Indian Philosophical Quarterly, Vol. VIII, *No.* 2, *Jan.* (1981).

THINGS IN THEMSELVES AND SCIENTIFIC EXPLANATION

Some time ago, Walsh found a witty simile to sum up the objections which, ever since Hegel, have greeted Kant's notion of the thing in itself : "Like some disreputable relative who persists in disgracing the family on whose charity he lives, it can neither be rendered respectable nor yet be got rid of altogether".[1] More recently Strawson has done his best to show us how to get rid of it as graciously as possible, by disentangling an analytically provable "metaphysics of experience" from what he sees as the unacceptable framework of transcendental idealism in the *Critique of Pure Reason*.[2] On the other hand, Wilfrid Sellars has said of Kant's distinction between appearances and things in themselves that "if it is ignored or misinterpreted little remains but a fancy-dress costume for quite unKantian, if fashionable, ideas", and has tried to render it respectable by metamorphizing Kant's transcendental idealism into his own version of scientific realism.[3] I hope to improve on Sellars by suggesting a role for that untrustworthy veteran, the thing in itself, which is not only more Kantian, but perhaps more respectable even by today's demanding standards.

The first point (which I share with Sellars) is to take the distinction between appearances and things in themselves , or phenomena and noumena, as marking a division not between two realms of entity, but between two kinds of property of the *same* things. Kant often expresses it in this latter way himself, most emphatically where he says that the distinction which his work has shown to be necessary is "between things as objects of experience and those same things as things in themselves" (B xxvii).[4] The thought that the things we perceive may have properties (indeed, constituents) which we cannot perceive is one that a modern philosopher should hardly choke on, for it is what physics and chemistry would seem to have proved. However, Kant wanted to say not just that perceptible things have *imperceptible* properties, but natures of which we can have no *knowledge* whatsoever (although considerations involving practical reason are supposed to justify certain hypotheses about them); yet the old objections apply to the suggestion of unknowable properties just as much as to unknowable

things. Admittedly, Kant sometimes mentions the logical possibility of some superhuman sort of being, with an "intellectual intuition", not limited by our forms of sensibility, who could know things as they are in themselves (B 71-2), but since he adds that we cannot comprehend the possibility of such an intuition (B 307), this is only to to replace one mystery with another.

The main feature of Sellars' theory, as I understand it, is to replace God (or whoever it might be that emjoys intellectual intuition) by science, or at least by a Peircian ideal of the future outcome of scientific enquiry.[5] Kant's "appearances" are taken to be the contents of the spatio-temporal material world as described in perceptual terms, using our ordinary everyday concepts (what Sellars calls the "manifest image" of the world); whereas Kant's "things in themselves" are taken to be the contents of the world as described by the theories of the physical sciences (Sellars' "scientific image"). This interpretation renders things as they are in themselves knowable (by us) after all, much to their philosophical benefit. It seems to cast them in the role of Locke's real essences, those "real internal, but generally in substances unknown, constitution(s) of things, whereon their discoverable qualities depend",[6] a role which will be all the more congenial now that we can moderate Locke's pessimism about our human abilities to discover such essences, in the light of the progress of microphysical explanation since the seventeenth century. We should also recognize that on this rehabilitation of things as they are in themselves, they will have whatever spatial and temporal properties that scientific theory finds it appropriate to attribute to them. Kant's thesis that they are neither in space nor in time will therefore have to be modified; but we can point out that spatio-temporal talk in physics (we hear of four-dimensional space-time, of space being curved and non-Euclidean, and even of time going backwards) has been getting ever more conceptually different from our everyday talk of the spatial and temporal relations of perceptible objects. So even here we have a partial vindication of Kant, for we can say that things as scientifically described do not have the spatio-temporal properties we ordinarily conceive of.

If we buy this revision of Kant from Sellars, we can use it to meet the two conditions Strawson rightly lays down for a significant contrast between appearance and reality.[7] Firstly, we have

already built in *identity* of reference between appearances and things in themselves, by interpreting the distinction as between two kinds of peoperty of the same set of things. Thus, in Eddington's famous example, there are not really two different tables, that of the common man and that of the atomic physicist,[8] there is of course one table, with the everyday description of it as solid being perfectly compatible with the scientific theory of it as consisting of tiny particles separated by empty spaces. Secondly, we can offer at least one fundamental reason why the scientific accounts may be said to give a *corrected view* vis-a-vis the ordinary perceptual descriptions, namely that the former can be employed to *explain* the latter, but not vice versa. As is often said, science aims to produce accounts of the material world which are more "objective" in this sense; that given the scientific descriptions of things (and of our human sense-organs) we can understand *why* the world should appear as it does to beings perceptually equipped as we are. But we cannot in the same way infer scientific theories from ordinary perceptual descriptions, although no doubt the latter must give the most ultimate evidence for accepting the former. There is, then, in the scientific project a necessary asymmetry of explanation between theory and observation; and the successes of the explanatory project are amply demonstrated in the technological control over the world that science has (for better or for worse) given us. But success in explanation should not lead us to deny the truth of what is explained — we do not have to follow Eddington into saying that the table is not really solid, brown, hard, etc. Sellars' neo-Kantian thesis would seem to be that the table and its perceptible properties are empirically real (not dependent for their existence on any particular perceiver or perceptual conditions) but transcendentally ideal (composed of, or explicable in terms of, entities or properties which are not perceptible, and not describable in our common sense vocabulary).

But however attractive this application of the Kantian terminology may seem, it would be idle to pretend that it is Kant's own. The whole trend of his critical philosophy is to confine science, along with ordinary perception, to knowledge of things as they appear, and to deny the very possibility of our *knowing* (rather than believing postulates about) things as they are in themselves. One passage where his difference from Sellars emerges very forcefully is that in which he rejects the suggestion that observational

astronomy gives us an account of the sensible world while theoretical astronomy yields knowledge of an intelligible world : in Kant's view "such a twisting of words is a merely sophistical subterfuge" (A 257/B 313). The question which he had discussed in the previous pages is whether we can *apply* the concept of a noumenon, " a thing which is not to be thought of as object of the senses but as a thing in itself" — and he answered it emphatically in the negative. The concept is not selfcontradictory, but by its very nature we cannot apply it to anything; it has only a "negative employment", as a "limiting concept" to "curb the pretensions of sensibility" (A 254-4/ B 310-1).

It may look, therefore, as if Sellars' charge of misappropriating Kant's terminology as fancy-dress for unKantian but fashionable ideas, may apply against himself. But before we find him guilty we had better examine more closely what he has done. Although he sometimes talks of *the* scientific image of the world, he is (as any competent philosopher ought to be) very much aware of the many-layered complexity, and the changing and developing nature of scientific theories. There are at least three aspects to be distinguished here — complexity, incompleteness and revisability. Firstly, there are many levels of scientific theory; descriptions in terms of populations, individual animals, organs, cells, genes, molecules, atoms and sub-atomic particles (or forces, or packets of energy, or warps in sapce-time) are consistent with each other, although not necessarily interdefinable. Secondly, these descriptions may be incomplete, not only in that there is room for more at each level, but in that there is no bar to the addition of further levels of theory (whether at either end, or in the spaces in between); in particular, induction over the history of physics gives us no reason to suppose that the subatomic is the most fundamental level there is. Thirdly, as we have come to realize through the work of Popper, Kuhn and others,[9] science does not progress simply by addition to existing results, for at any level revolutions can occur, in which crucial elements in hitherto accepted theory are rejected, or are allowed to survive into the new intellectual regime only in considerably modified form.

Given that we have to acknowledge the truth of this less superficial picture of science, what does Sellars make of it? He claims that in Chapter V of *Science and Metaphysics* he has decisively

clarified the relationship between what he has called the "manifest" and the "scientific" images of what there is. What we get there is an attempt to analyse the notion of truth within conceptual structures, such as scientific theories, which are subject to change and evolution. The fundamental notion of truth is said to be semantic assertibility (IV. 26), and since assertion must be done in some language or conceptual scheme (V. 48), the unqualified sense of 'true' can only be semantic assertibility in *our* conceptual structure (V. 50). But how can we reckon with the fact of change in conceptual structures (CS's for hort)? Sellars appeals to the intuitions that one concept can be a development of another (V.52), that one CS can be more "adequate" than another (V. 54, 65), and that expressions in different CS's can refer to the same things in the world, although picturing them in different ways (V. 62, 67). But I see no attempt in this chapter to give any criteria for *applying* the crucial notions of common reference between CS's, and greater adequacy of one CS over another; presumably Sellars is appealing to the natural view of scientific progress, that a later theory gives a better account of the same subject-matter than an earlier one — but a proper defence of this common sense position would have to find replies to the philosophical doubts which have been raised about it.[10] Sellars wants to go further than this, however, for he says that "we are haunted by the ideal of *the* truth about the world" (V. 55), and, far from exorcising any such ghost, he tries to find a respectable home for it in his philosophy. So he invites us to "conceive of a language which enables its users to form *ideally* adequate pictures of reality", to call it "Peircish" (V. 69), and to "conceive of Peircish speakers as a successor generation in a continuing scientific community" (V. 72). The concepts of ideal truth and of what really exists are to be defined in terms of this Peircian conceptual scheme (V. 75, 95).

In this way Sellars claims to have improved on C. S. Peirce's account of truth and reality : "The opinion which is fated to be ultimately agreed to by all who investigate is what we mean by truth, and the object represented in this opinion is the real" (where 'fate' means "that which is sure to come true, and can nohow be avoided").[11] Sellars does not require that there ever *be* a Peircish community (wisely, for even if the ideal were fulfillable, many mishaps could prevent it); he only asks us to conceive of the *possibility* of such an ideally adequate language. In requiring such a

conception in his account of the scientific project, he has been followed more recently by Bernard Williams, who in his book on Descartes sketches a connection between the very notion of knowledge (as knowledge of a reality which exists independently of that knowledge) and what he calls "the absolute conception of reality", namely, a conception which would enable us to understand how reality is related to *all* possible conceptualizations of it.[12] Williams endorses the search for such a conception as a reasonable goal, for the physical if not for the social sciences, although he argues that we must sharply distinguish this from a quest for Cartesian certainty. (Rather surprisingly, Williams thinks that realism in this sense is more fundamental to "the Cartesian outlook" than Descartes' beliefs in God, dualism, and the need for certainty — but one might rather say that the idea of science searching for *the* truth about the world has haunted the whole of Western philosophy since the seventeenth century).

But however attached we may be to such an ideal of absolute truth as expressible in some possible future scientific conceptual scheme, we must face up to hard philosophical questions about the very meaningfulness of such a notion, and the role it is supposed to play in scientific enquiry. Certainly, we need better accounts of the nature of progress in science (as noted above) — but it would surely seem that we can describe one theory as being superior to another in a certain domain without bringing in any notion of a *completely* adequate theory. Peirce and Sellars try to define this latter as an actual or possible limit to a series of increasingly adequate theories. But this notion of an ideal limit seems subject to contradictory demands : on the one hand, it should be a member of the series of theories, otherwise how can it be said to be more adequate than all the previous members (by whatever the criteria of theory-comparison should be), but on the other hand, it needs to be outside the series, for it is supposed somehow to escape the revisability which seems characteristic of scientific theories. If the very notion is thus contradictory, how can it play a defensible role in any kind of enquiry?

We have here an antinomy of characteristically Kantian form about a limit to a series which is required both to be and not to be a member of the series — and I want to propose a Kantian-style resolution of it. In fact, I think we must develop two quite different

notions — relative and absolute — out of Sellars' rehabilitation of Kant's thing in itself.[13] The relative notion is that relation which holds between two descriptions or theories when one of them gives a better account than the other of a certain subject-matter, 'better' meaning here that it explains everything asserted in or explained by the other, and some other phenomena too. To describe this relation as that between appearance and reality would perhaps be to give it too much metaphysical dignity, since the explaining theory may itself get explained; and, at the other end, nothing excludes the explained description being explanatory of some other account (the notion of *observation*, of descriptions based only on untutored perception, might be thought to exclude this, but there is reason to suppose that there can be no observation which is not imbued with some interpretation.)[14] But even if this is true, it does not prevent us applying the relative notion of one account being explanatory of another. We *could* hijack the Kantian distinction between things as they appear and as they are in themselves to mark the terms of this two-place relation, but there would be little point in doing so, since Kant's use of this terminology was so different. After all, he had his own vocabulary for what sounds like the explanatory relation between theories, where he talks in the Transcendental Dialectic of the condition for a given conditioned (A 307/B 364 ff).

The absolute notion I have in mind, though , does relate much more closely to Kant's thinking. Given the relative notion of one theory being better than another, it seems that we can at least form the concept of a theory which is ideally adequate, in the double sense of being the most *fundamental* theory there can be (in that all other physical theories are deducible from it, given appropriate bridging principles), and of being *unrevisable* (which seems to amount to saying that it is unqualifiedly true). But in thus idealizing our notion of scientific theory, we must realise that we have made it too perfect for this imperfect world, for we can see that even if we could (logically) have such a theory, we could not *know* that we had it, since what guarantee could there be that unexpected evidence would never turn up to upset our best-entrenched theory, or that some revolutionary genius may not come up with a better explanation of everything we accepted so far? So we could never apply this notion of an ideally adequate theory; yet it does not seem to be selfcontradictory (the antinomy noted above arose from

requiring it to be attainable by us). But we *can* apply the relative, comparitive, notion : we can say that one theory is more explanatory than another, we can tell when theoretical unifications take place in science, and we can even conceive of a future in which we might have a well-established, completely unified, physical theory (although for the reasons just noted, we could never exclude the possibility of further revision).

So we can make *progress* towards an ideally adequate theory, even though we could never reach it. In fact, the notion seems to have precisely the status of Kant's "transcendental ideas"—concepts of reason "to which no corresponding object can be given in sense-experience", which "overstep the limits of all experience" in that no object adequate to them can ever be found. But for Kant such ideas "are not arbitrarily invented; they are imposed by the very nature of reason itself, and therefore stand in necessary relation to the whole employment of understanding" (A 327 / B 384). For as he later explains in more detail, transcendental ideas have an excellent, and indeed indisputably necessary, regulative employment, namely, that of directing the understanding towards a certain goal upon which the routes marked out by all its rules converge, as upon their point of intersection. This point is indeed a mere idea, a *focus imaginarius*, from which, since it lies quite outside the bounds of possible experience, the concepts of the understanding do not in reality proceed; none the less it serves to give to these concepts the greatest possible unity combined with the greatest possible extension. (A644 / B672) The suggestion of this paper is that Sellars' scientific rehabilitation of the notion of things as they are in themselves can play this Kantian role of regulative ideal (and, perhaps, play it rather better than the ideas of soul, universe and God do in Kant's original version of the transcendental drama).[15]

University of St. Andrews, Scotland. **Leslie Stevenson**

NOTES

1. W. H. Walsh, *Reason and Experience* (Oxford : Clarendon Press 1947), pp. 69-70.
2. P. F. Strawson, *The Bounds of Sense* (London : Methuen 1966).

3. Wilfrid Sellars, *Science and Metaphysics : Variations on Kantian Themes* (London : Routledge 1968), p. viii.
4. All quotations from the *Critique of Pure Reason* are from the translation by Norman Kemp Smith (London : Macmillan 1929).
5. Sellars, *op. cit.*, II. 49ff, V. 79ff.
6. John Locke, *Essay Concerning Human Understanding*, III. iii. 15ff.
7. Strawson, *op. cit.*, p. 250.
8. A. Eddington, *The Nature of the Physical World* (Cambribge : University Press 1927; London : Dent & Dutton 1935) Introduction.
9. Karl Popper, *The Logic of Scientific Discovery* (London : Hutchinson 1959), T. S. Kuhn, *The Structure of Scientific Revolutions* (Chicago : University Press 1962).
10. See, for example, Kuhn, *op, cit.*, and Paul Feyerabend, *Against Method* (London : 1975, verso 1978).
11. C. S. Peirce, ' How to Make Our Ideas Clear ', §407, in *Collected Papers* (Cambridge, Mass. : Harvard University Press 1934), Vol. V, p. 268.
12. Bernard Williams, *Descartes* (London : Penguin 1968), see the index references to " absolute conception of the world ".
13. I here develop a theme I broached some time ago in ' Immanent Transcendence : Variations on a Logical Theme ', in *Religious Studies* Vol. 6 (1970), pp. 89-98.
14. See the works by Popper, Kuhn and Feyerabend mentioned above.
15. Various versions of this paper have been given in seminars at the Universities of St. Andrews, Aberdeen, Newcastle, Edinburgh; the London School of Economics; University College, Cardiff; and, in January 1980, at the University of Poona.
I have learnt from the discussions on all these occasions.