordrecht, c.1969.

N. Goodman, Fact Fiction and Forecast. 2nd ed., New York, c.1965.

H.P. Grice, "Meaning". Philosophical Review, 1957.

H.P. Grice, "Utterer's Meaning and Intentions". Philosophical Review, 1969.

H.P. Grice and P.F. Strawson, "In Defence of a Dogma". Philosophical Review, 1956.

Kant's gesammelte Schriften. Ed. by the Kgl. Preussischen Akademie der Wissenschaften, and its successors. Berlin, 1900- .

Immanuel Kant's Critique of Pure Reason. Trans. by N. Kemp Smith. London, 1929.

G.W. Leibniz, Die philosophischen Schriften. Ed. by C.J. Gerhardt. 7 vols., Berlin, 1875-90.

L. Linsky (ed.), Semantics and the Philosophy of Language. Urbana, Illinois, 1952.

E. Nagel, The Structure of Science. London, 1961.

D. Noble, "Charles Taylor on Teleological Explanation". Analysis, 1966-7.

D. Noble, "The Conceptualist View of Teleology". Analysis, 1967-8.

H.J. Paton, Kant's Metaphysics of Experience. 2 vols. London, 1936.

D.F. Pears, "Universals". Philosophical Quarterly, 1951.

P. Plaass, Kants Theorie der Naturwissenschaft. Göttingen,1960.

W.V. Quine, From a Logical Point of View. 2nd ed., Harvard,1961.

W.V. Quine, Word and Object. M.I.T., 1960.

W.V. Quine, The Ways of Paradox. New York, c. 1966.

K. Reach, "The Name Relation and the Logical Antinomies". Journal of Symbolic Logic, 1938.

P.F. Strawson, Individuals. London, 1959.

P.F. Strawson, The Bounds of Sense. London, 1966.

B. Stroud, "Transcendental Arguments". Journal of Philosophy, 1968.

C. Taylor, The Explanation of Behaviour. London, 1964.

C. Taylor, "Teleological Explanation". Analysis, 1966-7.

H. Vaihinger, Kommentar zu Kants Kritik der reinen Vernunft. 2nd ed. Berlin, Stuttgart, Leipzig, 1922.

H.J. de Vleeschauwer, La Déduction transcendentale dans l'OEuvre de Kant. Paris, Antwerp, the Hague, 1934-7.

L. Wittgenstein, Philosophical Investigations. 2nd ed., Oxford, 1958.

L. Wittgenstein, Remarks on the Foundations of Mathematics. Oxford, 1956.

C. Wolff, Philosophia Prima sive Ontologia. New ed., Frankfurt and Leipzig, 1736.

*These papers, which complement some of the things I say in chapter II, were not available to me before that chapter was finished.