

# **SELF-REFERENTIAL ARGUMENTS IN PHILOSOPHY**

Steven Yates

*Auburn University*

## 1.

Does philosophy have an identity, method, and subject matter enabling one to differentiate it from the sciences and from other activities such as politics, literature, poetry, art, and so on? Can it have results that are distinctly its own? The history of philosophy records many efforts to isolate the method, subject matter, and results of philosophy. Their results, however, seem at first glance to have been ephemeral: time and time again philosophy has been uprooted and transformed. Philosophy -- or supposedly central branches of it such as metaphysics -- has been declared dead on numerous occasions (by Hume, Comte, Wittgenstein, to name just three authors of such declarations). But the idea of a distinctly philosophical method won't go away; over the past two hundred years alone it has reappeared as the study of transcendental subjectivity, logical form, "marks of the mental," ordinary language, scientific method, and so on.

Today, however, philosophy as an autonomous discipline faces perhaps its greatest crisis, embodied in the recent critiques of foundationalism. Partly as a result of the reawakened interest in the history of philosophy, showing how the discipline has changed through time; partly due to the new interest in cultural diversity and in listening to the voices of those alleging disenfranchisement by the central strains of western philosophy;

and partly to the gradual convergence of fields like artificial intelligence, computer science, and the neurosciences into an all-embracing science of cognition; scientific philosophers, political activists, and purveyors of *difference* have again declared philosophy dead and are presently dividing up the spoils.

In this paper I wish to re-examine several recent results of a controversial but widely-used form of metaphysical argument -- or more precisely, class of arguments -- which has kept resurfacing throughout the history of philosophy despite the changes philosophy has undergone. These arguments all involve reflexivity or self-reference in one form or another. It is the apparently perennial nature of this class of arguments, I will finally suggest, that keep alive hope of identifying both a distinct method and distinct results that are deserving of being called philosophical knowledge. Indeed, if the argument of this paper is cogent, such knowledge already exists and has since Aristotle; we just haven't been looking for it in the right ways or in the right places.

Specifically, these arguments conclude that certain philosophical theses are self-referentially inconsistent (incoherent), or in some similar way self-refuting. Versions of them have been directed against a surprisingly wide array of modern philosophical positions, including epistemological relativism,<sup>1</sup> determinism,<sup>2</sup> behaviorism,<sup>3</sup> representative realism,<sup>4</sup> evolutionary epistemology,<sup>5</sup> ontological relativity,<sup>6</sup> antifoundationalism,<sup>7</sup> skepticism,<sup>8</sup> deconstructionism and other "postmodern critiques of reason,"<sup>9</sup> Quine's thesis that "no statements are immune to revision,"<sup>10</sup> Whorf's thesis of linguistic relativity,<sup>11</sup> the "strong thesis" in sociology of knowledge,<sup>12</sup> versions of cognitivism holding that the world is a construction of the brain,<sup>13</sup> eliminative materialism,<sup>14</sup> and many others.

How much can such arguments accomplish? Do they succeed at decisively refuting their targets? Answers to these questions vary from a determined Yes to an equally determined No. The former come from philosophers holding that the positions at stake have genuine reflexive properties whose consequences must be taken seriously; by virtue of the kind of generality they take as their subject domain they have direct implications for themselves: for their own truth, knowability, assertability, or rational justifiability. Such philosophers are then in a position to uphold the self-referential argument as a distinctively philosophical strategy productive of results as definitive as those in mathematics and geometry. Those who deny the validity of self-referential arguments employ strategies ranging from a denial that reflexive properties really exist to the claim that reflexivity exists but for one reason or another doesn't provide the basis for refuting its targets. My aim in this paper will be to defend the first of these views from the criticisms employed by the latter. The conclusion is bound to be provocative and controversial; for I have come to believe that self-referential strategies, if carried out

properly, can be productive of genuine philosophical knowledge - knowledge, that is, which cannot be had in any of the special sciences *qua* sciences, and may, in fact, post limits on what the sciences may discover the world (or the objects in their special subject domains) to be like. Philosophy will be seen to have both a method and results of its own, a genuine place in our epistemic "division of labor" apart from the mere analysis of language and the results of the sciences or simply "keeping the conversation of the West going."

Let us consider some examples.

## 2.

*Example 1.* The best known self-referential argument is that which has commonly been directed against various forms of epistemological relativism. Harvey Siegel recently developed a version of this argument,<sup>15</sup> which in one form or another dates back at least as far as Socrates' effort to refute Protagoras in the *Theaetetus*. The version I will present here owes more to Siegel than anyone else; it goes as follows:

Let  $\rho$  be any declarative statement. In that case, epistemological relativism asserts that the truth or justifiability of  $\rho$  is relative to the central propositions of the conceptual framework (scheme, model, paradigm, etc.) in which  $\rho$  is most at home, especially those expressing this framework's standards for evaluating truth and justifiability; there is no framework-independent way of evaluating the truth or justifiability of  $\rho$ .

Now epistemological relativism certainly seems to be a thesis about *all* declarative statements. For a statement to be an exception to this general formulation of epistemological relativism would mean that its truth or justifiability is framework-independent; this would contradict epistemological relativism and confirm instead epistemological absolutism. So epistemological relativism's domain of reference must be all declarative statements. In that case, the position is clearly self-referring. For epistemological relativism is itself a logically ordered sequence of declarative statements (about the general nature of truth, knowledge, and justification). As the mathematical logician Frederick B. Fitch wrote, "If a theory [or statement] is included within its own subject matter, we say that it is a *self-referential* theory."<sup>16</sup> Epistemological relativism then applies to itself, and must have the properties it predicates of all declarative propositions or be self-referentially inconsistent; Fitch went on to observe that "If a self-referential theory T implies that T has [some] property P, and if T does not have the property P, then we will call T self-referentially inconsistent."<sup>17</sup>

To find out whether or not epistemological relativism is self-referentially inconsistent, let us formulate the immediate consequence of its self-reference: The truth or justifiability of epistemological relativism is relative to the conceptual framework (scheme, model, paradigm, etc.) in which it appears; there is no framework-independent way of evaluating epistemological relativism regarding its truth or justifiability. In other words, if we begin by assuming the truth of epistemological relativism, we reach the result that its truth is relative to the conceptual framework in which it appears (presumably a philosophical one). It will follow that since epistemological relativism's truth is itself relative, there is at least one conceptual framework in which epistemological relativism is false. For were it true for all frameworks it would be true absolutely and be in the embarrassing position of straightforwardly being its own counterexample; its self-refutation would be absolute. But to say that epistemological relativism is false in at *least one framework* is to say that in this framework epistemological absolutism is true. But if we take epistemological absolutism seriously, it becomes simply redundant to say that it is true in at least one framework. For absolutism's content does not relativize truth or justifiability to conceptual frameworks. So if epistemological absolutism is true in at least one framework, it is true *tout court*, and it will follow that epistemological relativism is not merely false in at least one framework but false *tout court* (frameworks where epistemological relativism was central will also be false *tout court*).

To sum up, if epistemological relativism is assumed to be true, then epistemological relativism is false. It cannot have the properties it predicates of all declarative propositions and hence of itself. Therefore epistemological relativism is self-referentially inconsistent, and therefore false. Those philosophers who reject it are right to do so.

*Example 2* W. V. Quine's celebrated claim that "No statement is immune to revision" was shown by Carl R. Kordig to have similar difficulties.<sup>18</sup> The context of Quine's statement is, of course, the following classic passage:

The totality of our so-called knowledge or beliefs . . . is a man-made fabric which impinges on experience only at the edges. . . . A conflict with experience at the periphery occasions readjustments in the interior of the field. . . . Reevaluation of some statements entails reevaluation of others, because of their logical interconnectedness -- the logical laws being in turn simply further statements of the system, certain further elements of the field. . . .

. . . Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system. Even a statement very close to the periphery can be held true in the face of recalcitrant experience by pleading hallucination or by amending certain statements of the kind called

logical laws. Conversely, by the same token, no statement is immune to revision.<sup>19</sup>

Let us label this final statement  $Q$  (for Quine). The appropriate question then is: Is  $Q$  immune to revision or isn't it? It seems clear that Quine intends  $Q$  to refer to the totality of statements; after all, he says so in the first sentence. Besides, without such universality there could easily be statements outside its domain which are immune to revision and would constitute counterexamples. And if  $Q$  refers to all statements, it includes itself in its domain of reference; otherwise it would again be its own counterexample. So it follows that  $Q$  is not immune to revision. For Quine, to revise  $Q$  would be to change its truth value. So it is possible that  $Q$  could be discovered to be false; we might find a counterexample if we looked hard enough.

The introduction of a modality here requires a different formulation than was used for epistemological relativism; instead of conceptual frameworks let us adopt the conventions of *possible world semantics*. In that case, to say that *possibly*  $Q$  is false is to say that there is at least one possible world where there are agents capable of formulating  $Q$  and in which  $Q$  is false. It is *necessary* that there be some such possible world, for otherwise  $Q$  would be true in all possible worlds (true necessarily, that is), and this again would make  $Q$  immune to revision: again, it would be its own counterexample.<sup>20</sup> Consequently we are forced to say that in at least one possible world,  $Q$  is false. In other words, in this world, at least one statement --  $\neg Q$  -- is immune to revision. But to say that  $\neg Q$  is immune to revision in at least one possible world is to say that in this world, the truth of  $\neg Q$  is necessary and not merely contingent. And to say that a statement is necessarily true is not to restrict its truth to a given possible world or set of possible worlds but rather to say that it is true in *all* possible worlds. So at this point, the reference to *at least one possible world* drops out as redundant, as did the reference to frameworks in the statement, epistemological absolutism is true in at least one framework, and for the same reason. So if  $Q$  is assumed to be true, then  $Q$  is false. We reach the result that Quine's "No statement is immune to revision" is self-referentially inconsistent, and hence necessarily false. Some statement is immune to revision, *tout court*. Aristotle's Principle of Contradiction has most frequently been offered as the prime candidate for such a statement.<sup>21</sup>

*Example 3.* Eliminative materialism (sometimes called the "disappearance theory of the mind") is the most recent and widely discussed theory which has occasionally been charged with self-referential inconsistency. Eliminative materialism consists of the following theses: (1) Our commonsense

conception of mental states including beliefs, knowledge, and other intentional states or propositional attitudes, is an empirical theoretical framework and not a set of givens; the friends of eliminative materialism call this framework *folk psychology*. (2) Folk psychology is a radically false framework, so false that a completed neuroscience of cognition can expect to eliminate the entities it postulates rather than reduce them to particular brain states or explicate them as functional states. In other words, according to eliminative materialism, beliefs, knowledge, and other intentional states or propositional attitudes are not givens but postulates of folk psychology, and therefore need play no more a role in a completed cognitive science than does, say, phlogiston play in modern chemistry or the impetus in modern dynamics. Since intentional states and propositional attitudes are part of the conceptual framework of folk psychology, if this framework goes, they go with it.

The argument for the self-referential inconsistency of eliminative materialism is somewhat more complicated than for the above two cases. With epistemological relativism and Quine's "No statement is immune to revision," self-referential inconsistency resulted from the positions' own internal logic; their inconsistency was *semantic*. The charge against eliminative materialism holds not that it is inconsistent in this way but rather inconsistent with principles which must be accepted as necessary conditions of rational discourse, conditions for the assertibility, meaningfulness, and rational justifiability of any theory whatsoever. In other words, the content of eliminative materialism conflicts with some of its own presuppositions; its alleged inconsistency is *pragmatic*. A different version of the argument is possible for every condition of discourse; to simply matters, I will focus on *rational justifiability* as a typical condition of discourse aimed at establishing declarative statements and *belief* as a typical propositional attitude. The argument, then, goes essentially as follows:

All scientific theories stand in need of rational justifiability, and this presupposes that they be the kind of things that can be rationally justified. Eliminative materialism, then, as a purported theory or research program for cognitive science, must be the sort of thing that can be rationally justified. But eliminative materialism can be rationally justified only if, at the very least, it can be made worthy of belief as the best theory available given the scientific evidence. A theory can be made worthy of belief only if there really are beliefs. So let us assume that eliminative materialism is true (i.e., that it depicts our cognitive life as it really is, as opposed to what folk psychologists say it is). If eliminative materialism is true, then there really are no beliefs, any more than there was a natural kind called phlogiston which is imparted to the air in every case of combustion or an impetus which pushes an object along in every case of uniform rectilinear motion. But in that case, given that there are no beliefs, it is actually

mistaken to hold that theories can be made worthy of belief. If no theory can be made worthy of belief, then no theory can really be rationally justified. Hence eliminative materialism cannot be rationally justified. We reach the result that if we assume eliminative materialism to be true, eliminative materialism cannot be rationally justified, even in principle. Result: if our theory permits the derivation of results that conflict with the possibility of ever rationally justifying any theory, then something is seriously wrong with the theory and it is appropriate to reject it as false (if not actually meaningless). As R. G. Swinburne put it in a review of the most elaborate defense of eliminative materialism, Paul Churchland's *Scientific Realism and the Plasticity of Mind*,<sup>22</sup>

If knowledge and justified belief are not to be had, Churchland does not have them and so his conclusions cannot be regarded as worthy of our belief. The general moral to be drawn from this is that the inanimate universe cannot be understood by someone who is no more than a very complicated part of it.<sup>23</sup>

Eliminative materialism, too, then, is self-referentially inconsistent in a pragmatic sense, and should rightly be rejected as describing a logically impossible state of affairs.

These three examples, then, direct self-referential arguments at theories in three basic subject domains: *epistemology*, the *philosophy of language*, and *cognitive science*. Epistemology, they assert, takes as its subject matter knowledge-claims but also consists of knowledge-claims, with the latter being a subset of the former. Likewise, claims about language are formulated -- how else? -- in language, and so must share any properties ascribed to all language. Regarding cognition, we must remember than any general theories about cognition are products of cognition; cognitive science is in this way self-referential. A completed cognitive science, then, cannot discover just anything about cognition. It could not discover, for example, that human beings are for whatever reason incapable of believing, discovering, or knowing factual truth. For this would constitute a belief, discovery, or factual truth, and the position would be defeated from within. So in this sense, philosophical argument limits cognitive science. To declare that cognition has no products (beliefs, factual knowledge) may be actually unintelligible; it would have the absurd consequence that cognitive science itself does not exist!

### 3.

Responses to such arguments fall into five isolable categories: (1) a chuckling dismissal at what is perceived to be substanceless dialectical

cleverness; (2) rejection on the grounds that the kind of reflexivity of form required by self-referential arguments does not really exist, and so such arguments cannot really get started; (3) rejection on the grounds that they are successful only against simplistic or artificially formalized versions of their targets: although such reflexivity might exist it is not to be found in any reasonable version of the position targeted; (4) rejection on the grounds that they beg the question against their targets, taking for granted theses which go hand in hand with a substantive theory the self-referential argument assumes and which is optional, not necessary; and (5) admission that self-referential arguments occasionally are successful, but fail to accomplish anything positive or useful except mere avoidance of contradiction.

Regarding (1), Henry W. Johnstone wrote:

To the chuckle we need not reply. It is the response of the unreflective man when confronted with any reflective analysis, and in fact represents his adjustment to an intellectual environment rather than a responsible argument.<sup>24</sup>

Accordingly we will say no more about it here. Strategy (2) is considerably more challenging. The most famous version of (2) began with Russell's realization that reflexivity of form lies at the heart of many logical and set-theoretical paradoxes. One of the most important strategies for getting rid of the paradoxes has therefore consisted of efforts to ban reflexivity of form from both philosophy and mathematics with a Theory of Types.<sup>25</sup> Russell saw that self-referential universal affirmative propositions about all propositions would include themselves in their own domain of reference, believed this to be the source of paradox. Therefore:

Whatever we suppose to be the totality of propositions, statements about this totality generate new propositions which, on pain of contradiction, must lie outside the totality. It is useless to enlarge the totality, for that equally enlarges the scope of statements about the totality. Hence there must be no totality of propositions, and "all propositions" must be a meaningless phrase.<sup>26</sup>

This identical situation applied to propositions about "all sets," "all relations," "all definitions," etc. He concluded with the following rule:

Whatever involves all of a collection must not be one of the collection; or, conversely. If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection had no total.<sup>27</sup>

Or, to put the matter more bluntly, self-referential propositions are

simply nonsensical. But then how are we to handle propositions such as, "All propositions about matters of fact are either true or false," which certainly seems to be (1) about all propositions and (2) true, not senseless.<sup>28</sup> Here the Ramified Theory of Types came to the rescue; propositions of this sort, in order to exclude themselves from their own domain of reference, would be said to be of a higher *type* than those included in their scope. Russell defined a type as "the range of significance of a propositional function,"<sup>29</sup> thus limiting its generality to a specific domain. Reflexivity of form could be avoided, then, with the "vicious-circle principle":

No totality can contain members defined in terms of itself. . . . Thus whatever contains an apparent variable must be of a different type from the possible value of that variable; we will say that it is of a higher type.<sup>30</sup>

According to the Ramified Theory of Types, then, we can distinguish a hierarchy of order among propositions and propositional functions. The lowest type consists of the totality of individual propositions, elementary propositions of the subject-predicate form containing no variables. A new totality can be formed by generalizing propositions about individuals, given that the class of individuals and the class of propositions are mutually exclusive. This yields the totality of *first order* propositions, the second type. Another way of saying this would be to say that first order propositions are universals about nonlinguistic entities, and since to be reflexive they would have to be about at least some linguistic entities they are trivially non-reflexive. Propositions of the form, "All propositions are x," refer to this totality, but are still nonreflexive since they are really truncated ways of saying, "All *first order* propositions are x." In that case, such propositions form a new totality of *second order* propositions, the third type which takes as its domain all first order propositions but none of any higher type. This process distinguishes successive types according to the general rule that no proposition or propositional function can contain a quantifier ranging over propositions or propositional functions of the same or of higher type than itself.<sup>31</sup> The result is that propositions such as "This sentence is false," or "All sentences are uncertain," or "This set is a member of itself," are not genuine propositions since they violate this rule. Hence according to the Ramified Theory of Types genuine reflexivity of form cannot exist; and self-referential argument, which depends on a special case of reflexivity of form, cannot get off the ground:

The  $n + 1$ th logical type will consist of propositions of order  $n$ , which will be such as contain propositions of order  $n - 1$ , but of no higher order, as apparent variables. The types so obtained are mutually exclusive, and thus no reflexive fallacies are possible so long as we remember that an apparent

variable must always be confined within some one type.<sup>32</sup>

In this case, the arguments for the self-referential inconsistency of epistemological relativism, "No statement is immune to revision," and eliminative materialism, will all be invalidated since the application of each of these positions to itself (or to conditions of its own rational justifiability) will be vitiated.

It is worth noting, first, that the Ramified Theory of Types does not succeed in exorcizing paradox since not all versions of paradox depend on self-reference. For example, the following well-known case depends on what might be called circular reference:

- (1) Sentence (2) is false.
- (2) Sentence (1) is true.

Second, it might be added that not all forms of self-reference generate paradox. Consider:

- (3) This sentence is in English.
- (4) All the sentences in this paper are carefully considered.

These are not paradox-generating and so are entirely innocuous. This suggests that banning self-reference to avoid paradox amounts to philosophical overkill. Far better to avoid paradox by eliminating paradoxical (because self-destroying) sentences piecemeal.

These observations, though, don't go to the heart of the matter. Paul Weiss, in an unjustly neglected paper, was the first to show in detail what happens as soon as we turn our attention to the machinery of the Ramified Theory of Types itself and pose the question of its place in the hierarchy it proposes. Weiss formulated the problem as a classic dilemma:

1. [The Ramified Theory of Types] is either about all propositions or it is not.

A. If it were about all propositions it would violate the Theory of Types and be meaningless and self-contradictory.

B. If it were not about all propositions, it would not be universally applicable. To state it, its limitations of application would have to be specified. One cannot say that there is a different theory of types for each order of the hierarchy, for the propositions about the hierarchies introduces the difficulty over again.<sup>33</sup>

In other words, the Ramified Theory of Types faced a self-referential dilemma before it could get off the ground. It alleged to describe logical features of the entire hierarchy of sensible propositions, but it is itself formulated in propositions (how else could it be formulated?). Hence it becomes not too difficult to show that the propositions comprising the

Ramified Theory of Types can have no type whatever, and so must be meaningless by one of the Theory's own criteria of meaningfulness (which was that every meaningful proposition can be assigned a type).

Assume that the propositions comprising the Ramified Theory of Types are of type  $n$  (where  $n$  can be any natural number greater than two). In this case, given the above rule, the Theory of Types cannot include in its domain any propositions of type  $n$ , or any of type  $n + 1$  or higher, since then it would be in violation of its own edicts. But in that case, it is at least possible that some propositions of this or of higher types are self-referential, and it will not have banished self-reference. To eliminate the possibility of self-reference from the totality of type  $n$  propositions, the Theory of Types will have to move up to type  $n + 1$ . But then the same difficulty arises for this order of proposition, and for the next, and so on; the result is a vicious infinite regress. The dilemma, then, is this: the Ramified Theory of Types cannot ban self-reference without violating its own principles and applying to the totality of propositions, and by applying to this totality it would apply to itself and hence be self-referential after all, in violation of its own edicts; conversely, a hypothetically successful Ramified Theory of Types (successful, that is, at banning self-reference) could have no type at all, and hence again be in violation of its own principles (that every meaningful proposition has a type). To say, as some have, that if one accepts the Theory of Types one does not allow criticisms of this sort to arise because one never refers to propositions in the required unrestricted sense,<sup>34</sup> is clearly to beg the question. Prior to the establishment of the Theory of Types there is no reason for the restriction on the generality of propositions. Or, as Fitch observed at the conclusion of a similar argument,

the ramified theory of types cannot assign a type to the meaning of the word "type," and yet it must do so if the theory applies to all meanings. In a similar way, no "order" can be assigned to a proposition which is about all propositions, hence no order can be assigned to the proposition which states the ramified theory of types.<sup>35</sup>

Consequently the Theory of Types cannot be considered successful in banning self-reference from philosophy; it is not a legitimate objection to self-referential argument, and we will have to be wary of type-based strategies which attempt to get around self-referential efforts. The contention that there are propositions, theories, etc., which are included in their own scope seems unavoidable, and Objection (2) above is answered.

## 4.

In other words, reflexivity of form seems vindicated and, indeed, ineliminable. Self-referential arguments are therefore at least possible. But it may yet be the case that they fail for other reasons. Objection (3) held that self-referential arguments are only effective against highly formalized and oversimplified versions of the positions at stake; so, even if successful, they are successful only against strawmen.

Paul Feyerabend, for example, claims to have articulated a version of Protagorean relativism which avoids self-refutation.<sup>36</sup> Feyerabend maintains that the argument against epistemological relativism went awry at the start by treating the position as a set of abstract propositions and arriving at a position easy to refute. His claim is that neither Protagoras nor any other serious relativist has ever had anything so precise in mind as that, the first philosopher to make this mistake being none other than Plato himself. In the *Theaetetus* Plato consistently has Socrates interpret Protagoras' relativistic remarks as abstract, well-formed propositions with definite logical consequences. According to Feyerabend, what we may call (for lack of a better term) *rhetorical* relativism "is not about concepts . . . but about human relations. It deals with problems that arise when different cultures, or individuals with different habits and tastes, collide."<sup>37</sup> Accordingly rhetorical relativism does not consist of abstract statements (abstract in the sense that they are meaningful apart from the particular context in which they are presented). Its statements:

are not 'universal truths'; they are statements which I, as one member of the tribe of Western intellectuals, present to the rest of the tribe (together with appropriate arguments) to make them doubt the objectivity and, in some forms, also the feasibility of the idea of objective truth.<sup>38</sup>

Relativists who try to utter 'universal truths' (e.g., anyone who would defend relativism as a thesis in epistemology) therefore misconstrue their own position and come up for typically Feyerabendian abuse:

Strangely enough there are relativists who . . . do not merely want to air their own opinions . . . they want to make general and -- god help us! -- 'objective' statements about the nature of knowledge and truth.

But if objectivism while perhaps acceptable as a particular point of view cannot claim objective superiority over other ideas, then the objective way of posing problems and presenting results is not the right way for the relativist to adopt. A relativist who deserves his name will then have to refrain from making assertions about the nature of reality, truth, and knowledge, and will have to keep to specifics instead.<sup>39</sup>

In other words, no relativist paying attention to what he is doing will allow his position to be formulated as we did epistemological relativism. His concerns are quite different. He wishes not to establish truths, either his own or anyone else's, but to undermine the claims others make to

truth, as a way of undercutting the intellectual authoritarianism which usually follows such claims. For all of Feyerabend's fun, games, and "dadaism," his work has a serious side; his broader aim is to help protect non-Western cultures and non-scientific traditions from being overwhelmed by what he regards as Western rationalistic and scientific imperialism.<sup>40</sup> Since the basis of these ideologies is one form or another of epistemological absolutism (what Feyerabend calls objectivism), if absolutism can be undercut the real targets fall with them. Rhetorical relativism, unlike the epistemological thesis, is not self-refuting since it has none of the definite logical consequences self-refutation requires; it presents no precise, general position for refutation because it questions both the possibility and desirability of precise, general positions.

But has Feyerabend really avoided self-refutation? If we pay close attention to what is going on, I think we will see that he has not. Self-refutation, as we already suggested, may result from circumstances other than pure, semantic self-referential inconsistency; it can arise on pragmatic grounds as well. Let us consider Feyerabend's own presentation of rhetorical relativism, including his declarations of his aims and of the restrictive nature of his propositions. There is good reason, I will argue, for suspecting that the very fact of this presentation forces him into a position at least as awkward as the epistemological thesis, the result being what we might call self-defeat. The passage begun above continues:

Debating with objectivists, [the relativist] may of course use objectivist methods and assumptions; however, his purpose will not be to establish universally acceptable truths . . . but to embarrass the opponent. He is simply trying to defeat the objectivist with his own weapons. Relativistic arguments are always *ad hominem*; their beauty lies in the fact that the homines addressed, being constrained by their code of intellectual honesty, must consider them and, if they are good (in their sense), accept them as objectively valid.<sup>41</sup>

So rhetorical relativism is addressed to those who accept absolutism, and is couched in terms which its adherents ought to understand and (if the rhetoric is successful) ought to find compelling. But if absolutists find grounds for not regarding rhetorical relativism as a serious or cogent thesis or its conclusions as true (as the *former* understand these terms), then the *ad hominem* backfires. Rhetorical relativism is left in the position of being, on its own terms, ignorable. Is Feyerabend in such a position? I believe he is. He has just told his readers openly that what matters for the success of his position is not the truth of its conclusions but the efficacy of its rhetoric. Since absolutists are interested in truth (again, as *they* understand the term), what more do they need?! Rhetorical relativism can achieve its aim only by offering absolutists a

compelling argument in absolutist terms; but absolutists will not be persuaded if there is *direct textual evidence* that they would be hoodwinked by taking the position seriously. The absolutist, contrary to Feyerabend, seems perfectly justified in treating rhetorical relativism, and perhaps all similar positions, as being clever but uninteresting wordplay; reasons for taking them seriously on their own terms just aren't there.<sup>42</sup> Should Feyerabend appeal to absolutists' "code of intellectual honesty," all they need do is retort that allowing themselves to be hoodwinked is not part of this code. So Feyerabend may avoid the semantic self-refutation of epistemological relativism, but his position is still pragmatically self-defeating: the full statement of the aims of the position undercuts whatever reason we may have for taking it seriously. Of course, Feyerabend could simply refrain from declaring such intentions. But then his position risks reverting back to old-fashioned epistemological relativism (or a position indistinguishable from it to his readers). It seems that in the case of rhetorical relativism, the position's self-referential properties have resulted not so much in falsehood as pointlessness. The absolute skeptic can utter the equivalent of "No one knows anything" and fall into self-refutation, or else clam up altogether; likewise, the advocate of rhetorical relativism, rather like an Erik Satie composition, is as ignorable as he is listenable. We can elect to go about our business as systematic philosophers as if he isn't even there.<sup>43</sup> This, I submit, takes care of Strategy (3).

## 5.

Objection (4) held that self-referential arguments, if not directed against strawmen, simply beg the question against their targets. Jack W. Meiland, for example, has argued that self-referential arguments against epistemological relativism beg the question against the relativist by assuming an absolutist conception of truth. Meiland argues that the self-refutation of relativism

is a myth which must be laid to rest. It *would* be inconsistent for the relativist to say both that all doctrines are relatively true and that relativism is not relatively true but instead is absolutely true. However, the careful relativist would not and need not say this. He would either say that all doctrines except relativism (and perhaps its competitors on the metalevel) are relatively true or false, or else he would say that his own doctrine of relativism is relatively true too. And saying that relativism is only relatively true does not produce inconsistency.<sup>44</sup>

It is clear that the first of the proposed strategies will not work; for it invokes an epistemic Theory of Types to make a distinction between "first order" doctrines whose subject domains are nonlinguistic and noncogni-

tive states of affairs and “second order” doctrines such as relativism and absolutism whose subject domains are first order doctrines and the conditions of their acceptability, or justifiability, with these two classes being mutually exclusive. And then, all we need ask is the position of Meiland’s meta-meta-level, and we have the same regress as we saw above.

Meiland no doubt realized this and opted for the second, which was to declare relativism true only relatively and try to cash out a notion of relative truth that itself can avoid inconsistency.<sup>45</sup> He did not, in my view, succeed, and for the reasons given above where we showed that the making the truth of relativism relative to a given conceptual framework results in its being compatible with relativism’s falsehood in some other framework (in fact, requires its falsehood in at least one framework); here, Meiland might argue, is where the absolutist conception of truth enters the picture. Can we do without it? Meiland’s strategy was to cash out “*p* is true in *W*” as “*p* is true-in-*W*,” where *W* is some conceptual framework.

[The hyphens] are extremely important. For they show that the relativist is not talking about truth but instead about truth-for-*W*. Thus, one can no more reasonably ask what ‘true’ means in the expression true-for-*W* than one can ask what ‘cat’ means in the word ‘cattle.’ ‘True-for-*W* denotes a special three-term relation which does not include the two-term relation of absolute truth as a distinct part.<sup>46</sup>

This, as it turns out, will not work either. Meiland believes he has isolated a three-term relation which will express a coherent relativist notion of truth. Presumably, then, the three items being related are statements, conceptual frameworks (*W*’s), and the actual world. But as Siegel wondered,

What . . . is the status of the world on the three-term conception? Is it clearly distinguishable from the other two relata? Unfortunately, the answer is no. On the relativist conception, the world is not distinguishable . . . What are related by the alleged three-term relation are statements and the world-relative-to-*W*. . . . On the relativist conception, the world cannot be conceived as independent of *W*; if it is so conceived, the relativist conception collapses into an absolutist one, for it is granted that there is a way the world is, independent of statements and of *W*’s. This is precisely what the relativist must deny, however.<sup>47</sup>

So on Meiland’s conception, the actual world simply drops out. It can never be known or talked about; what can be known or talked about is the world (or some part of it) as conceived by the community which believes *W*, thus blurring the distinction between the world and *W*.<sup>48</sup> So the formulation of relative truth as a three-way relation contains the seeds of its own destruction no less than did epistemological relativism, in our original formulation. If we can talk about the world as one thing and

conceptual frameworks as another, then why not just talk about the world (or some part of it) and treat frameworks as, perhaps, psychological or sociological entities with no necessary epistemic significance? Meiland, it seems, has no other option than to drop his third relation to the world and speak of truth as framework-relative, period. But in this case epistemological relativism collapses into full-blown conceptual idealism. I conclude that truth-for-W is as logically impossible as epistemological relativism itself, and hence hardly fitting as a means by which to rescue the position from self-refutation. The only clear formulation of true-in-W might read something like *believed to be true by those who believe W*. But this latter notion is trivial; it amounts to the commonplace observation that different peoples have different beliefs, or that different communities of scientists have promoted different and conflicting research programs at different times. Relative truth, to be at all credible, must be articulated in such a way that it does not collapse into absolute truth; otherwise the notion is as self-contradictory as epistemological relativism. Meiland failed to avoid this basic dilemma.

Friends of eliminative materialism have retorted that the argument from self-referential inconsistency is question-begging. Here the response looks to be, at first glance, considerably more formidable since eliminative materialism is a more complex position and has been defended with a great deal of skill. Andrew D. Cling recently systematized the self-referential line of argument sketched above as follows:

1. Eliminative materialism (EM) can be articulated and defended.
2. EM can be articulated and defended only if it can be justified, only if it is the sort of thing which can be justified.
3. EM can be justified only if it can be made worthy of belief.
4. A theory can be made worthy of belief only if there are beliefs.
5. EM is true (assumption for *reductio*).
6. There are no beliefs. (from 5)
7. No theory can be made worth of belief. (from 4, 6)
8. EM cannot be justified. (from 3, 7)
9. EM cannot be articulated and defended (from 2, 8)
10. If EM is true, then EM both can and cannot be articulated and defended. (from 5-9, 1)
11. EM is not true. (from 10)<sup>49</sup>

Cling calls this the argument from justificatory presuppositions.<sup>50</sup> He smokes out (4) as the argument's most controversial link. According to Cling,

(4) is a strange sort of statement. (4) claims that the possession by a theory

of a certain normative property, belief worthiness, in some way requires the truth of certain descriptive psychological statements to the effect that beliefs exist. . . . What [this] says is that a theory's possession of the capacity to be justified depends upon the existence of states individuated within the descriptive confines of what may turn out to be an idiosyncratic account of human cognition and behavior.<sup>51</sup>

In other words, the self-referential criticism begs the question by presupposing the reality of beliefs as a condition for justifying a theory; it presupposes a central tenet of folk psychology, the theory at issue.

However, as Cling also notes (and Churchland before him), for eliminative materialism to eliminate beliefs and other propositional attitudes it is imperative that it provide an alternative account of justification which makes no use of such entities but does all the work beliefs do. Here is where the trouble starts; for the question invariably arises, not, Does the criticism of eliminative materialism presuppose the reality of beliefs? but, Must it *necessarily* presuppose the reality of beliefs? Cling argues that the position can get by without beliefs.

Why can't we say simply that a theory is worthy of belief only if it is more likely than not to be true in light of the evidence? Here there is no explicit reference to beliefs at all. On this way of looking at things, talk of the belief-worthiness of a theory does not commit us straightforwardly to any particular way of describing beings who theorize. Which account we do adopt is left up to such things as predictive and explanatory power.<sup>52</sup>

This, though, is puzzling. It suggests we are to make a hard and fast distinction between *belief* and *belief-worthiness* in such a way that the first is a folk psychological concept and the second at home in eliminative materialism? In this view, what adjudicates theories is whether they are "more likely than not to be true in light of the evidence." But it is human beings, "beings who theorize," who decide this; theories do not adjudicate themselves, after all. This suggests it will be impossible to separate human beings, their decisions, and whatever forms the basis of these decisions, from the adjudication process. The friend of eliminative materialism might assert, dogmatically, that eliminative materialism is not a matter of belief at all but of scientific truth. But to my knowledge no one has taken this route, nor would they; even for Churchland, eliminative materialism's most formidable defender, eliminative materialism is the just the most reasonable research program available for cognitive science, not something he or anyone else can claim to have shown to be true. But this is just a roundabout way of saying that eliminative materialism, taken at face value, is a candidate for our allegiance. It is, in other words, a body of *belief*, a candidate for *belief-worthiness*. To say that we have cashed out *belief-worthiness* in a manner making no reference to *belief* hence obscures instead of clarifies.

At this point, the friend of eliminative materialism might employ a different strategy. He might argue that what is being eliminated is not *all* forms with propositional content, just distinctively *mental* ones. He might then be in a position to say, not that he *believes* eliminative materialism to be worthy of pursuit but that he *believes\** eliminative materialism to be worthy of pursuit, where *believes\** functions as a placeholder for something to be articulated more fully within a more developed neuroscience. That the friend of eliminative materialism might find this a credible strategy is indicated by Peter Smith's remarks to the effect that the friend of eliminative materialism

does not believe his thesis; for by his light there are no beliefs. [And] this only leads to paradox when taken together with the claim that he is asserting his thesis -- and our materialist rejects this too. His position is rather that he is asserting\* a proposition which he believes\*. Thus our materialist can consistently describe his situation, echoing from outside the framework of folk psychology the insider's description of what is going on, while continuing to insist that beliefs\* are no more to be identified with beliefs than states of demoniacal possession\* (i.e., what are in fact hallucinatory psychoses) really are states of possession.<sup>53</sup>

In this case, what we need is an account of the ways in which beliefs\* differ from beliefs in addition to the trivial one that the former is a "nonmental" neuroscientific concept and the latter a "mental" folk psychological one. Now it would be unfair to place too high a burden on eliminative materialism at this point; for *belief\** cannot very well be given a detailed explication in the absense of facts about the brain as yet undiscovered and within a theoretical framework as yet undeveloped. But I suspect that the relation holding between members of the pair *beliefs/beliefs\** will be different from that between the pairs *demon possession/psychotic states* and *instances of phlogiston being imparted to air/instances of oxygen being taken up from air*, etc. For as Cling notes, eliminative materialism does not propose to eliminate all propositional content as modern psychiatry eliminated all demons and chemistry eliminated all chemical principles.

Eliminative materialism does not entail the claim that there are no states with propositional content, it only entails the thesis that there are no mental states with propositional content.<sup>54</sup>

But this only serves to increase our puzzlement. Unfortunately there is not sufficient space here to explore the issue of just what is eliminated when an out-of-date theory is replaced by a successor.<sup>55</sup> But we can make some admittedly cursory remarks, as a prelude to a more detailed investigation. During the time of the chemical revolution of the late eighteenth century, at issue (for Lavoisier, anyway) was the adequacy of a certain theory of combustion which postulated a specific natural kind,

phlogiston, as the key to explaining every instance of combustion (as well as other natural phenomena such as the common properties possessed by all metals). The chemical revolution eliminated phlogiston as a referring term; it certainly did not eliminate the observable phenomenon, combustion. Likewise, as psychiatric science advanced it eliminated demons (or demonic possession) as a referring term; it did not eliminate the states which demons (or demonic possession) had been invoked to explain.<sup>56</sup>

I submit that the situation with beliefs is very similar. That we have a mental life in some sense of this expression is no less observable than that there is combustion: all who are parties to this debate can observe their own, pre-analytic mental lives for themselves by direct introspection. How we explain or offer a scientific account of that mental life is a different matter (one where introspection may not be of much help, any more than direct observation gives us the microstructural properties of physical and chemical processes such as combustion). If all eliminative materialism purports to do is eliminate the view that beliefs and other mental phenomena consist of mysterious, nonphysical entities inside our heads (perhaps made of some kind of Cartesian “mental stuff” or perhaps just not capable of a physicalist account) then its success would hardly be a new or groundbreaking achievement for few philosophers and practically no cognitive scientists believe we have a “mental life” in this sense. But if we can elaborate a more up-to-date theory of what beliefs are; if, say, we propose that they are manifestations of complex neurophysiological (i.e., essentially material) processes capable of storing information in a referential manner,<sup>57</sup> then we have a notion that does the work of our traditional concept of belief but without Cartesian or some other dualistic ontology. But it is clear that we have not eliminated beliefs, only the outdated ontology; for were we to eliminate beliefs *per se*, we would have a notion incapable of doing the above work.<sup>58</sup> In short, the friends of eliminative materialism have conflated two separate things, our everyday *experience* of belief, and dualistic or neodualistic ways of *understanding* this phenomenon. The latter we can part company with and avoid pragmatic self-referential inconsistency; not so with the former.

This kind of argument, I will submit, should also enable us to grapple with one of the Churchlands’ primary efforts to defuse self-referential criticisms of eliminative materialism. Churchland, in defending his position from the charge of self-referential inconsistency, drew the following analogy between the self-referential argument against eliminative materialism and that which a hypothetical philosopher might have made against *vitalism* a century ago (he actually credits his wife and colleague Patricia Smith Churchland for having originated the analogy):

The anti-vitalist says that there is no such thing as vital spirit. But this claim is self-refuting. The speaker can expect to be taken seriously only if his claim cannot. For if the claim is true, then the speaker does not have vital spirit and must be dead. But if he is dead, then his statement is a meaningless string of noises, devoid of reason and truth.<sup>59</sup>

This argument, meant to be taken as obviously invalid, would eliminate the self-referential argument against eliminative materialism by logical refutation, and at first glance, quite powerfully. For clearly no one today asserts that possessing a vital spirit is a condition of (or explanation of) being alive. And it is this analogy, between having a vital spirit as a condition of being alive and having beliefs as a condition of being able to justify or meaningfully assert one's theories, that eliminative materialism's defenders want to press. Will the analogy work? The vitalist would have maintained that the sentence, "I am alive although there is no vital spirit," is self-contradictory. In other words, *being alive* and *having vital spirit*, were, according to the vitalist, synonymous and coreferential. Interpreting the terms this way would make the antivitalist's argument as sound as the ones against epistemological relativism and "No statement is immune to revision." But this interpretation would be odd, given that the former refers and the latter does not. Let us ask, though, what task was the concept *vital spirit* intended to perform? This seems clear: to explain the observed phenomenon of life prior to the arrival of concepts revealing life's chemical and biological conditions. So we cannot conflate the observed phenomenon *life* with concepts invoked to explain it. If we do so, we can substitute into the above statement and end up with the equivalent of, "I am alive, but I am not alive," which is obviously self-contradictory. So when we dropped *vital spirit* from our vocabulary we did not at the same time eliminate the concept *life*; to eliminate the former was not to eliminate the latter. Thus, "I am alive although there is no vital spirit," is consistent. Is this the case with *belief*? The critic of the eliminative materialist's equivalent sentence is (put in the new vocabulary with its placeholder), "I believe\*  $p$  although there are no beliefs." Is this statement self-contradictory? To find out, we must pinpoint which of the two senses of *belief* are meant. Given the Cartesian (or neo-Cartesian) usage, there will be no inconsistency; this will not be the case for the pretheoretical usage. Or to put the matter another way, *belief\** does not eliminate pretheoretical belief but only the Cartesian (or neoCartesian) theory of belief (of our mental life generally); it incorporates and explains pretheoretical belief in the way earlier, less radical forms of materialism purported to do. In this case, the statement "I believe\*  $p$  although there are no beliefs," will indeed be self-contradictory if the pretheoretical sense of belief is meant; the theoretical *belief\** must, of necessity, contain and explain the phenomenon of *belief*, not eliminate it. Thus the *reductio* of the self-referential criticism proposed by the Churchlands rests on a confusion of theoretical and nontheoretical notions, and so does not

succeed. The issue, in their terms, is not whether the friend of folk psychology begs the question against eliminative materialism but rather whether he is *forced* by the internal logic of the debate to "beg the question." For one of the implications of this result is that certain concepts eliminative materialists locate in what they call the theoretical framework of folk psychology (e.g., knowledge, belief) may well turn out to be pragmatic necessities. There may be no other intelligible way of describing our cognitive lives as "beings who theorize;" in this case, cognitive science will be faced with the choice of accomodating this by virtue of its status as a product of human cognition or fall into pragmatic inconsistency. There are indeed descriptions which are barred to rational forms of cognitive science, one of which is that there isn't really any such thing as belief (in its pretheoretical sense, without the asterisk). This will mean that Objection (4) is answered.

## 6.

At this point it might appear that the critics of self-referential argument in philosophy are, if not in full retreat, at least on the defensive. But they have one last gambit to play. Objection (5) did not deny that self-referential arguments occasionally score direct hits; it suggested that nothing useful or positive is accomplished by their doing so: if we arrive at the view that some version of epistemological absolutism must be true we have not added anything scientifically concrete to our knowledge of the world; by concluding that "some statement is immune to revision" we have not identified *what* statement is immune to revision (other than this statement itself); the claim that beliefs are necessary does not give us an adequate account of what beliefs really are, neurologically speaking, nor does it tell us how they ought to be fitted into an adequate science of cognition; indeed, one of the genuine merits of eliminative materialism has been to show us that we still lack such an account. In summary, self-referential arguments accomplish nothing more than avoiding contradiction. As J. L. Mackie argues at the conclusion of his formal analysis of self-refutations,

We might be tempted to believe that there is a special form of philosophical argument which enables us to establish positive conclusions by showing that certain contrary statements would be self-refuting. This would go against empiricism, for if any view would literally refute itself, its denial would be a necessary truth. However, our analysis shows that this challenge to empiricism evaporates on closer inspection.<sup>60</sup>

Mackie's statement is extremely valuable for its identification of what is really at stake here: *empiricism*. If we consider the structure of each of

the positions above, their shared commitment to empiricism as a theory of the origins of knowledge should be evident. The defender of epistemological relativism (or similar positions) frequently relies, for example, on the empirical observation that different peoples and different scientific communities have used different methods and standards and sometimes described their observations in quite different ways from those of our own communities, noting that although this by itself is not a refutation of absolutism it at least makes sense of relativism. Feyerabend, for example, draws liberally on episodes from the history of science; he also makes use of the findings of anthropological linguists such as B. L. Whorf, and his arguments occasionally even include forays into art history.<sup>61</sup> He has, moreover, explicitly labelled his views as empiricist.<sup>62</sup> Quine's position, while different from that of the logical empiricism he criticized and rejected, is still empiricist in the broader sense that it relies on such entities as "sensory stimuli" and "surface irritations" as the means by which the truth-values of those propositions describing phenomena at the periphery of our "web of belief" are revised.<sup>63</sup> And the friend of eliminative materialism is clear about his commitment to empiricism as a component of his confidence that neuroscience will eventually do away with such philosophical disciplines as epistemology. Eliminative materialism is, in fact, just the latest in a long line of philosophical theses resulting from the assumption that empirical science has the final word in matters cognitive (a thesis sometimes called *scientism*). I will submit that commitment to increasingly radical forms of empiricism by modern philosophers beginning with Hume and extending to Feyerabend, Quine, and eliminative materialists (and, often, by scientists as well<sup>64</sup>), is the main reason why self-referential arguments are generally regarded as wrongheaded. For self-referential arguments are not empirical; they are (contrary to Cling) a species of *a priori* argument. While not taking issue with specific, concrete scientific findings (which, as everybody knows, always underdetermine theory), they reach the result that there are certain empirical states of affairs which science could not, even in principle, discover to hold, because the propositions describing them are necessarily false -- either false in all possible worlds or false in all those worlds where there exist beings capable of formulating and rationally defending them. In short, self-referential arguments rest on an apriorist epistemology and philosophy of logic; this puts them quite at odds with the most influential doctrines of the twentieth century.

Be that as it may, it does not answer Mackie's central challenge, which is to produce some positive results of self-referential arguments in philosophy. Here we must be careful. We must realize that although the self-referential argument places logical/conceptual limits on what science can discover, in no other sense does it compete with science.<sup>65</sup> If anything, a more detailed study of self-referential relations than can be attempted

in a paper of this length should be able to clarify the differences which emerge between philosophy and the sciences given apriorism. For while the sciences are domain-specific and their results discovered empirically, the results of self-referential argument in philosophy are highly general and discovered *a priori*; they do not yield concrete scientific results but rather help delineate the forms to which scientific results (and, indeed, all other cognitive enterprises) must conform. Can we isolate such accomplishments at high levels of generality? I believe we already have, and that the results should shed light on the dispute between foundationalists and antifoundationalists.

In Section 2, we reached the result that "Some statement is immune to revision" is necessarily true (true in all possible worlds); at the end of that section we proposed Aristotle's Principle of Contradiction, "that the same attribute cannot at the same time belong and not belong to the same subject and in the same respect . . ."<sup>66</sup> as the most likely first candidate for a statement immune to revision. In twentieth century philosophy, particularly among logical positivists, it has been standard to treat the Principle of Contradiction as having no empirical consequences but rather as being a tautologous combination of signs -- a formal or linguistic convention rather than a metaphysical law of reality. A. J. Ayer wrote that "the principles of logic . . . are universally true because we never allow them to be anything else."<sup>67</sup> But to call a definition, theory, or logical principle a convention -- to say of it that "we allow it" to be such and such or "do not allow it" to do such and such - is to imply that "we" could have stipulated otherwise, i.e., that we could have devised a logical system with no Principle of Contradiction, and in which there are results that ignore the Principle of Contradiction. (Some might even say that Hegel's system does just that.)

This, however, has bizarre consequences. If we assume that the Principle of Contradiction applies only to certain combinations of signs formulated and used by human beings, then does it not follow that genuinely contradictory but no less *real* states of affairs are possible? Consider a proposition such as "It is the case both that there are houses on Elm Street and that there are no houses on Elm Street." Conventionalism in logic (and Quinean universal revisionism) would permit it to be true in at least one possible world that there both are and are no objects of a particular kind in a specific place.<sup>68</sup> But this is clearly absurd! Were someone to claim that he had observed or even conceived of some such state of affairs, he would be considered joking or insane (most likely the former, since not even the clinically insane hallucinate contradictory states of affairs). So whatever else we might say, the Principle of Contradiction seems not to be a convention we could revise on the basis of recalcitrant experience.

Aristotle himself gave what at least one commentator has concluded is

the best argument ever devised both for why the Principle of Contradiction could not be otherwise and for why we are justified in taking it as a law of reality.<sup>69</sup> Aristotle's argument consisted of demonstrating the unintelligibility of any denial of the principle of contradiction (and, hence, of any logical system which claims to dispense with it). Aristotle pointed out that any significant particular utterance, e.g., "All humans are rational animals," presupposes that one definite kind of thing is meant by the word *humans* and another definite thing is meant by the categories *rational* and *animals*. To presuppose such is to acknowledge the Principle of Contradiction; not to presuppose it would imply that these words could have arbitrarily different meanings on one and the same occasion, the result being a breakdown of intelligible discourse.<sup>70</sup> As Aristotle himself said,

If . . . one were to say that the word has an infinite number of meanings, obviously speech would be impossible; for not to have one meaning is to have no meaning, and if words have no meaning our reasoning with one another, and indeed ourselves, has been annihilated, for it is impossible to think of anything if we do not think of one thing.<sup>71</sup>

So the argument boils down to the following: in order for language to be meaningful or communicative at all, it must have definite content; and in order for it to have definite content it must have noncontradictory content; hence any meaningful and communicative use of language presupposes the Principle of Contradiction.

This, of course, is not a "proof" that the Principle of Contradiction is a law of reality; it is what Aristotle called a negative demonstration and hence is dialectic. As a 'first principle,' the Principle of Contradiction is presupposed in the very concept of a proof; were it subject to proof, it would not be a 'first principle.' So it might seem, again, that the Aristotelian argument just begs the question. However, we have no alternative except to use the Principle of Contradiction in its own defense. So in a sense, any defense of the Principle of Contradiction is indeed circular. But if any intelligible use of language presupposes the Principle of Contradiction, then clearly any attempt to philosophize in its defense will necessarily presuppose it; circularity is unavoidable. It is, however, not fallacious, since it is not part of an attempt to prove the Principle of Contradiction true. The dialectic shows, if any thing, that we cannot imagine what things would be like if it were false! We can, of course, utter sentences like, "It is the case both that there are houses on Elm Street and that there are no houses on Elm Street." But we cannot conceive of a factual situation they would describe. This seems to establish it as immune to revision, and put us on the road to answering Mackie's challenge. We do not merely avoid contradiction but can state affirmatively that no possible worlds contain contradictions.

## 7.

Since this may still not seem like much of an achievement, it is worth concluding by going back and reiterating the rest of our results in the context of their implications for philosophy as a genuine cognitive enterprise. If the various lines of argument throughout this paper are cogent, we have demonstrated that some statements and theories are semantically self-referential (contain themselves in their domain of reference); certain others are pragmatically self-referential (contain in their domain of reference the necessary conditions of their own meaningfulness, assertibility, rational justifiability). Both must not yield consequences which conflict with the assumption of their truth; if they do, they must be rejected as self-referentially inconsistent. Type-based strategies designed to avoid self-reference quickly get entangled in the very difficulties they are designed to avoid, as no "type" can be assigned to the propositions in which these strategies themselves are formulated. That the Theory of Types is false is, therefore, immune to revision. If self-reference is combined with the above rejection of conventionalism about the Principle of Contradiction, we reach the more specific result that self-referentially inconsistent statements and theories actually purport to describe states of affairs which are necessarily false, cannot hold in any possible world (or, in some cases, cannot hold in any possible world which also contains agents capable of formulating and rationally defending them). Their denials describe states of affairs which, conversely, must hold necessarily. Steven J. Bartlett recently stated that "A postulate is self-validating if its denial will result in self-referential inconsistency."<sup>72</sup> The denial of epistemological absolutism is epistemological relativism; since the latter was found to be self-referentially inconsistent, the former is validated. The denial of the claim that no statement is immune to revision was likewise found to be self-referentially inconsistent; so it must be the case that some statements are immune to revision, and that this statement itself is immune to revision. The denial of the contention that there really are beliefs, however we explicate them, resulted in pragmatic inconsistency. So it must be the case that beliefs are ineliminable, and that -- however we come to understand them scientifically -- "There are beliefs" is also immune to revision.

Actually, if the strands of argument comprising this paper are sound, they suggest new and potentially quite fruitful directions for philosophical research on the part of philosophers dissatisfied with the state of affairs sketched at the outset of this paper, with "continuing the conversation of the West" (as Richard Rorty puts it)<sup>73</sup> -- or just with their standard status as linguistic/conceptual underlaborers. These results suggest the possibility that foundationalism, despite having taken some hard knocks over the past few years, is still very much alive and kicking. For what is validated,

for example, in the validation of epistemological absolutism is the view, quite startling in an age of historicism, relativism, and "playful nihilism," that genuine knowledge and justification do not depend on one's culture, conceptual framework, model, theory, paradigm, or some other contingent factor. It does not follow from this, of course, that cultural differences, conceptual frameworks, models, paradigms, etc., do not exist or do not influence the beliefs, actions and practices of scientists and others, for quite obviously they do and are often confused with the actual world. However, my contention will be that these phenomena are best understood psychologically and sociologically, not logically or epistemologically.<sup>74</sup> They may influence one's thought and, up to a point, one's perceptions - but as a matter of logic, they can be transcended (a fact without which intellectual change and progress of whatever sort would obviously not be possible). Cognitive/epistemic determinism is, in short, false, and necessarily so; that we may, in principle, transcend whatever framework in which we find ourselves working is another statement immune to revision. In an age where social theorists are so quick to assume quite dogmatically that one's race, gender, class, up bringing, etc., all function in some combination as determinants on the thinking of the individual,<sup>75</sup> this seems to me a discovery of the first importance.

We stated that "some statements are immune to revision" is necessary because its denial is self-referentially inconsistent. This, I submit, suggests an important aim for philosophy: *the attempt to identify and improve our understanding of logical necessities holding in language, thought, and reality*. The domain of philosophy differs from the domains of the sciences in that the sciences are domain-specific, whereas philosophy seeks laws and concepts which apply across the board to all domains. It cannot of course prove to the satisfaction of all skeptics that such laws and concepts exist to be discovered, for, again, the concept of a proof in whatever sense we choose requires them. But I am assuming (and perhaps liberalizing my basic Aristotelianism with a pinch of Peirce<sup>76</sup>) that the mere possibility of doubt is not a positive reason for doubt, and so argue that philosophers can work under the reasonable belief that such laws and concepts exist to be discovered. These laws will be expressed as statements which are irreversible in the sense that their denials will sometimes result in self-referential inconsistency and sometimes simply in nonsense. In this case, there is a sense in which philosophy "stands above" or outside of science in just the way denied by Quine, and can be made foundational in just the sense denied by Rorty. Beyond this, of course, philosophy does not legislate specific methods and content to the sciences; it is up to scientists to discover and apply the methods most suitable to their particular domains. As for content, it will be true (and immune to revision) that a scientific discipline cannot discover just anything about its subject domain; for philosophy sets the logical-

conceptual boundaries of the world science can discover. There will be a clear-cut division of labor between the two, with plenty of work for everybody.<sup>77</sup>

Two concluding remarks are in order. (1) I do not claim to have done more than scratch the surface here. At the outset I mentioned, but due to space limits could not explore, self-referential arguments against a variety of positions in addition to those considered here. Ultimately a comprehensive account of the different forms of self-reference and their consequences for the various branches of philosophy and those areas of science directly connected to human beings (cognitive science and the so-called social sciences) will be needed, as well as those cognitive conditions which make reference of any sort possible.<sup>78</sup>

(2) These ideas, as I also noted at the outset, are admittedly not new; in fact, they go back to Plato and have been preserved or developed in one form or another by many twentieth century philosophers of a variety of persuasions: Weiss, Fitch, Kordig, Bartlett, Siegel, and many others. But these voices have been all but drowned out by the postmodern chorus of historicism, relativism, and antifoundationalism. One of the motive forces of this investigation has been this writer's growing concern that these paths can lead nowhere except to the further weakening of philosophy as a discipline: increasing its level of overspecialization, vulnerability to irrationalist ideologies and special interest groups (militant feminists come to mind), and the ultimate irrelevance for which academic philosophy is sometimes justifiably criticized. Philosophy, many have argued plausibly, should have as one of its larger aims the critical evaluation of culturally significant worldviews with the ultimate aim of achieving personal and social *wisdom*.<sup>79</sup> but in our century it has failed in this mission. Philosophy as a discipline has in recent years suffered a loss of nerve. Professional philosophers do not like to use such an expression, but many would not deny its aptness. I find it interesting and significant that this problem began around the time positivism, empiricism and scientism became the dominant views in epistemology (while emotivism and other forms of noncognitivism became the corresponding dominant modes of thought in meta-ethics). Hence it concerns me little that others have walked the conceptual paths I am walking now if these paths have the potential to lead our discipline out of crisis and offer it a new identity.<sup>80</sup>

1. Examples of self-referential arguments see Donald Davidson, "On the Very Idea of a Conceptual Scheme", recent examples see Donald Davidson, "Determinism's Dilemma", "Review of Metaphysics" 23 (1986): 225-59.
2. See e.g., James N. Jordan, "Determinism's Dilemma", "Review of Metaphysics" 23 (1969): 48-66; Joseph M. Boyle, Jr., German Giese and Olaf Tollefse, "Determinism, Freeedom, and Self-Referential Arguments", "Review of Metaphysics" 26 (1972), pp. 3-37; Peter Psycological Laboratory?", "Philosophy of the Social Sciences" 4 (1984): 149-57.
3. Formulated almost immediately upon the appearance of J. B. Watson's views by Arthur O. Lovejoy in "The Paradox of the Thinking Behaviorist", "Philosophical Review" 31 (1922): 135-47.
4. Stephen Willcox and Stuart Katz, "Can Indirect Realism Be Demonstrated in the Psycological Laboratory?", "Philosophy and Phenomenological Research" 20 (1983), pp. 207-16.
5. Carl R. Kordig, "Self-Reference and Philosophy", "American Philosophical Quarterly" 20 (1983), pp. 209-11.
7. See my, "Rorty's Foundationalism", "Reason Papers" 14 (1989): 130-36, and the literature cited there.
8. See my, "The Skeptic's Dilemma", "Reason Papers" 14 (1989): 143-46.
9. See James L. Marsh, "Strategies of Evasion: The Paradox of Self-Referentiality and the Post-Modern Crisis of Rationality", "International Philosophy Quarterly" 29 (1989): 339-49.
10. Carl R. Kordig, "Some Statements Are Immune to Revision", "The New Scholasticism" 55 (1981): 69-76.
11. See Lewis S. Feuer, "Sociological Aspects of the Relation Between Language and Philosophy," "Philosophy of Science" 20 (1953), esp. pp. 95-96; Franklin Henton, "An Examination of the Conceptions of Benjamin Whorf in the Light of Theories of Perception and Cognition and Cognition in Culture", ed. Harry Holger (Chicago: University of Chicago Press, 1954), esp. pp. 53-55; cf. also the remarks in Davidson, op. cit.
12. See e.g., Barry Gureneberg, "The Problem of Reflexivity in the Sociology of Science," "Philosophy of the Social Sciences" 8 (1978): 321-343.
13. Stuart Katz and Gordon Foss, "The Origins of Knowledge in Two Theories of Brain: The Cognitive Paradox Revealed", "Behaviorism" 7 (1979): 35-44.
14. Cf. Nicholas Everitt, "A Problem for the Eliminative Materialist", "Mind" 90 (1981), pp. 428-34; R. G. Swinburne, review of Paul Churchland's *Solventive Realism* and the *Plasticity of Mind*, "Philosophy" 55 (1980): 273-75; Lynne Rudder Baker, "Cognitive Solidide," in *Contents of Thought*, eds. R. H. Giurm and D. Mennell (Tucson: University of Arizona Press, 1987), pp. 1-30; cf. also her *Sharing Belief: A Critique of Physicalism* (Princeton: Princeton University Press, 1987), esp. ch. 7.
15. Cf. Kordig, "Some Statements Are Immune to Revision", esp. pp. 75-76, op. cit., n. 1.
16. Fredrick B. Fitch, "Self Reference in Philosophy", "Mind" 55 (1946): 65; reprinted in his *Symbolic Logic: An Introduction* (New York: The Ronald Press, 1952), Appendix C, 16. The modalities introduced below are mine, not Kordig's.
17. Ibid., p. 66.
18. Cf. Kordig, "Some Statements Are Immune to Revision", esp. pp. 75-76, op. cit., n. 16. Willard Van Orman Quine, "Two Dogmas of Empiricism", in *From a Logical Point of View* (Cambridge: Harvard University Press, 1953, 1961), pp. 42-43.

20. We might add, paradoxically, that it is also necessary, in our terms, that there be some such world in order to conform to Quine's rejection of the very idea of necessary truth - suggesting that one cannot consistently reject necessary truth.
21. See Section 6 for details.
22. Paul M. Churchland, *Scientific Realism and the Plasticity of Mind* (Oxford: Cambridge University Press, 1979).
23. Swinburne, op. cit., n. 14, p. 274.
24. Henry W. Johnstone, "Philosophy and *Argumentum ad Hominem*," reprinted in his *Validity and Rhetoric in Philosophical Argument* (University Park, PA: The Dialogue Press, 1978), p. 5.
25. See Bertrand Russell's classic article, "Mathematical Logic as Based on the Theory of Types," *American Journal of Mathematics* 30 (1908): 222-62.
26. *Ibid.*, p. 224.
27. *Ibid.*, p. 225.
28. I am, of course, leaving aside the possibility of 'three-valued logics' such as 'quantum logic' which purport to dispense with the law of the excluded middle; at any rate, the possibility of alternative logics does not affect the main argument of this section, for even if we decided that self-referentially inconsistent theories and propositions had some truth value other than true or false, they would still not be true.
29. *Ibid.*, p. 236.
30. *Ibid.*, p. 237.
31. *Ibid.*, pp. 236-38. For a good secondary summary cf. Irving Copi, *The Theory of Logical Types* (London: Routledge and Kegan Paul, 1971).
32. *Ibid.*, p. 238.
33. Paul Weiss, "The Theory of Types," *Mind* 37 (1928): 339. Similar points are made by Frederic B. Fitch, "Self-Reference in Philosophy," op. cit., and "Remarks on the Theory of Types," *Mind* 56 (1947): 184.
34. Cf. Andrzej Mostowski, "Review of Fitch's 'Self Reference in Philosophy,'" *Journal of Symbolic Logic* 11 (1946): 95-96; cf. Fitch's reply "Remarks on the Theory of Types," *ibid.*
35. Fitch, "Self-Reference in Philosophy," op. cit., p. 71.
36. Paul Feyerabend, *Farewell to Reason* (London: Verso, 1987), esp. ch. 1, "Notes on Relativism," pp. 19-89.
37. *Ibid.*, p. 83.
38. *Ibid.*, p. 73.
39. *Ibid.*, p. 78. Actually, readers of earlier Feyerabendiana (e.g., Part Three of *Science and a Free Society* [London: New Left Books, 1978] which is entitled "Conversations With Illiterates") will consider this excessively mild.
40. Cf. *Farewell to Reason*, op. cit., pp. 1-17, 273-79, as well as Part Two of *Science in a Free Society*, *ibid.*, and ch. 18 of *Against Method* (London: New Left Books, 1975).
41. *Farewell to Reason*, op. cit., p. 78.
42. It seems to me that Stanley Fish's recent anti-foundationalist rhetoric is in a situation almost identical to that of Feyerabend; cf. his tome *Doing What Comes Naturally* (Durham NC: Duke University Press, 1989). I submit that if Fish were really "doing what comes naturally" he would give up deconstructionist jargon immediately and go back to talking about the World Series. For whatever else we may say of it, going on and on for hundreds of pages about language is not "natural" (in Fish's sense of this word, assuming he has one). Natural language is about tables and chairs, dogs and cats, people, the World Series, sunsets, wars, hurricanes, final exams, and other such; only in rare circumstances do our linguistic usages become self-referential and reflect on its own structure and "natural" properties. When this happens, we are by definition no longer using language "naturally".

43. Cf. my "The Skeptic's Dilemma," op. cit., n. 8. This article was a response to Professor Davis's criticism of my earlier "Rorty's Foundationalism," op. cit., n. 7. (cf. William H. Davis, "In My Opinion That's Your Opinion: Is Rorty a Foundationalist?" *Reason Papers* 14 (1989): 137-42. Professor Davis had wondered why an antifoundationalist (or deconstructionist or relativist) could not speak the language of foundationalists (or absolutists), inviting them to open their eyes to the history of ideas and survey the "strife of systems" (p. 138). This misses the point; for obviously he can speak any language he wants. But if the antifoundationalist's (or deconstructionist's or relativist's) conclusions are meant to be taken seriously, then why should we believe his description of discourse? On what grounds can discourse be "transparent" to the eye of the antifoundationalist (or deconstructionist or relativist) if, as he maintains, there are no privileged vocabularies (which, self-referentially, would include vocabularies about discourse)?
44. Jack W. Meiland, "On the Paradox of Cognitive Relativism," *Metaphilosophy* 11 (1980): 121. Emphasis his.
45. Jack W. Meiland, "Concepts of Relative Truth," *The Monist* 60 (1977): 568-82.
46. *Ibid.*, p. 574.
47. Siegel, "Relativism, Truth and Incoherence," op. cit., p. 234.
48. Cf. also *ibid.*, p. 256, n. 30.
49. Andrew D. Cling, "Eliminative Materialism and Self-Referential Inconsistency," *Philosophical Studies* 56 (1989), esp. pp. 58-59. This valuable article gives similar logical reconstructions for both meaningfulness and assertibility as conditions of scientific discourse, all in accordance with a general schematic for constructive self-referential arguments.
50. *Ibid.*, p. 59.
51. *Ibid.*, p. 60.
52. *Ibid.*
53. Peter Smith, "Eliminative Materialism - A Reply to Everitt," *Mind* 91 (1982): 439.
54. Cling, op. cit., p. 71
55. I intend to take this issue up in more detail in my "Eliminative Materialism and the Incommensurability Thesis," in progress.
56. Some philosophers, friends of eliminative materialism among them, might object at this point by reiterating that part of their argument is that there are no theory-neutral observations; as Norwood Russell Hanson insisted, following Duhem, Wittgenstein and Ryle before him, all scientific observations are "theory-laden," (*Patterns of Discovery* (Cambridge: Cambridge University Press, 1958)) or fully theoretical (Paul Feyerabend, "An Attempt At a Realistic Theory of Observation," *Proceedings of The Aristotelian Society, New Series* 58 (1958): 143-70). But critics of the theory-ladenness thesis have made clear that if taken literally or interpreted in too strong a form it would have the result that no theory can ever be criticized on the basis of observations, and that each scientist (or scientific community) would be effectively isolated in the "world" created by his theoretical system. This would give the friend of eliminative materialism no foothold in folk psychology from which to launch the criticisms he wants, much less establish pragmatic consistency. See Israel Scheffler, *Science and Subjectivity* (Indianapolis: Bobbs-Merrill, 1967) or Carl R. Kordig "The Theory-Ladenness of Observation," *Review of Metaphysics* 24 (1971): 448-84.
57. This notion is, of necessity, vague; part of the problem is that we don't really know as of yet how our brains store information and relate it to the world outside our nervous systems; we simply have to wait for neuroscience to catch up to whatever philosophical speculations we have. My sense is that the friends of eliminative materialism, contrary to what they say, have not been willing to wait; for clearly eliminative materialism is not a scientific thesis but a philosophical research program embedded in a distinct philosophy of science (for details cf. my "Eliminative Materialism and the Incommensurability Thesis," op. cit.).

58. An argument similar to this was presented by Richard Bernstein in "The Challenge of Scientific Materialism," *International Philosophical Quarterly* 8 (1968): 252-75, in criticism of the two seminal articles articulating the "disappearance theory of the mind," Paul Feyerabend, "Materialism and the Mind-Body Problem," *Review of Metaphysics* 17 (1963): 49-66, and Richard Rorty, "Mind-Body Identity, Privacy and Categories," *Review of Metaphysics* 19 (1965): 24-54.
59. Paul Churchland, "Eliminative Materialism and the Propositional Attitudes," *Journal of Philosophy* 78 (1981): 89-90. Emphasis in original. Cf. also Patricia Smith Churchland, "Is Determinism Self-Refuting?" *Mind* 90 (1981): 99-101.
60. J. L. Mackie, "Self-Refutation: A Formal Analysis," *Philosophical Quarterly* 14 (1964): 203.
61. Cf. for example his main tract *Against Method: Outline of an Anarchistic Theory of Knowledge* (London: New Left Books, 1975), ch. 17. This intricate, and highly intriguing chapter, intended to illustrate the alleged incommensurability of cross-cultural discourses, makes use of all of these and others besides.
62. Cf. his "Putnam on Incommensurability," reprinted in *Farewell to Reason*, op. cit., p. 267.
63. Cf. his *Word and Object* (Cambridge: MIT Press, 1972), chs. 1 and 2.
64. One thinks here in particular of the defenders of the Copenhagen Interpretation of Quantum Mechanics.
65. For example, the self-referential argument against eliminativematerialism above may have established that we cannot eliminate beliefs; but it did not give us a theory of beliefs and other intentional states. That is a task of cognitive science (in which the philosopher, of course, is free to participate). A good place to start putting together an appropriate theoretical framework, I will suggest, might be a number of articles by the biologist and neuroscientist R. W. Sperry, particularly "A Modified Concept of Consciousness," *Psychological Review* 76 (1969): 532-36; "An Objective Approach to Subjective Experience: Further Explanation of a Hypothesis," *Psychological Review* 77 (1970): 585-90; "Mental Phenomena as Causal Determinants in Brain Function," in *Consciousness and the Brain*, eds. G. Globus, G. Maxwell and I. Savodnik (New York: Plenum Press, 1976); "Changing Concepts of Consciousness and Free Will," *Perspectives in Biology and Medicine* 20 (1976): 9-19; et al. A close study of the implications of Sperry's research findings for analytical philosophy of mind and cognitive science has not to my knowledge been attempted but I suspect that such a study would reveal that philosophy's materialists (whether reductive, functionalistic or eliminative) have all been barking up the wrong tree.
66. *Metaphysics*, IV, 1005b, in *Basic Works of Aristotle*, ed. Richard McKeon, trans. W. D. Ross (New York: Random House, 1941), p. 736.
67. A. J. Ayer, *Language, Truth and Logic*, 2nd. ed., (New York: Dover Books, 1946), p. 77.
68. We are assuming, of course, that both tokens of *house* and of *Elm Street* are tokens of the same type, and so refer to the same items and place in both occurrences; otherwise the given statement will not really be contradictory but will only maintain an appearance of contradiction.
69. See Douglas B. Rasmussen, "Aristotle and the Defense of the Law of Contradiction," *The Personalist* 54 (1973): 149-62.
70. I do not deny or overlook the fact that it is now commonplace among deconstructionists in particular to hold correctly that in a non-trivial sense words indeed do mean different things to different people on the same occasion of use. But it is also worth noting that deconstructionists, rather like Feyerabendian relativism, present us with a position so slippery that it manages to avoid systematic criticism with evasions like. "If you have attempted a systematic exposition and criticism this shows that you haven't

really understood deconstruction." Deconstructionism is, however, also notorious for having presented a jargon-laden no-man's-land of impenetrability - perhaps in unintentional confirmation of Aristotle's view.

71. Aristotle, op. cit., 1006b, 5, p. 738.
72. Steven J. Bartlett, "The Idea of a Metalogic of Reference," *Methodology and Science* 9 (1976): 89.
73. Richard Rorty, *Philosophy and the Mirror of Nature* (Princeton: Princeton University Press, 1979), Part Three.
74. This is not to say that philosophers can safely dispense with investigating them; they cannot.
75. Cf. Chapters Eight, Ten and Eleven of my book *Civil Wrongs: The Errors and Consequences of Affirmative Action* (under consideration for publication), or the paper "Why Philosophers Should Not Support 'Multiculturalism'", in preparation.
76. C. S. Peirce offers his argument against Cartesian certainty as a criterion for knowledge in "Some Consequences of the Four Incapacities," in J. Buchler, ed., *Philosophical Writings of Peirce* (New York: Dover Books, 1940, 1955), esp. pp. 228-29.
77. I owe this way of putting the matter to Tibor Machan (private conversation).
78. Steven J. Bartlett has made some strides in the latter direction; cf. his "Varieties of Self-Reference," in *Self-Reference: Reflections on Reflexivity*, eds. Steven J. Bartlett and Peter Suber (Dordrecht: Martinus Nijhoff, 1987), pp. 5-29.
79. Cf. e.g., John Kekes, *The Nature of Philosophy* (Totowa: Rowman and Littlefield, 1980); for related views see Nicholas Maxwell, *From Knowledge To Wisdom: A Revolution in the Aims and Methods of Science* (London: Basil Blackwell, 1984) and Mary Midgley, *Wisdom, Information and Wonder: What Is Knowledge For?* (London: Routledge, 1989).
80. I am grateful to Greg Johnson for comments which led to significant improvements in the final version of this paper.