Feyerabend and Incommensurability

By James Graham Ryan

for the M. Litt degree at the University of Durham.

This thesis is the result of research done in the Philosophy Department.

Submission date: 2 October 2002

Volume 1 of 1

Contents

ABSTRACT

Title: Feyerabend and Incommensurability

Author: Graham Ryan

I consider only the semantic claims of Paul Feyerabend's incommensurability thesis. These semantic claims are that incommensurable scientific theories, taken paradigmatically as successive theories: (1) are inconsistent; (2) the terms of one theory differ in meaning to those of another incommensurable theory; and (3) the claims of one theory are largely logically independent of the other. Since the inconsistency claim (1) is essential to Feyerabend's argument (against the Received View on theory reduction and explanation), I claim that (2) and (3) must be understood in the light of (1), and that (3) must be revised to avoid contradiction with (1). Feyerabend's semantic theory supporting (3) is presented and found wanting. Two other main arguments against (3) are also considered. The first is the causal theory of reference (of Putnam and Devitt), including causal descriptive theories advocated by Kitcher and Psillos; none of these theories is found to offer compelling reasons to reject (3). The second main argument against (3) is Donald Davidson's essay 'On the Very Idea of a Conceptual Scheme', and a close reading of Davidson's paper is offered. I find that Davidson does offer convincing reasons for rejecting any implication by (3) that the languages of incommensurable theories are not intertranslatable, or that such theories are closed cognitive frameworks. However, I agree with Larry Laudan that Davidson does not deliver a fatal blow to the semantic incommensurability thesis because: (a) incommensurability need not entail nontranslatability; and (b) Davidson's semantic arguments do not succeed in demonstrating that the very notion of a conceptual scheme is incoherent. I present briefly two versions of the semantic incommensurability thesis which are consistent in an interesting way with (1), (2) and a revised (3), namely taxonomic incommensurability and a model of misinterpretation in intractable conflicts.

Preface

I would like to thank Dr. Robin Hendry for his well-informed supervision and for having the perseverance to read not just this version but an untold number of drafts.

My biggest debt of gratitude is to my wife Irene. Without her support and interest this thesis would either never have been started or, if started, probably never have been completed.

Introduction

> "...incommensurability... although my ideas on the matter are pitiful, the objections are even more pitiful."[1]

It is now forty years since Paul Feyerabend published 'Explanation, Reduction and Empiricism', his first paper proposing the incommensurability thesis. In the same year, 1962, Thomas Kuhn proposed his own version of the incommensurability thesis, but Kuhn's *The Structure of Scientific Revolutions* has received the lion's share of academic attention.[2] The relative lack of attention given to Feyerabend's incommensurability thesis provides a reason for reconsidering his proposals and revisiting the debates around them.

Here, I consider the early semantic claims of Feyerabend's incommensurability thesis. These semantic claims are that incommensurable scientific theories, taken paradigmatically as successive theories: (1) are inconsistent; (2) the terms of one theory differ in meaning to those of another incommensurable theory; and (3) the claims of one theory are largely logically independent of the other. Since the inconsistency claim (1) is essential to Feyerabend's argument against the Received View on theory reduction and explanation, and the incommensurability thesis is a part of that argument, I claim that (2) and (3) must be understood in the light of (1), and that (3) must be revised to avoid contradiction with (1). Another reason for revising (3) is that Feyerabend's semantic theory supporting (3) is found wanting.

Opposition to (3) has come from many quarters. Two of the main arguments against (3) are considered. The first is the causal theory of reference (of Hilary Putnam and Michael Devitt), including causal descriptive theories advocated by Philip Kitcher and Stathis Psillos; none of these theories is found to offer compelling reasons to reject (3). The second main argument against (3) is Donald Davidson's essay 'On the Very Idea of a Conceptual Scheme', and a close

---

[1] Feyerabend, in a letter written to Imre Lakatos in 1971, Feyerabend (1999b), p. 237

[2] In 1977, Frederick Suppe believed he spoke for many when he said: "Feyerabend's philosophy of science has little to recommend itself and is losing whatever importance and influence it once had within philosophy of science." Suppe (1977), p. 643.

reading of Davidson's paper is offered. I find that Davidson does offer convincing reasons for rejecting any implication by (3) that the languages of incommensurable theories are not intertranslatable, or that such theories are closed cognitive frameworks. However, I agree with Larry Laudan that Davidson does not deliver a fatal blow to a slightly revised semantic incommensurability thesis because: (a) incommensurability need not entail nontranslatability; and (b) Davidson's semantic arguments do not succeed in demonstrating that the very notion of a conceptual scheme is incoherent. I present briefly two directions that a revised semantic incommensurability thesis, or theses, might take. These directions, namely taxonomic incommensurability and a model of misinterpretation in intractable conflicts, are consistent in an interesting way with (1), (2) and a revised (3).

Chapter 1

Problems With The Relations

[I]t is usually possible for the primitive concepts of an axiomatic system such as geometry to be correlated with, or interpreted by, the concepts of another system, e.g. physics. This possibility is particularly important when, in the course of the evolution of science, one system of statements is being explained by means of a new – a more general – system of hypotheses which permits the deduction not only of statements belonging to the first system, but also of statements belonging to other systems. In such cases it may be possible to define the fundamental concepts of the new system with the help of concepts which were originally used in some of the old systems.

Karl Popper (1959), *The Logic of Scientific Discovery*, p. 75.

I think that incommensurability *turns up* when we sharpen our concepts in the manner demanded by the logical positivists and their offspring and that it undermines their ideas on explanation, reduction and progress [...] but Kuhn used a different approach to apply the same term to a similar (not identical) situation. His approach was historical, while mine was abstract.

Paul Feyerabend (1993). *Against Method*, pp. 211-2.

Introduction

In 'Explanation, Reduction and Empiricism' (1962) Paul Feyerabend presents the incommensurability thesis (IT) as a denial and some additional proposals. That is, Feyerabend's presentation of the IT in 'Explanation, Reduction and Empiricism' is founded on the argument that the highly influential[1] Received View of scientific theories and theoretical change is false. The 'Received View' is a standard appellation given to logical positivist and logical empiricist views which regard scientific theories as languages with a clearly specified vocabulary and structure.

The Received View's formal description of the language of a scientific theory makes two general claims pertinent to Feyerabend's paper. *First*, "it embraces a 'hypothetico-deductive' view of theories"[2] such that a theory is a set of theoretical principles from which logically follow observation statements (and the observable consequences they state). *Second*, it assumes that observation statements "are scientifically and theoretically neutral, and nonproblematic with respect to truth"[3] and "that observational data are the bedrock on which theories ultimately rest"[4]. The two claims combine to assert that "a scientific theory is a deductively connected bundle of laws which are applicable to observable phenomena in ways specified by the correspondence rules."[5] Put this way, the Received View appears almost innocuous. However, Feyerabend argues that the Received View extends the two synchronic claims made above to diachronic assertions. For example, Ernest Nagel's view of theory reduction and Carl Hempel's (and Paul Oppenheim's) deductive-nomological model of scientific explanation augment the first claim by positing deductive relations *between successive theories*. The second claim implies that "highly confirmed theories are relatively immune from subsequent disconfirmation."[6] It is these further claims and the assumptions which underlie them that Feyerabend challenges.

---

[1] "It is little exaggeration to say that virtually every significant result obtained in philosophy of science between the 1920s and 1950 either employed or tacitly assumed the Received View." Suppe (1977), p. 4.

[2] Lambert, K. & Brittan, Gordon G. (1992), p. 92.

[3] Suppe (1977), p. 48.

[4] Lambert, K. & Brittan, Gordon G. (1992), p. 97.

[5] Suppe (1977), p. 36.

[6] Suppe (1977), p. 56.

Before Feyerabend argues against the Received View, he first describes it by attributing to it "the thesis of development by reduction"[7], the claims, mentioned in the previous paragraph, that "old theories are not rejected or abandoned once they have been accepted; they are just superseded by more comprehensive theories to which they are reduced."[8] Feyerabend maintains that this thesis of theory development by reduction, as well as deductive-nomological explanation, place three constraints (to be described in the coming Parts) on the relations between successive theories, namely:

the derivability condition

the consistency condition

the meaning invariance condition

Feyerabend then criticises these three alleged conditions. Having shown that the three conditions are untenable for general theories in a common domain, Feyerabend then argues that pairs of successive theories which do *not* meet the above conditions, and which also meet some further conditions which he stipulates, are *incommensurable.*

Part 1 of this chapter considers whether Feyerabend is right to attribute the derivability and consistency conditions to the Received View of theory reduction; and Part 2 considers the same question with respect to the Received View of explanation. Part 3 deals with the attribution of the meaning invariance condition to the Received View. From the first three Parts, I conclude, with certain reservations, that Feyerabend rightly attributes the derivability, consistency and meaning invariance conditions to the Received View.

In Part 4, Feyerabend's objections to the Received View are stated and his arguments are judged valid, though doubts are raised about their soundness: some of the premises expressing proposals of the meaning variance thesis (MVT) are problematic, particularly the claim that successive theories may be logically independent in a common domain. Part 5 has three main tasks. The first is to state the four main claims of the MVT and to highlight problems. The second task of Part 5 is to claim that the MVT expresses the semantic claims of IT. It is the *semantic* problems of the IT, that is, the MVT, which will be addressed in the subsequent

[7] Suppe (1977), p. 56.

[8] Suppe (1977), p. 56.

chapters. Part 5's third task is to argue that, as a result of Feyerabend's arguments against the Received View, what he calls two 'incommensurable' theories are first and foremost mutually inconsistent; so the claim of logical independence will need to be revised.

Part 1: Reduction

Feyerabend attributes the derivability and consistency conditions to Nagelian reduction as a representative of the Received View. The burden of Part 1 is whether these two conditions are legitimately attributed to Nagel and the Received View of theory reduction. Throughout this and subsequent chapters, let 'T_1' and 'T_2' signify two general theories which have a common domain[9] and are such that T_2 succeeds, and is wider[10] than, T_1.

The derivability condition is the claim that the sentences of T_1 are a logical consequence of the sentences of T_2. Attributing this condition to Nagelian reduction is not entirely straight-forward. Nagel himself maintains that "[t]he objective of reduction is to show that the laws, or the general principles of the secondary science are simply logical consequences of the primary science."[11] Here, it looks as if Nagel posits derivability between sciences, not theories. A (branch of) science may be defined by the problems which that science is concerned with, along with characteristic methods and techniques. A theory, according to the Received View, is an explanatory and predictive system composed of:

> an abstract calculus whose postulates [...] 'implicitly define the basic notions of the system' [...and] correspondence rules, relating theoretical notions to 'observational procedures' [...] or 'experimental concepts'.[12]

Reduction of one science, such as Biology or Psychology, to another, such as Chemistry or Physics, is therefore not simply the same as reduction of one theory to another. A further apparent deviation from the derivability condition is that Nagel's words make no mention of the *succession* of one science by another (T_2, it will be recalled, is T_1's successor).

---

[9] Let us say that a domain is a body of information of a problematic nature, but with a suspected underlying unity. See Suppe (1977), p. 239.

[10] 'Wider' in the sense of applying to more phenomena and therefore able to make a greater variety of predictions.

[11] Nagel, quoted by Feyerabend (1962), p. 33.

The gist of the two concerns just raised is that Nagelian reduction is a thesis about the logical unity of the sciences, and not about the historical development of theories. Nagel illustrates such a unity in his layer cake model where scientific knowledge is structured as sentences of different kinds, the different kinds pictured in different layers. The bottom layer contains sentences expressing facts; the next layer contains sentences expressing empirical generalisations; then there are sentences expressing theoretical laws; and on the top there are increasingly more abstract or general theories. The sentences in one layer are linked deductively to those in the next layer so that sentences of the lower layers can be derived from the upper layers, but not vice-versa. The layer cake model of science illustrates the claims that "science tends towards a more unified structure"[13] (the *ne plus ultra* of which is "a theory which holds all natural phenomena in its deductive embrace"[14]) and that science is based on empirically knowable facts.

I will now reply to the two concerns expressed about the applicability of the derivability condition to Nagel and, in so doing, will make reference to the layer cake model. The first concern was that reduction in Nagel's view concerns the derivability of one *science* from another, not of one theory from another. Nagel does not seem to regard the distinction between sciences and theories as important in this context. For example, the layer cake model of the logical unity of the *sciences* does not contain the notion of *a* science! Hence the criticism that the Received View "suggests that to reduce one branch of science [...] to another [...] is simply to reduce one theory to another."[15] The second concern was that Nagel's notion of theory reduction has no historical import. But the layer cake model implies that science *does* develop more and more general theories which stand in a particular relation to their less general predecessors; so from our *current* well-confirmed theories we *can in principle* (the 'layer cake principle') derive our previous well-confirmed theories. *In practice*, such derivability may occur only after suitable connecting statements have been ascertained, but the establishment of such connecting laws merely offers empirical confirmation of the 'layer cake principle'; for Nagel expresses no doubt about the historical nature of theory reduction:

---

[12] Feyerabend (1964b), in PP2, p. 53, using quotations from Nagel.
[13] Lambert, K. & Brittan, Gordon G. (1992), p. 93.
[14] Lambert, K. & Brittan, Gordon G. (1992), p. 159.
[15] Lambert, K. & Brittan, Gordon G. (1992), p. 159.

> the phenomenon of a relatively autonomous theory becoming absorbed by, or reduced to, some other more inclusive theory is an undeniable and recurrent feature of the history of modern science.[16]

I conclude that Feyerabend is justified in attributing the derivability condition to Nagel. The view just stated in Nagel's own words will now be considered further, a view described in the introduction as 'the thesis of development by reduction'.

Frederick Suppe helpfully lists four conditions for Nagelian theory reduction:

(a) The theoretical terms of T_1 and T_2 must have "meanings unambiguously fixed by codified rules of usage or by established procedures appropriate to each discipline."[17]

(b) For each theoretical term, *a*, of T_1 not found in T_2, "assumptions must be introduced which postulate relations between whatever is signified by *a* and traits represented by theoretical terms"[18] in T_2's vocabulary.

(c) Using, if need be, the assumptions in (b), all the laws of T_1 "must be logically derivable from the theoretical premises and their associated correspondence rules"[19] in T_2.

(d) "these additional assumptions must have evidential support."[20]

Condition (c) expresses the derivability condition. When condition (b) is required, the reduction is termed 'inhomogeneous'; otherwise 'homogeneous'. Condition (b) is employed to ensure that (c); hence Thomas Nickles' remarks that "Nagel's strategy, in effect, is to turn heterogeneous [i.e. inhomogeneous] reduction into homogeneous reduction"[21], and that "Nagel's treatment of all reduction [is] derivational; in the final analysis he too casts all reduction in essentially the same mould."[22] While such comments serve to support Feyerabend's application of the derivability condition to Nagel, these comments merely support a conclusion that we have already reached. Of more interest is Nickles' further remark that "Nagel's analysis of reduction is best regarded as a treatment of domain-combining reduction only."[23] By 'domain-combining' Nickles means that T_1 is *not* shown as defective within a certain domain by T_2, and that, instead, "ontological reduction and consolidation of theoretical postulates"[24] occur. So Nagelian theory reduction proposes that T_2, in reducing T_1,

---

[16] Nagel, quoted in Preston (1997), p. 81.
[17] Suppe (1977), p. 55, quoting Nagel.
[18] Suppe (1977), p. 55.
[19] Suppe (1977), p. 55.
[20] Suppe (1977), p. 55.
[21] Nickles (1973), p. 186.
[22] Nickles (1973), p. 187.
[23] Nickles (1973), p. 187.
[24] Nickles, in Suppe (1977), p. 586.

absorbs rather than fundamentally replaces T_1. A little more detail about this view of theory development or scientific progress will now be given.

The Received View understands scientific progress in three ways[25]:

(e) T_1 was once highly confirmed, but subsequent developments, such as better measuring instruments, revealed T_1 to be predictively inadequate. T_2 is the well-confirmed alternative.

(f) T_1 remains predictively adequate within its original domain, but T_2, also well-confirmed, encompasses the original domain and more. So T_2 is an expansion of T_1 by using correspondence rules which increase the scope of T_1.

(g) "various disparate theories, each enjoying high degrees of confirmation, are included in, or *reduced to*, some more inclusive theory"[26], T_2. Here the theoretical principles of the previous theories are altered, and possibly their correspondence rules.

However, the Received View regards way (e) as improbable because it holds that, once T_1 is confirmed, "it is highly unlikely that the theory can ever be disconfirmed."[27] Correspondence rules are, according to the Received View, partly constitutive of a given theory; new measuring instruments, and such like, would entail additional correspondence rules for T_1, thereby constituting (according to the Received View), a new theory T_1*. The disconfirmed theory will therefore be T_1*, not T_1. In this way, the Received View holds: "once it enjoys a high degree of confirmation, a theory [T_1] is unlikely to be disconfirmed; rather, any disconfirmation will be of extensions of [T_1] to scopes wider than that of [T_1]."[28] This formal flourish discounts (e), leaving (f) and (g) as the Received View's preferred descriptions of scientific progress. In Nagelian terms, (f) describes homogeneous reduction, and (g) inhomogeneous.[29] So the thesis of development by reduction proposes a cumulative view of scientific theory development in which old successful theories are extended, or absorbed, by new successful theories which make a greater range of successful predictions. Crucially (for Feyerabend's argument), there is more to the thesis of development by reduction than cumulativity – there is also the relation of reduction. In (f) and (g), where Nagel and the Received View meet is in

[25] See Suppe (1977), p. 53.
[26] Suppe (1977), p. 53.
[27] Suppe (1977), p. 54.
[28] Suppe (1977), p. 54.
[29] See Suppe (1977), p. 54.

the claim that old theories are "superceded by more comprehensive theories to which they are reduced."[30]

The foregoing comments serve two main purposes. The first purpose is to clarify that the Received View and Nagel, in holding the thesis of development by reduction, hold not merely the claim T_1 and T_2 make the same *predictions* within a common domain, but that the Received View and Nagel also hold the reductionist claim that *theory* T_1 is derivable from T_2 within the common domain. When D_1 is the domain of T_1, and d "expresses, in terms of [T_2] the conditions valid inside D_1",[31] the thesis of development by reduction then implies the derivability condition:[32]

T_2 & d |- T_1

The second purpose of the previous comments is to support Feyerabend's attribution of the second condition, the consistency condition, to the Received View of theory reduction.

The consistency condition states that T_1 and T_2 are mutually consistent within the common domain. Since consistency is a semantic relation, it may be useful to state the derivability condition semantically:

T_2 & d |= T_1

The difference, then, between the derivability condition and the consistency condition is as follows. If the sequent expressing the derivability condition is semantically valid, then there is *no* interpretation under which (T_2 & d) is true and T_1 is false. The consistency condition claims: T_1 and T_2 are mutually consistent only if there is an interpretation under which T_1 is true *and* T_2 is true. We have seen previously that the thesis of development by reduction proposes just this interpretation: T_1 and T_2 are *both* true inside the common domain. Or we might put matters thus: since the Received View maintains that, in the common domain, T_1 is itself consistent; and that T_2 is a consistent theory; and that T_1 is a logical consequence of T_2; then T_1 and T_2 are mutually consistent. So because of the thesis of development by reduction, Feyerabend is right to attribute the derivability and consistency conditions to the Received View.

---

[30] Suppe (1977), p. 56.

[31] Feyerabend (1962), p. 46.

Part 2: Explanation

The second designated representative of the Received View is the Hempelian deductive-nomological (D-N) model of scientific explanation. This model requires that a scientific explanation take the form of a deductively valid argument. The premises of the argument, the explanans statements, must express at least one general law of nature, and may include statements of antecedent conditions. The conclusion of the argument, the explanandum statement, describes the explanandum phenomenon.

In order to ascribe the derivability condition to Hempel, it must be shown that his D-N model of scientific explanation is also a model of the historical development of scientific theories; in which case:

(h) T_2 is the explanans

(i) T_1 the explanandum.

Requirement (h) presents no major problem: explanans statements will be statements of laws from T_2 (and possibly condition(s) allowed by T_2). Part of requirement (i) is that Hempel must allow the explanandum phenomenon to be a predecessor *theory*.[33] Suppe thinks that this view *is* attributable to the Received View, for, "[o]n the D-N model, the explanandum, E, may be either a (description of an) event or a law or theory."[34] I have not found any place where Hempel explicitly allows the explanandum to be a theory. However, comments by J. Alberto Coffa link the explanation of a theory with the explanation of its laws: T_2 explains T_1 if and only if T_2 implies the laws of T_1.[35] Coffa maintains that this claim "is obvious"[36] and is part of "Hempel's deductive model"[37]. Hempel *does* explicitly claim that "Newton's theory accounts for Galileo's law of free fall"[38], and speaks of "the explanation, by the kinetic theory, of Boyle's law"[39]; and these examples show at least that Hempel allows that the *laws* of T_1 may be the explanandum of T_2. So it seems fair to allow that Hempel permits that one theory explains another, if by that we mean one theory may explain the laws of another.

---

[32] Stated formally in Feyerabend (1962), p. 46.

[33] "Explanation [...of T_1] consists in the derivation of [T_1] from [T_2] and initial conditions, which specify the domain [D_1] in which [T_1] is applicable." Feyerabend (1965a), p. 164.

[34] Suppe (1977), p. 620. There is more discussion of this later.

[35] Coffa (1967), p. 503.

[36] Coffa (1967), p. 503.

[37] Coffa (1967), p. 503. Coffa's claims are problematic.

[38] Hempel (1966), p. 76.

[39] Hempel (1966), p. 73.

Feyerabend maintains that the Received View's D-N explanation is a model of the past historical development of theories; after all, the Received View explicitly equates D-N explanation with theory reduction.[40] For example, Nagel makes the direct assertion that *reduction* is "the *explanation* of a theory or a set of experimental laws established in one area of inquiry by a theory usually, though not invariably, formulated for some other domain."[41] Thomas Nickles concurs that Nagelian theory reduction "amounts to a *deductive explanation* of the reduced theory."[42] Since it has been shown that the derivability condition can be ascribed to the Received View of theory reduction; and that the Received View equates theory reduction and D-N explanation; the conclusion follows that the derivability condition can also be ascribed to the Received View of D-N explanation. While it seems to be the case that, for the *Nagelian* aspect of the Received View, the Received View implies that D-N explanation is a model of the historical development of theories, it is not as easy to pin this view on Hempel. This is because Hempel does not regard it as a foregone conclusion (as Nagel seems to in his 'layer cake' model) that T_1 is reducible to T_2:

> Generally, then, the extent to which biological laws are explainable by means of psycho-chemical laws depends on the extent to which suitable connecting laws can be established. And that, again, cannot be decided by *a priori* arguments; the answer can be found only by biological and biophysical research.[43]

Insofar as T_1 is reducible to T_2, the latter explains the former; but Hempel does not seem to buy into the claim that all past theories are, even in *principle*, reducible to or explained by their current successors[44]. Where does this leave the derivability condition vis-à-vis Hempelian D-N explanation? What Feyerabend does show (rather obviously) is that cases which Hempel claims *are* cases of intertheoretic D-N explanation[45] are also cases which lay claim to the derivability condition.

In order to show that the consistency condition applies to those cases where the D-N model of explanation does obtain, it must be shown that Hempel proposes an

---

[40] See Suppe (1977), p. 623.

[41] Nagel, quoted in Feyerabend (1962), p. 34, n. 14. My italics.

[42] Nickles (1973), p. 184. Kenneth Schaffner also remarks: "Intertheoretic explanation in which one theory is explained by another theory, usually formulated for a different domain, is generally termed 'theory reduction'." Schaffner (1967), p. 137.

[43] Hempel (1966), p. 105.

[44] In his discussion of emergence in his 1948 paper 'Studies in the Logic of Explanation', reprinted in Hempel (1965), Hempel is also careful to *avoid* the claim that emergent phenomena in biology or psychology *will ever* be explained by psycho-physical theories.

[45] For example, Newtonian mechanics explaining the laws of Galilean free-fall or the kinetic theory explaining Boyle's law.

interpretation whereby both explanans and explanandum statements are true at the same time. Suppe points out that "Hempel and Oppenheim [1948] required that they [the explanans statements] be true"[46]; and if the premises in a deductively valid argument are true under a given interpretation, then the conclusion must also be true under that same interpretation[47]. Feyerabend makes the same point, but puts it more carefully: "the consequences of a satisfactory explanans [T_2] inside [d] must be compatible with the explanandum [T_1]"[48]. Since Hempel's model of scientific explanation implies that explanans and explanandum statements are both true, it interprets T_2 and (the laws of) T_1 in such a way as to meet the consistency condition (though Hempel does *not* hold that D-N explanation – or therefore, the consistency condition – hold *in principle* between any current theory and its predecessor).

Attributing the derivability and consistency conditions to Hempel has met with more limited success than their attribution to Nagel. Matters become even more strained when Feyerabend attributes to the Received View the following claim:

(j) "only theories are admissible (for explanation and prediction) in a given domain which either contain the theories already used in the domain, or are at least consistent with them."[49]

I will argue that, because of the Nagelian thesis of development by reduction, Feyerabend is justified in attributing (j) to the Received View. However, Hempel does not subscribe to the thesis of development by reduction, and that means that pinning (j) on him will be more problematic. I will also say why I think the attribution of (j) to the Received View has proved such a hot potato. A little later, I will consider a couple of objections to (j).

It seems to me that Feyerabend's remarkable move of attributing a forward-looking methodology to the derivability and consistency conditions can be justified on the grounds of the thesis of development by reduction. In holding that scientific progress is characterised by either the expansion of the scope of T_1 (i.e. (f)) or by the absorption of T_1 into T_2 (i.e. (g)), the Received View is expressing views consistent with (j).

---

[46] Suppe (1977), p. 620, .n 6. And Hempel [1948] in his (1965), p. 248: "The sentences constituting the explanans must be true."

[47] "[A] true explanation, of course, has a true explanandum as well." Hempel (1965), p. 338.

[48] Feyerabend (1965a), p. 164.

[49] Feyerabend (1962), p. 44.

The derivability and consistency conditions ((j) mentions *containment* and consistency conditions) were generally held, by members of the Received View, to be *retrospective* descriptions of the historical development of theories. Now Feyerabend wants to pin on the Received View the belief that the derivability and consistency conditions are adequacy conditions on *future* theory development. To claim that they are such has proved to be a contentious and even startling move on Feyerabend's part. Commenting on (j), John Preston agrees with comments by Cliff Hooker:

> Cliff Hooker, in an excellent discussion of this issue, correctly suggests that superficial criticisms of Feyerabend are made [...] because most philosophers of science, deploying the distinction between the 'context of discovery' and the 'context of justification', assume that the rules for a prospective methodology of science need not be related to those for a retrospective assessment of science.[50]

Hooker's argument – as given by Preston - in support of attributing (j) to the Received View, eschews mention of contexts of discovery and justification, and argues that it is only rational to expect that rules which apply retrospectively to our best theories will have prospective implications for our best theories. I find Hooker's argument convincing, and it shows in part why attribution (j) was so contentious: Hooker suggests that supporters of the Received View were being less than completely rational in denying the forward-looking methodology (attributed to them by Feyerabend) while at the same time holding the derivability and consistency conditions.

I wonder if Cliff Hooker's comments do not make clear *why* supporters of the Received View are so vehemently opposed to (j). For the Received View, whether in formalising theories, or in stipulating how they are to be tested, claims to deal with theories as finished products, and *therefore* the Received View holds derivability and consistency as relations between past, or current and past, *accepted* theories. The Received View does not view derivability and consistency as conditions for a theory to *reach* acceptance: the two conditions describing scientific progress are not, for supporters of the Received View, conditions used to *fashion* new theories, but rather to describe *successive finished products.* Hooker's argument, compelling though it is, does not (as Preston presents it) quite explain *why* the attribution of (j) as a forward-looking methodology was anathema to supporters of the Received View.

---

[50] Preston (1997), p. 85.

Feyerabend's attribution of (j) to the Received View is contentious because it implies that, if it is to avoid self-contradiction, then the Received View must acknowledge that its views on reduction and explanation (and the conditions of derivability and consistency) do not apply *only* to theories which are, or have been, accepted already; rather, such views on reduction and explanation also imply a forward-looking methodology. (j) is discomfiting for the Received View because the *rejection* of (j) would appear to be inconsistent with the Received View of theory reduction and explanation.

Coffa rebutts (j) by claiming that "Hempel believes that scientific progress consists in the explanation of more inclusive sets of facts [...not in] the explanation of preexistent theories"[51]. I think that Coffa's rebuttal is more than half right: Hempel does think that when progress occurs, T_2 explains more facts than T_1, and that the explanation of predecessor theories is not a *hard and fast* adequacy condition on progress. (It is not a hard and fast condition because Hempel did not think it a logical truth that T_1 is deducible from and consistent with T_2 in the common domain.) But I am inclined to think that, for Hempel, (j) is a soft and optional condition on progress. By this I mean that Hempel allows that there are occasions when a progressive successor theory does explain the laws of its predecessor, and on those occasions the derivability and consistency conditions must hold because:

> [I]n a sound explanation, the content of the explanandum is contained in that of the explanans. That is correct since the explanandum is a logical consequence of the explanans; but this peculiarity does not make scientific explanation trivially circular since the general laws occurring in the explanans go far beyond the content of the specific explanandum.[52]

If the derivability and consistency conditions sometimes apply retrospectively, then it seems reasonable (à la Hooker) to expect that the laws of our current theories will sometimes be derivable from and consistent with future theories. But, in a later paper, Hempel rejects even this soft view of (j):

> [Feyerabend] is completely mistaken in his allegation [...] that the conception of explanation by deductive subsumption under general laws or theoretical principles entails the incriminated methodological maxim [j]. Indeed, the D-N model of explanation concerns simply the relation between explanans and explanandum and *implies nothing whatever about the compatibility of different explanatory principles that might be accepted successively in a given field of empirical science.* In particular, it does not imply that a new explanatory theory may be accepted only on condition that it be logically compatible with those previously accepted.[53]

---

[51] Coffa (1967), p. 506.

[52] Hempel [1948] (1965), p. 276, n. 36.

[53] Hempel (1965), p. 347, n. 17. My italics.

The claim in italics and the claim which follows it are certainly not the same claim. The latter claim seems to me consistent with Hempel's other views, but the italicised claim strikes me as too strong. For on those occasions where T_2 will be found to explain the laws of T_1, the D-N model of explanation will imply something about the different explanatory principles of the successor theory, namely, that T_2 logically implies and is consistent with the laws of T_1.

While Coffa is surely right to deny the attribution of (j) to Hempel, is there not a toned-down version of (j), (j'), which does fit with Hempel's views, to wit:

(j') It is reasonable to expect that sometimes a theory which is under consideration explains (and so logically implies and is consistent with) our currently held theory (in a common domain). When this explanatory relationship is discovered, the theory under consideration is in a stronger position to be accepted than before the explanatory relationship was discovered.

Feyerabend is not therefore justified in attributing (j) to the Hempelian view of D-N explanation. But if one takes the received view to include the thesis of development by reduction (which Hempel did not hold), and since reduction and D-N explanation are logically the same, then it seems to me that (j) is generally attributable to the Received View.

William Newton-Smith concludes that Feyerabend's attribution of (j) to the Received View fails:

(k) (j) is not one of "the rules that philosophers and/or scientists have tended to assume are used in theory choice."[54]

Yet Newton-Smith's two claims in support of this conclusion (k) make little or no attempt to address the Received View, or any particular view. For example, Newton-Smith's first claim in support of (k) is:

> not even the most conservative of rationalists will deny that an unacceptable theory may have gained ascendancy.[55]

Newton-Smith's *stated* conclusion from this first premise is:

> Hence it cannot be a constraint on one who wishes to evaluate a new theory critically that it must agree with any *de facto* accepted theories.[56]

---

[54] Newton-Smith (1981), p. 129.

[55] Newton-smith (1981), p. 129.

[56] Newton-Smith (1981), pp. 129-30.

It seems to me that Newton-Smith fails to maintain a clear distinction between on the one hand what can or cannot (read 'ought' or 'ought not') be a constraint on theory choice (for even strict rationalists), and on the other hand what constraint those rationalists actually do hold. I made such a distinction already in the presentation of (j) and pointed out that the two issues are related in the following way: members of the Received View may not, or do not, actually hold (j); but since (j) is consistent with a large body of the beliefs of members of the Received View, the members of the Received View could logically hold (j), and should do because (j) follows from[57] their other views on reduction and explanation.

Newton-Smith's second claim in support of (k) is:

> What one wants to preserve when faced with a choice between new rival theories is not the old theory itself but the observational successes of that theory.[58]

Even if 'one's view' is taken to refer to the Received View, Newton-Smith's claim here is open to the distinction I have drawn between the Nagelian Received View and the Hempelian view. What William Newton-Smith *objects* to about (j) is this: it requires that "new hypotheses agree with accepted theories"[59] *is or was* "a rule which he [Feyerabend] takes to have been standardly held by philosophers of science"[60]. But I do not think that Feyerabend makes this claim; rather, I think Feyerabend argues that (j) is a consequence of the Received View of reduction and explanation. (Besides, Newton-Smith's remarks do not sit well with Hempel's comment "that conflict with a broadly supported theory militates against a hypothesis".[61])

I conclude that the attribution of (j) to the Received View is problematic for two main reasons. First, (j) is not implied by Hempel's views on explanation, though a soft version of (j) is. Second, (j) is on the one hand *generally* consistent with the Nagelian Received View's claims about theory reduction and explanation; on the other hand, (j) is *particularly* inconsistent with the Received View that reduction and explanation are adequacy conditions *only* on *past* theory development. With these reservations, then, it seems to me that Feyerabend is justified in ascribing the

---

[57] If their views include or imply the thesis of development by reduction.

[58] Newton-Smith (1981), p. 130.

[59] Newton-Smith (1981), p. 129. He is quoting Feyerabend whose italics I have removed.

[60] Newton-Smith (1981), p. 129.

[61] Hempel (1966), p. 40. Though, of course, a hypothesis is not a theory.

derivability and consistency conditions to the Received View's (but not Hempel's) model of D-N explanation. That the derivability and consistency conditions do ascribe prospective conditions on the Received View of theory development is a problem which is surely is to be laid at the door of the Received View, not at the door of Feyerabend.

Part 3: The Meaning Invariance Condition

The meaning invariance condition states that "the meanings of (observational) terms are invariant with respect to both reduction [of T_1 to T_2] and explanation [of T_1 by T_2]."[62] In presenting this condition I will first give an account of the Received View of observation language and its role in scientific theories. This brief description of the double-language model of scientific theories seeks to explain why, in stating the meaning invariance condition, Feyerabend encloses 'observational' in parentheses in the above quotation. Having considered the meaning invariance of observational and (in a slightly different way) theoretical terms, and concluded that such meaning invariance follows from the derivability condition, I then try to establish the relation between the meaning invariance condition and the consistency condition.

Examples of actual scientific theories would make the points made here in Part 3 clearer; but such examples will not be introduced until Part 4, for two reasons. First, I wish to try and avoid needless repetition of the same examples in Parts 3 and 4. Second, I think it is important to *formulate* carefully (as opposed to *illustrate*) what the meaning invariance condition is, so that it is clear what *proposal* Feyerabend will later oppose. Since the IT denies the meaning invariance condition, it is hoped that a clear idea of the meaning invariance condition will aid a clear notion of the meaning variance thesis embodied in the incommensurability thesis [IT].

According to the Received View, scientific theories can be expressed in a first-order language (L). The nonlogical primitive terms of L are divided into two classes: observation terms (V_O) and theoretical terms (V_T). V_O has a domain of interpretation consisting of "concrete observable events, things, or things-moments; the relations

[62] Feyerabend (1962), p. 43.

and properties of the interpretation must be *directly observable.*"[63] So sentences of L which contain V_O, but not V_T, are completely interpreted. V_T is partially interpreted by two kinds of postulates: first, theoretical postulates (TP) are the theoretical laws which interpret a theoretical term in terms of other theoretical terms; second, correspondence rules (C) are mixed sentences in the sense that each rule contains "at least one V_O term and at least one V_T term essentially or nonvacuously"[64]. An important point (which will be returned to) concerning the relation of C to TP is:

(I) C must be logically compatible with TP.

C must also be finite in number, and have a domain which "may be construed as the sum total of admissible experimental procedures for applying the theory to observable phenomena."[65] All the descriptive terms of L are expressed in TP & C, and "a theory is the set of all logical consequences of the conjunction of [TP-] and C-postulates."[66]

The interpretation of V_T terms by C is only partial, and is so in two ways. First, C does not define any V_T term: C sets constraints on the meaning of any V_T term. Second, not every V_T term is in C; the rest of the V_T terms will be interpreted in TP so that "as used in [TP & C], the theoretical terms must admit of such and such observational manifestations of the systems described by [TP & C]."[67] This limited interpretative role of C is described by Carnap:

> All they [i.e. C-rules] do is, in effect, to permit the derivation of certain sentences of [the observation language] from certain sentences of [the theoretical language] or vice versa. They serve indirectly for the derivations of conclusions in [the observation language], e.g., predictions of observable events, from premises in [the observation language].[68]

Without C, no V_T term would have an observational interpretation; and so without C "a theory would have no explanatory power [...] it would also be incapable of test"[69] as far as the Received View is concerned. Since V_O terms are fully interpreted by *direct observation*, their meanings are theory neutral: "any two observers who possess the words from V_O used in [observation statements], regardless of their scientific or theoretical background, will be able to agree upon the truth of such V_O assertions."[70] This theory-neutrality allows the performance of crucial tests of a

---

[63] Suppe (1977), p. 51. My italics.
[64] Suppe (1977), p. 25.
[65] Suppe (1977), p. 25.
[66] Psillos (1999), p. 41.
[67] Suppe (1977), p. 103, n. 213.
[68] Carnap, quoted in Suppe (1977), p. 87.
[69] Hempel (1966), p. 74.
[70] Suppe (1977), p. 48.

theory by the comparison of statements of observational predictions with statements about what is observed.

It is now clear that in the Received View, V_O in T_1 will have the *same* interpretation as V_O in (T_2 & d). Since its domain is the directly observable, the interpretation of V_O also tends to be incorrigible; and so V_O has not just a *common* interpretation (in T_1 and T_2) it has an historically *invariant* interpretation. This would explain why Feyerabend ascribes the meaning invariance of observational terms to the Received View.

Yet Feyerabend's presentation of the Received View does not stop at the meaning invariance of V_O. The invariant observation language ensures that, for each theory (T_1 and T_2) in a common domain, V_T terms partially interpreted by C receive common partial interpretations; these common, partial, observational interpretations of V_T are fed into the *other* V_T terms via TP. All *interpreted* V_T terms, as far as the Received View is concerned, have observational import either directly via C, or indirectly in TP via V_T terms from C.[71] Carnap notes that TP alone does not, strictly speaking, interpret any V_T term:

> All the interpretation (*in the strict sense of this term, i.e. observational interpretation*) that can be given for [V_T] is given in the C-rules, and their function is essentially the interpretation of certain sentences containing descriptive terms, and thereby [V_T]."[72]

Nagel also remarks that without C, the descriptive terms of L are nothing more than bound variables:

> Without correspondence rules a theory is not even a statement ... as its descriptive terms (or rather those for which no rules of correspondence are given) have the status of variables"[73].

So C is necessary for the interpretation of all V_T terms. All interpreted theoretical terms correspond, in C, or in TP & C, with observational terms; and since the observational terms are meaning invariant, the interpretation (in the strict sense above) of theoretical terms will also be meaning invariant. Since C only partially interprets V_T terms, it is sometimes said that theoretical terms are meaning invariant insofar as their partial interpretation includes an observational core. It is

[71] "The terms of [V_T] obtain only an indirect and incomplete interpretation by the fact that some of them are connected by correspondence rules with observational terms, and the remaining terms of [V_T] are connected with the first ones by the postulates of T." Carnap, quoted by Feyerabend (1962), p. 41.

[72] Carnap, quoted in Suppe (1977), p. 86.

[73] Nagel, quoted by Feyerabend (1964) in PP2, p. 53.

probably for this reason that Feyerabend places 'observation' in parentheses when he states the meaning invariance condition: observation terms have fully interpreted invariant meanings; and all interpreted theoretical terms are partially interpreted by observation terms; so all theoretical terms are partially[74] meaning invariant in theory transition.

Nagel points out that theoretical terms do partially change meanings, and partially do not. Concerning the reduction of classical thermodynamics to the kinetic theory, he writes:

> It is certainly possible to redefine the word 'temperature' so that it becomes synonymous with 'mean kinetic energy' of molecules. But it is certain that on this redefined usage the word has a different meaning from the one associated with it in the classical science of heat, and therefore a meaning different from the one associated with the word in the statement of the Boyle-Charles law. However if thermodynamics is to be reduced to mechanics, it is temperature in the sense of the term in classical science of heat which must be asserted to be proportional to the mean kinetic energy of gas molecules. Accordingly ... the state of bodies described as 'temperature' (in the classical thermodynamical sense) is also characterized by 'temperature' in the redefined sense of the term.[75]

Nagel allows that a theoretical term's theoretical meaning postulates (TP) may change in the move from T_1 to T_2; hence the need for bridge principles (in the reduction of T_1 to T_2) to relate the sense of 'temperature' in T_1 to the sense of 'temperature' in T_2. But the main point of interest in Nagel's comments, as far as the meaning invariance condition is concerned, is not what changes, but what remains invariant with respect to the theoretical terms of T_1 and T_2. If inhomogeneous reduction is to occur, if bridge principles relate 'temperature' in T_1 with 'temperature' in T_2, then ("accordingly" as Nagel puts it) 'temperature' in each theory is co-referential. (The co-reference of theoretical terms is taken as given in homogeneous reduction). Nagel's requirement that theoretical terms co-refer in successive general theories expresses the meaning invariance condition for V_T terms. This requirement, even with regard to the particular example of 'temperature', will be discussed further by Feyerabend and by Putnam in Chapter 3.

Hempel also can construe meaning invariance as referential continuity when he writes about attempting the inhomogeneous reduction of Biology to Physics and Chemistry:

---

[74] Lambert et al., maintain that "the condition of 'meaning invariance' requires that the meaning of theoretical terms does not shift as new phenomena are described and explained." Lambert, Karel & Brittan, Gordon G. (1992), p. 98. In the text I claim that this claim is true if 'meaning' can be substituted by 'reference'; otherwise, the statement is only partially true.

> It would be very difficult to name even one biological term for which a physico-chemical synonym can be specified [...] But descriptive definition may also be understood in a less stringent sense, which does not require that the definiens have the same meaning, or intension, as the definiendum, but only that it have the same extension or application.[76]

Bridge principles are simply ways of capturing such co-references. Hempel's bridge principles posit not a relation of sense between identical sentences of T_1 and T_2 employing theoretical terms, but a preservation of the sentences' truth values. Such a preservation of truth values would be sufficient to secure the derivability condition. It will be recalled that the Received View holds not only that T_1 and T_2 (within a common domain) have the same observational consequences, but also that T_1 is derivable from T_2 within the common domain. In holding the derivability condition, the Received View *must* also hold the meaning invariance condition if it is to avoid an equivocation fallacy.[77]

While the Received View holds that the theoretical postulates (TP) are not subject to the meaning invariance condition in the same way as C (which links a theoretical term with a core observational meaning), any changes to TP in the transition from T_1 to T_2 are constrained by C. As already stated in (1), the Received View held that TP and C must be consistent. For example:

> [If] I incorporate an experimental procedure into [TP & C] as a correspondence rule involving the V_T term corresponding to electrons and assert [TP & C] so interpreted, I am committing myself to using 'electron' in such a way that its observational content includes that specified by the correspondence rule. [78]

What consistency between TP and C states is that *any* meaning postulate of a theoretical term must be consistent with (statements of) the observational consequences attributed to that term. But the Received View holds that the observational consequences of T_1 and T_2 are the *same* within the common domain. Consequently, the meaning postulates of all V_T terms of T_1 are *consistent* with the meaning postulates of all V_T terms of T_2. In Parts 1 and 2, the consistency condition has been described as consistency between successive theories; now it is clearer that this means consistency not merely between the observational consequences of T_1 and T_2, but between the theoretical axioms or laws of T_1 and T_2:

> the demand for meaning invariance implies that the laws of later theories be compatible with the principles of the context of which the earlier theories are part.[79]

---

[75] Nagel, quoted by Coffa (1967), p. 507.

[76] Hempel (1966), p. 103.

[77] As Suppe (1977), p. 172 observes.

[78] Suppe (1977), p. 92.

[79] Feyerabend (1962), p. 81.

So Feyerabend regards the meaning invariance condition as "a special case of"[80] the consistency condition.

I conclude that Feyerabend is right to ascribe the following views to the Received View. The observational and theoretical terms of T_1 and T_2 are "unambiguously fixed"[81] in each theory. This fixing is a complete and theory-independent interpretation in the case of observational terms, and is a partially theory-independent interpretation in the case of theoretical terms. For any theoretical term, theoretical meaning postulates must be consistent with C-rules (the observational core). In the transition form T_1 to T_2, then, observational terms will not undergo any change of meaning; and theoretical terms will not endure any change of meaning which would conflict with their (ineradicable) correspondence rules. In ascribing the meaning invariance condition to the Received View, Feyerabend attributes to that View the claim that the meanings of the primitive descriptive terms of T_1 "will not be affected by the processes of reduction"[82]. What is *absolutely* not affected by the process of reduction are the meaning postulates of V_O terms; but I have tried to show that, according to the Received View, the meaning postulates of V_T terms may be affected, but only to a *limited* extent (i.e. TP may change but must always be *consistent* with the meaning postulates of V_O terms). I venture that this *difference* in meaning invariance between the terms of V_O and V_T is marked by Feyerabend placing 'observational' in parentheses when stating the meaning invariance condition (given at the beginning of Part 3).

Part 4: Arguments Rejecting The Received View

Feyerabend has established (I have argued) that the Received View espouses the derivability condition, the consistency condition and the meaning invariance condition as conditions on the development of successive, well-confirmed theories. In opposing the Received View, however, Feyerabend does not argue against each of these conditions in turn. Instead, he chooses to argue against the consistency condition, implying that, in arguing against the consistency condition, he is also

[80] Feyerabend (1962), p. 81.
[81] Nagel, quoted by Suppe (1977), p. 55.
[82] Feyerabend (1962), p. 33.

arguing against the derivability condition.[83] Then he asserts that the arguments against the consistency condition can also be used against the meaning invariance condition. This final move is problematic, however, because having asserted that T_1 and T_2 are inconsistent, he cannot then logically claim (as the meaning variance thesis does go on to do) that T_1 and T_2 are logically independent.

Feyerabend asserts:

(m) the derivability condition "leads to the demand [...] that all successful theories in a given domain must be mutually consistent."[84]

If (m) is true (if the derivability condition implies the consistency condition) then Feyerabend can kill two conditions by aiming at one target - simply arguing against the consistency condition. The Received View requires that the theories T_1 and T_2 are *each* consistent. Placing each on either side of the turnstile in the derivability condition does then 'lead to the demand' that T_1 and T_2 are *mutually* consistent. So (m) is true with respect to the Received View of reduction and explanation. Consequently, Feyerabend can argue against the derivability condition by aiming at the consistency condition.

The second element in Feyerabend's attack on the Received View is the claim:

(n) "Using our [...] arguments against [the consistency condition] we may now infer the untenability, on methodological grounds, of meaning invariance as well."[85]

In Part 3 I tried to show that, according to the Received View, the meanings of all theoretical terms "are a function of observational consequences"[86] such that the theoretical axioms (TP) of a theory must be *consistent with* C postulates; and that since the meanings of observation terms in C are common to T_1 and T_2 (in the common domain), TP in T_1 will be consistent with TP in T_2 (in the common domain). Part 3 therefore supports Feyerabend's assertion that the meaning invariance condition, in proposing consistency between the theoretical postulates of T_1 and T_2 (and equivalence of observational postulates), *is* a special case of the consistency condition, so that claim (n) is true – provided that Feyerabend's arguments against the consistency condition are sound.

---

[83] "It is in this form [the form of the consistency condition as stated in (j)] that [the derivability condition] will be discussed". Feyerabend (1962), p. 44.

[84] Feyerabend (1962), p. 30.

[85] Feyerabend (1962), p. 81.

[86] Suppe (1977), p. 92.

The two most general supporting elements of Feyerabend's attack (in 'Explanation, Reduction and Empiricism') against the Received View are therefore the short-cut strategy just explained and the arguments produced against the consistency condition. I will now present the three kinds of argument against the consistency (and derivability) conditions. The meaning invariance condition will be addressed *after*.

The first type of argument against the derivability and consistency conditions is the argument from example. It states:

> most of the cases which have been used as shining examples of scientific explanation do not satisfy [the consistency condition] and [...] it is not possible to adapt them to the deductive schema.[87]

Feyerabend selects two examples; the first is where T_1 is Galilean physics and T_2 is Newtonian celestial mechanics. It will be recalled that D_1 is then the domain of T_1, and d "expresses, in terms of [T_2] the conditions valid inside D_1."[88] The laws of free-fall in T_1 maintain that the acceleration of a falling body as constant, whereas the laws of T_2 hold that acceleration increases the closer to Earth the body falls. According to Nagel, the derivability condition holds in this example, so that:

$T_2 \& d \mid\!- T_1$

However, Feyerabend points out that T_1 and T_2 give quantitatively different predictions,[89] for T_1 posits constant acceleration and T_2 posits variable acceleration. Feyerabend concludes:

> It is therefore impossible, for quantitative reasons, to establish a deductive relationship between [T_1] and [T_2], or even to make [T_1] and [T_2] compatible [i.e. consistent].[90]

The question is: has Feyerabend demonstrated what he claims to have?

Feyerabend has certainly pointed out that T_1 and T_2 are not mutually consistent, in which case:

(o) $T_2 \& d \mid\!- \sim T_1$

Taking up this matter of inconsistency, Hempel writes:

---

[87] Feyerabend (1962), p. 46.

[88] Feyerabend (1962), p. 46.

[89] Unless "the earth's radius is infinitely large, which it is not." Schaffner (1967, p. 138.

[90] Feyerabend (1962), p. 47.

> It might therefore be tempting to say that theories often do not explain previously established laws, but refute them. But this would give a distorted picture of the insight afforded by a theory. After all a theory does not simply refute the earlier empirical generalizations in its field; rather it shows that within a certain limited range defined by qualifying conditions, the generalizations hold true in fairly close approximation.[91]

Though "everybody would admit that explanation may be by approximation only"[92], Feyerabend maintains that everybody who holds the Received View is wrong to do so. Either one can derive T_1 from T_2 or one can not, for the turnstile does not allow for approximate derivation. Therefore the Received View's use of approximation in its formal account of reduction and explanation is, as it stands, in need of emendation. Lawrence Sklar, for example, feels that "somewhat more needs to be done to clarify the matter of approximate derivational reduction"[93]; and Schaffner[94] and Nickles[95] agree in their own ways. In the argument from example, Feyerabend claims that T_2 & d |- T_1, does not express a valid sequent when T_1 and T_2 are interpreted as in the free-fall example. So as a result of disproving the *consistency* condition, Feyerabend has shown that the sequent expressing the *derivability* condition is invalid.[96]

The second example (in the argument from example) takes T_1 is the medieval impetus theory and T_2 is Newtonian theory of motion. In considering whether T_1 and T_2 are consistent, Feyerabend considers the terms 'impetus' in T_1, and 'momentum' in T_2. "It has been suggested", writes Feyerabend, "that the momentum of the moving object is the perfect analogue of the impetus."[97] Three reasons in support of this Received View suggestion are, first, that it is possible to substitute 'momentum' for 'impetus' in such statements as: "The impetus of a body in empty space which is not under the influence of any outer force remains constant."[98] The first reason, then, for claiming meaning invariance of the terms 'impetus' and 'momentum', is that they are intersubstitutable salve veritate. Underlying the substitution claim is the assumption that Newton's first law and the impetus theory are consistent (within a common domain). Consistency between statements containing 'impetus' in T_1, and all but identical statements containing instead 'momentum' in T_2, is claimed by the

---

[91] Hempel (1966), p. 76.

[92] Feyerabend (1962), p. 48.

[93] Sklar (1967), p. 111.

[94] Schaffner (1967), p. 142.

[95] Nickles (1973), p. 188.

[96] One worry about this argument is the exact nature of the relation between the domain expressions D_1 and d. This point pre-empts discussion about a problem associated with the claim that theories are incommensurable, namely, in what sense can it be said that incommensurable theories have a common domain?

[97] Feyerabend (1962), p. 56.

Received View because the measurement of impetus, according to T_1, in all cases equals that of momentum in T_2. That impetus and momentum take the same numerical value is the second reason for claiming meaning invariance. Thirdly, even operationally considered, the procedures for measuring impetus and momentum are the same (in the common domain).

Feyerabend then points out that 'impetus' is *not* substitutable *salve veritate* between statements of T_1 and T_2 in the common domain. According to T_1:

> [I]mpetus is the force responsible for the movement of the object that has ceased to be in direct contact, by push or pull, with the material mover. If this force did not act, i.e., if the impetus were destroyed, then the object would cease to move and fall to the ground.[99]

In short, impetus is the force sustaining all motion, but momentum is not. This certainly appears to show inconsistency[100] between T_1 and T_2, even within the common domain. Feyerabend explains:

> what is being asserted is not the inconsistency of, say, Newton's theory and Galileo's law, but rather the inconsistency of some consequences of Newton's theory in the domain of validity of Galileo's law, and Galileo's law.[101]

This comment is useful because it says something about the nature of the common domain. The common domain of T_1 and T_2 is the domain were both theories are empirically adequate. Furthermore, as Chapter 2 will show, the common domain is the domain of *common causes* of the utterance of observation sentences of T_1 and T_2. More explication of the causal nature of the common domain is given in Feyerabend's pragmatic theory of observation, presented in Chapter 2.

Turning to the second reason given for equating 'impetus' and 'momentum', Feyerabend points out that they would *not* have the same numerical values in T_1 and T_2. Since impetus is the force sustaining motion, it is determined as the product of force and acceleration in T_2, not, as in T_1, as the product of mass and velocity (the value of momentum). So in a state of uniform motion in empty space, 'impetus' would have value 0, *according to* T_2, and such a value would *not* be equivalent to that given by the impetus theory, T_1.

---

[98] Feyerabend (1962), p. 54.

[99] Feyerabend (1962), p. 55.

[100] "Note that what is being asserted here is logical inconsistency", Feyerabend (1963b), p. 13.

Thirdly, an operational notion of impetus and momentum may claim that the way to measure each magnitude is to bring a body "to a stop in an appropriate medium (such as soft wax) and then noting the effect of such a maneuver"[102]. The degree to which the object is embedded in the wax will be an indicator or measure of the object's momentum just before impact; the degree of impact will also be a measure of the object's impetus. To try to use this operational definition of impetus and momentum as a bridge principle for the purpose of reducing medieval impetus theory to Newtonian mechanics, however, is problematic. The operational definition presupposes the further bridge hypothesis "that wherever momentum is present, impetus will also be present, and [...] the measure will be the same in both cases."[103] Feyerabend objects to this bridge hypothesis on the grounds that that T_2 forbids it. The theoretical postulates of 'impetus' (in T_1) conflict with axioms in T_2: T_2 claims that a body moving at constant velocity is *not* acted on by *any* force; but 'impetus' (according to T_1) is the force which sustains *any* motion. Bridge laws between the T_1 term 'impetus' and the T_2 term 'momentum' cannot be adopted because the axioms of T_2 imply that the magnitude impetus does not exist[104]. As was pointed out in Part 3, bridge principles attempt to capture co-references. If it can be plausibly argued that 'momentum' and 'impetus' *do* refer to the same magnitude, then the truth of statements about impetus will be preserved across theories in statements about momentum (and the substitution and numerical claims will be shown to be true). Such an argument, the causal theory of reference, will be considered in Chapter 3.

Feyerabend's second kind of argument against the derivability and consistency conditions is a conceptual argument about 'empiricism'. If the derivability and consistency conditions are adequacy conditions on scientific progress, as Feyerabend has alleged they *are* in the Received View, then only a theory which is consistent with the current theory, and from which the current theory can be derived, will reach acceptance. Empiricism requires that observations confirm laws and theories; and the Received View claims to be empiricist; yet the derivability and consistency conditions (for example, (j)) are such that a theory is rejected "not because it is inconsistent with the facts, but because it is inconsistent with another, and as yet unrefuted, theory whose confirming instances it shares."[105] As pointed out

---

[101] Feyerabend (1963b), p. 13.
[102] Feyerabend (1962), p. 54.
[103] Feyerabend (1962), p. 58.
[104] Or that, if impetus does exist, it has a constant value of zero.
[105] Feyerabend (1962), p. 64.

earlier, members of the Received View would probably deny that they follow any such test condition; and I argued (in Part 2) that, be that as it may, such a test condition is largely implied by the derivability and consistency conditions: they are anti-empiricist conditions.

The third argument against the derivability and consistency conditions is a methodological one and follows from the conceptual argument. The methodological *claim* is that "a strict empiricism will [admit] theories which are factually adequate and yet mutually inconsistent"[106] (such theories are called 'strong alternatives' by Feyerabend). The *warrant* for this claim is that "the basic principle of empiricism is to increase the empirical content of whatever knowledge we claim to possess."[107] The *'data' or ground* for the warrant for the methodological claim is Feyerabend's semantic views about observation statements. Feyerabend *denies* that "the facts which belong to the content of some theory are available whether or not one considers alternatives to this theory"[108]; rather, he believes:

(p) "[e]xperimental evidence does not consist of facts pure and simple, but of facts analyzed, modeled [sic], and manufactured according to some theory."[109]

Premise (p) claims that "the description of every single fact [is] dependent on some theory"[110]; and the methodological argument which (p) supports argues for the use of strong alternatives on the grounds that "[t]here exist also facts which cannot be unearthed except with the help of alternatives to the theory to be considered"[111]. For empiricists, factual statements must be derivable from observational statements, and in the Received View the latter are statements whose meanings are foundational. For Feyerabend, the meanings of observational sentences are *not* basic or foundational, and he argues such when presenting his pragmatic theory of observation (PTO). For Feyerabend, a consequence of the PTO is that "[m]eaning comes from ideas. Meaning, therefore 'trickles down' from the theoretical level toward the level of observation."[112] So for empiricists such as Feyerabend, the PTO will have (p) as a consequence *because* the PTO supports what Feyerabend labels *thesis I*:

---

[106] Feyerabend (1962), p. 67.
[107] Feyerabend (1962), p. 66.
[108] Feyerabend (1963b), p. 22.
[109] Feyerabend (1962), pp. 50-1.
[110] Feyerabend (1963b), p. 22.
[111] Feyerabend (1963b), p. 22.
[112] Feyerabend (1995), p. 118.

> *the interpretation of an observation language is determined by the theories which we use to explain what we observe, and it changes as soon as those theories change.*[113]

So the methodological argument against the derivability and consistency conditions, particularly premise (p), depends on *thesis I. Thesis I* (and the PTO) will be considered in Chapter 2.

Given (p) (for the sake of Feyerabend's argument), it follows that if only those theories consistent with the test theory are considered, then the range of facts or empirical evidence which can be considered is limited - without empirical reason. It is therefore only good empirical procedure to consider strong alternatives - those theories which are "partly overlapping, factually adequate, but mutually inconsistent"[114] with the test theory. The use of strong alternatives as a test procedure could yield successor theories which are *inconsistent* with their predecessors; so allowing the use of strong alternatives undermines the claim that the consistency and derivability conditions are necessary conditions on theory development.

Reviewing the three arguments against the consistency (and derivability) condition(s), it seems that all three kinds of argument are valid; and I have claimed that the arguments from example and the conceptual argument are sound. The methodological argument is more uncertain because *thesis I*, (p)'s support, has yet to be judged (in Chapter 2). But is clear that *thesis I* stands in contradiction of the meaning invariance condition, and in this way the methodological argument attacks the meaning invariance condition, as Feyerabend claims:

> It is also clear that the methodological arguments against meaning invariance will be the same as the arguments against the derivability condition and the consistency condition.[115]

It will be recalled that Feyerabend regarded the meaning invariance condition as a special case of the consistency condition. It comes as no surprise, then, that arguments against the consistency (and derivability) condition(s) are also used against the meaning invariance condition.

---

[113] Feyerabend (1958), p. 31. This claim will be considered further in Chapter 2.
[114] Feyerabend (1963b), pp. 22-3.
[115] Feyerabend (1962), p. 81.

Turning now to the arguments against the meaning invariance condition, that condition (described in Part 3) simply stated that the meanings of observation terms do not change in the transition from T_1 to T_2, and that the theoretical meaning postulates of terms of T_2 do not contradict those of T_1. From the argument from example it is clear that the theoretical meaning postulates (TP) of terms of T_2 *do* contradict (or imply a contradiction of) those of T_1; and the methodological argument claimed that the TP of T_2 generally *ought to* do so. This is where Feyerabend's argument against the meaning invariance condition begins:

> Our argument against meaning invariance is simple and clear. It proceeds from the fact that usually some of the principles involved in the determination of the meanings of older theories or points of view are *inconsistent* with the new, and better, theories.[116]

For example, impetus is the force which sustains motion, but momentum is not. This contradiction at the level of TP will, because there are correspondence rules, have an affect on the *observational meaning postulates*, otherwise each theory would not be internally consistent and each theory's predictions would not logically follow from that theory's laws. "Thus descriptions of the theories' observable predictions depend upon some theory (or theories)."[117] About this argument, Frederick Suppe tells us:

> It is worth noting that Feyerabend's view here comes as close as possible to a complete reversal of the Received View's picture of a one-way flow of meanings from the observation language to the theoretical language.[118]

This reversal of the 'flow of meanings' is essential to Feyerabend's main argument against the meaning invariance condition, and *thesis I* is the warrant Feyerabend uses to yield such a reversal.[119]

The argument from example and the methodological argument against the meaning invariance condition each work by pointing out that the theoretical postulates of T_2 contradict those of T_1; and that, because of *thesis I*, the observational postulates do also. Consequently, the meaning invariance condition is false.

The methodological argument's use of strong alternatives clearly assumes that "one and the same set of observational data is compatible with very different and

---

[116] Feyerabend (1962), p. 82. My italics.

[117] Suppe (1977), p. 176.

[118] Suppe (1977), p. 176, n. 430.

[119] But to challenge the meaning invariance condition, it is not necessary to postulate *thesis I* or a *reversal* of any flow of meanings. Having shown that the theoretical postulates of T_1 and T_2 are inconsistent, it would be sufficient for Feyerabend to argue that observational meaning postulates are not fixed or 'given' extra-theoretically to validly argue that the meaning

mutually inconsistent theories."[120] Carl Hempel also admits that "[a]ny type of empirical findings, however rich and diverse, can in principle be subsumed under many different laws or theories."[121] He cites as an example the particle and wave theories of light which were empirically adequate up until the crucial experiments of the nineteenth century. Where Feyerabend and Hempel differ (with respect to strong alternatives) is that, for Hempel, the situation where there are two empirically adequate and inconsistent theories is an anomaly which will eventually be rectified by an observationally based crucial experiment.[122] Feyerabend believes that, for global theories, "the alternatives do not share a single statement with the theories they criticize. Clearly a crucial experiment is now impossible."[123] What Feyerabend means by the theories' not sharing a single statement is:

(q) T_1 and T_2 "may not possess any comparable consequences, observational or otherwise".[124]

Claim (q) asserts that statements of T_1 and T_2 are semantically incomparable, and this is sufficient to warrant the denial of the ability to perform an observationally based crucial experiment for two global theories. However, claim (q) is problematic because, in addition to ruling out a crucial experiment between T_1 and T_2, it also implies that T_2 is not a strong *alternative* to T_1; for a consequence of (q) is that statements of T_1 and T_2 cannot be mutually inconsistent. This problem of combining Feyerabend's assertions that statements of T_1 are logically inconsistent with *and* logically independent of statements of T_2 crops up time and time again. This problem suggests that one of the two assertions has to go. Feyerabend tries to revise the claim of logical inconsistency (and so alter the notion of a strong alternative), as Chapters 2 and 3 will show. But it seems to me that Feyerabend *cannot* drop his inconsistency claim because his *opposition* to the *consistency* condition is the hinge on which turns his arguments for the IT.

The semantic incomparability of the terms of T_1 and T_2 expressed in statement (q) is made possible by *thesis I* permitting a complete meaning change in each and every term of successive general theories. Feyerabend muses:

---

invariance condition is false. And as Chapter 2 will show, Feyerabend does in fact argue that his contemporaries' two main accounts of extra-theoretically fixed meanings are false.

[120] Feyerabend (1962), p. 48.

[121] Hempel (1966), p. 80.

[122] See Hempel (1966), p. 80. This is another reason for suggesting that Hempel does sign up to some soft version of (j).

[123] Feyerabend (1963b), p. 8.

[124] Feyerabend (1962), p. 94. See also Feyerabend (1963b), p. 8: "there is no statement [of T_2] capable of expressing what emerges from the observation[s] [made by one who holds T_1].

> I interpreted observation languages by the theories that explain what we observe. Such interpretations change as soon as the theories change. I realized that interpretations of this kind might make it impossible to establish deductive relations between rival theories".[125]

The meaning change proposal concerns both the *number* of meanings changed and the *nature* of their change. Numerically, "there is a change in the meanings of *all* descriptive terms of [T_1] (provided these terms are still employed)"[126]. The nature of the meaning changes is such that it is:

> completely impossible either to reduce [the theories] to each other, or to relate them to each other with the help of an empirical hypothesis, or to find entities which belong to the extension of both kinds of terms.[127]

The proposal that *all* the terms of a theory are affected in this way has been called the *radical* MVT.[128] I will distinguish this from the MVT, where not all terms of T_1 and T_2 are affected in the way proposed. This distinction will come in useful in Chapter 3, where it will be shown that Feyerabend drops the radical MVT for the MVT.

Part 5: The IT and the MVT

As regards Feyerabend's first presentation of the IT, given in 'Explanation, Reduction and Empiricism', Frederick Suppe is of the opinion that "a legitimate objection lurks buried in this discussion"[129]:

> [T]he reduced theory often is false whereas the reducing theory is true, which precludes the required sort of sound deduction of the former from the latter augmented by further definitions and hypotheses.[130]

The Received View does tend to overlook or downplay inconsistency between successive theories. For example, the predictions of Newtonian physics concerning a body in uniform motion in empty space contradict those of the medieval theory. But for the Received View, claims Feyerabend:

> [I]t is natural to resolve this contradiction by eliminating the troublesome and unsatisfactory older principles and to replace them by principles, or theorems, of the new and better theory.[131]

The Received View ignores certain theoretical postulates of T_1 and the observational consequences of the laws which contain them, so that the observational

---

[125] Feyerabend (1978), p. 67.
[126] Feyerabend (1962), p. 59. My italics.
[127] Feyerabend (1962), p. 90.
[128] See Newton-Smith (1981), p. 155.
[129] Suppe (1977), p. 624.
[130] Suppe (1977), p. 624.
[131] Feyerabend (1962), p. 82.

consequences of T_2 will be regarded as including those of T_1. In uncovering the inconsistency to which the Received View had turned a blind eye, Feyerabend shows that the three conditions on theory development do not always hold. In this respect, Feyerabend's 1962 paper may be judged a success.

What Feyerabend's 1962 paper has argued is that T_1 and T_2 may be incommensurable by showing that they may be mutually inconsistent. Feyerabend himself describes the relations between two incommensurable theories (T_1 and T_2) thus:

> "[T]he use of [T_2] will necessitate the elimination both of the conceptual apparatus of [T_1] and the laws of [T_1]. The conceptual apparatus will have to be eliminated because it involves principles [...] which are *inconsistent* with the principles of [T_2]; and the laws will have to be eliminated because they are *inconsistent* with what follows from [T_2] for events inside D_1."[132]

It seems to me that the IT must retain the successful inconsistency claim and revise the logical independence claim which conflicts with it. So I would like to say that it is the interplay between the notions of inconsistency and something *like* logical independence (as notions which relate the statements of T_1 and T_2) which compose the IT. Such an interplay might be that T_1 *appears* illogical or irrational to a holder of T_2. This conception of the IT will come up again in Chapters 3 and 4.

It is surely not an uncommon opinion that the Received View's meaning invariance condition is (like the derivability and consistency conditions) too strict, and that the meanings of some terms of T_1 are altered in T_2. For example, Michael Devitt can write:

> I am sympathetic to the view that theory change often leads to meaning change and find the discussions of Kuhn and Feyerabend illuminating on that issue.[133]

The big issue, then, is not *that* meaning change occurs, but "the *nature* and *degree* of the semantic changes"[134] which Feyerabend proposes. The degree of meaning change is a question of numbers: are all terms changed or only some? The nature of meaning change proposed is the more problematic issue of the MVT, for two reasons. First, the mechanism of meaning change, Feyerabend maintains, is *theory* change; the MVT therefore includes the very contentious *thesis I*. Second, the proposed meaning change is such that statements in T_1 and T_2 are logically

---

[132] Feyerabend (1962), p. 59. My italics.
[133] Devitt (1979), p. 33.
[134] Ramberg (1989), p. 118.

independent (even in the common domain). Feyerabend seems convinced that incommensurability occurs when "none of the usual logical relations (inclusion, exclusion, overlap) can be said to hold between [two successive theories]"[135].

There are a number of reasons for regarding the logical independence claim as highly problematic. For example, it implies that a successor theory neither supports nor rivals its predecessor. It also implies that T_1 and T_2 are not about the same things, raising the question of how they could have a common domain. Such problems have been used as reasons for rejecting the logical independence claim. But I argue that the logical independence claim ought to be rejected on the grounds that Feyerabend opposes the consistency condition, and because the MVT is, according to Feyerabend, a special case of inconsistency between theories. The IT claims that the derivability, consistency and meaning invariance conditions are false; any additional claims made by the IT must not conflict with those foundational assertions.

That there can be meaning change in theory transition is a logical consequence of opposing the Received View's meaning invariance condition; but this consequence alone is not the MVT. A fuller statement of the MVT would include the following:
(1) *thesis I:* "the interpretation of an observation language is determined by the theories which we use to explain what we observe, and it changes as soon as those theories change."[136]
(2) meaning change affects all or some of terms of the predecessor theory
(3) meaning postulates of T_1 cannot be true of the same things of which meaning postulates in T_2 would be true (and to assume they are true of the same thing gives rise to inconsistency)
(4) in the common domain, statements of T_1 can be logically independent of those of T_2 (so the truth of T_1 has no logical consequences for the truth value of T_2.)
Taken together, these claims amount to the main *semantic* claims of the *IT*. The IT itself, as *Against Method* shows, is an even more general claim concerning the history of science, scientific method, rationality, anthropology, and a wealth of other issues. These topics each present their own sets of problems for the IT; but the MVT is the set of claims which compose the *semantic* IT and which *together* undermine

[135] Feyerabend (1975), p. 223.
[136] Feyerabend (1958), p. 31. I have removed the italics.

the semantic IT. It is this problem – that of the MVT, or the semantic IT if you like - which the remaining chapters will address. Chapter 2 will look at what motivates the general notion of the fluidity of meaning and the particular claims made in (1) and (2). Chapter 3 will address arguments from causal theories of reference that attempt to show that (1), the strong form of (2), and (4) are false. Chapter 4 looks at a further argument against (4).

Chapter 2

The Meaning Variance of (Observation) Terms

If you are distressed by anything external, the pain is not due to the thing itself but to your own estimate of it; and this you have the power to revoke at any moment.

Marcus Aurelius, *Meditations*, p. 131.

A new theory of pains will not change the pains; nor will it change the causal connection between the occurrence of pains and the production of 'I am in pain', except perhaps very slightly. It will change the meaning of 'I am in pain'.

Paul Feyerabend, 'Materialism and the Mind-Body Problem' (1963a), in PP1, p. 169.

Introduction

In the first chapter, it was shown that two successive general theories need not meet the derivability condition, the consistency condition, or the meaning invariance condition. These arguments showed:

(a) If it is assumed that T_1 and T_2 have a substantial common ontology, then statements of T_1 and T_2 may be inconsistent in the common domain.

According to Feyerabend's presentation, failing to meet the three conditions is necessary, but not sufficient, for T_1 and T_2 to be incommensurable. The semantic incommensurability thesis (IT) includes a further proposal about the holistic nature of the meaning change which takes place, namely:

(b) *thesis I: "the interpretation of an observation language is determined by the theories which we use to explain what we observe, and it changes as soon as the theories change."*[1]

The IT also proposes that the terms of T_1 and T_2 have not merely different meanings, but different such that:

(c) statements of T_1 may be logically independent – even in the common domain - to statements of T_2, so that the truth of T_1 may have no logical consequences for the truth of T_2.

I will take the meaning variance thesis (MVT) to be the semantic claims of the IT and to consist of claims (a), (b) and (c). The radical MVT, which is proposed and eventually withdrawn, replaces (c) with (c'):

(c') *All* descriptive statements of T_1 and T_2 may be logically independent in the common domain.

Chapter 1 has already established (a). The case for (b) is mostly presented in Feyerabend's pragmatic theory of observation (PTO), and is considered in this chapter. Feyerabend regards claim (c) as a possible result of (b), so the PTO lies behind both (b) and (c). Claim (c') is returned to at the beginning of Chapter 3.

The PTO is not generally regarded as a credible theory of meaning, so the point of this chapter is not to show that it is. To draw such a conclusion at this stage is not intended to 'take the wind out of our sails', but to bring into focus what are the aims of this chapter. First, there is the descriptive aim of recording Feyerabend's arguments for the PTO and against opposing views. Second, there is the critical aim of stating shortcomings of the PTO and the arguments against competitor theories of

[1] Feyerabend (1958), in PP1, p. 31

meaning. Thirdly, and perhaps most interestingly, there is the evaluative aim. The intention here is to ask whether, and to what extent, the PTO is a Quinean theory of meaning, as Feyerabend claimed it was, and to judge how much support (b) and (c) have.

Feyerabend first presents the PTO in English in his 1958 paper 'An Attempt At A Realistic Interpretation Of Experience'. In that paper he does not use the phrase 'meaning invariance'; instead, he speaks of the 'stability thesis'. The stability thesis is the view:

(d) *"interpretations ... do not depend upon the status of our theoretical knowledge."*[2]

Feyerabend uses the word 'interpretations' in a very particular way, meaning the assertoric content of (observation) sentences (or that which makes an observation sentence a statement). Since the stability thesis rejects the idea that the meaning of an observation term is determined holistically by its embedding theory, and yet the stability thesis maintains that such terms are meaningful, Feyerabend will need to justify (d) by stipulating what the interpretation of observation terms *does* depend on. Feyerabend thinks that proposers of the stability thesis will tend to maintain that the meanings of observation terms are relatively stable because either the meanings of such terms are determined by sense-data, or because the meanings of such terms are explained by the way the terms are used. Interpretation as a function of experience is labelled the 'principle of phenomenological meaning' and interpretation as a function of linguistic convention is called the 'principle of pragmatic meaning'.

In this chapter, Part 1 addresses Feyerabend's arguments against the stability thesis' principle of phenomenological meaning; and Part 2 considers his arguments against the principle of pragmatic meaning. Part 3 presents the PTO and evaluates to what extent it provides a Quinean view of language. Part 4 considers criticisms of the PTO and the degree of support it gives to the semantic claims of the IT, claims (b) and (c).

[2] Feyerabend (1958), in PP1, p. 20.

Part 1: Sense-Datum Theories

The principle of phenomenological meaning states that the meaning of an observation statement is determined by sense-data: "the acceptance (or the rejection) of any description of those things is uniquely determined by the observational situation."[3] So:

> in order to explain to a person what 'red' means, one need only create circumstances in which red is experienced. The things experienced, or 'immediately perceived', in these circumstances completely settle the question concerning the meaning of the word 'red'.[4]

Feyerabend offers three sets of argument against this principle. Set 1 considers the claim that knowledge by acquaintance underlies the ability to denote. Set 2 considers whether the meanings of introspective statements are given solely by sense-data. Set 3 contains three arguments against the principle of phenomenological meaning: an infinite regress, which in turn is part of a *reductio ad absurdum*, and a third argument about phenomena not imparting meaning but being grounds for *selecting* a meaningful statement from a group.

With regard to set 1, one of the objects of knowledge by *acquaintance* is sense-data. In contrast to knowing by acquaintance, we know an individual by *description* "when we know that it is *the* object having some property or properties with which we are acquainted"[5]. The relation, then, between the two types of knowledge (i.e. by acquaintance and by description) is such that "in the case of particulars, knowledge concerning what is known by description is ultimately reducible to knowledge concerning what is known by acquaintance."[6] It was Bertrand Russell's view that the sense-data statement 'there is a canoid patch of colour' was more certain (or less prone to error) than 'there is a dog'. Russell thought that, with his canoid patches, it was possible to 'break down' the denotation of a statement about a dog into constituent sensations[7] with which we could be unproblematically acquainted. About this view, Feyerabend remarks:[8]

> Russell seems to assume that the statement 'there is a canoid patch of colour', while being true whenever 'there is a dog' is true, satisfies the condition of being logically simpler than 'there is a dog' *because it is about a simpler phenomenon*

---

[3] Feyerabend (1958), in PP1, p. 25.

[4] Feyerabend (1965a), pp. 203-4.

[5] Russell (1917), p. 166.

[6] Russell (1917), p. 158.

[7] "[I]n order to understand such propositions, we need acquaintance with the constants of the description, but do not need acquaintance with its denotation." Russell (1917), p. 166.

[8] How accurately Feyerabend's remark represents Russell's views will not be of much concern here. It is the position, rather than the accuracy of its attribution to Russell which is the main interest.

(a patch of colour is two-dimensional, does not bark, a dog is three-dimensional, barks etc).[9]

Feyerabend objects to the appeal to sense data on three grounds. First, to state 'there is a canoid patch of colour' in the presence of a dog is to give a phenomenologically inadequate observation statement, for a canoid patch of colour is not a dog. Second, if there actually is a dog, then the statement 'there is a canoid patch of colour' is false, "for a picture of a dog is not a dog."[10] Third, the allegedly simpler statement contains the category 'canoid' which involves the category 'dog': 'dog-like' is hardly more simple than 'dog'. Feyerabend concludes that the attempt, on the basis of knowledge by acquaintance, to build up an observation language whose interpretation is immediately given by sense-data is unsuccessful.

Here, Feyerabend's criticisms of the sense-data view do not seem to me to succeed in their goal of having the stability thesis dismissed. Feyerabend's criticisms are in danger of being criticisms only of statements about individuals known by description, whereas Russell's comments distinguish between the physical objects known by description and sense-data known by acquaintance.[11] Of course, collapsing Russell's distinction is Feyerabend's intention. But Feyerabend's simply not following this distinction does not produce convincing arguments against Russell's making it. Feyerabend's failure to engage with Russell's distinction seems most marked in the third criticism. I can sense a canoid patch of colour without there being a dog present (such as when a wolf or a jackal is present); but when I look at a dog, I will always sense a canoid patch of colour. So the sense-data statement 'There is a canoid patch of colour' *is* logically weaker than the descriptive statement 'There is a dog'. Inasmuch as Feyerabend's criticisms are interesting, they are so because there is some *disparity* between my experiencing a dog and my experiencing a canoid patch of colour. But Feyerabend does not enlighten us as to the nature of that disparity. The next set of arguments involves an area where, it is claimed, there is no such disparity.

---

[9] Feyerabend (1958), in PP1, p. 28.

[10] Feyerabend (1958), in PP1, p. 28. Does Feyerabend here assume that a patch of canoid colour is a *picture* of a dog?

[11] Whether physical objects are a function of sense-data or vice-versa is an issue of which Russell is keenly aware: "In Physics as commonly set forth, sense-data appear as functions of physical objects: when such-and-such waves impinge upon the eye, we see such-and-such colours, and so on. But the waves are in fact inferred from the colours, not vice-versa [...] Thus if Physics is to be verifiable we are faced with the following problem: Physics exhibits sense-data as functions of physical objects, but verification is only possible if physical objects can be exhibited as functions of sense-data.""Russell (1917), p. 109.

Set 2 arguments against the principle of phenomenological meaning concern psychological phenomena. Proponents of the principle can claim that, when I feel pain, then the statement 'I am in pain' is true and must be so: the truth of the statement (and hence the meaning of 'pain') is determined by the sensation. The meaning of such statements are therefore fixed by the sensation accompanying the urge for me to say that I am in pain.

Feyerabend has three objections in this set concerning psychological phenomena. The first is that the above comments about psychological phenomena claim:

> the existence of either an urge to produce a sentence of a certain kind or the existence of psychological phenomenon, can without further ado transfer meaning upon a sentence.[12]

Feyerabend objects to this claim on the grounds that it is "unacceptable to any philosopher who takes seriously the distinction between facts and conventions."[13] Feyerabend wants to insist that facts are in some sense independent of opinions, while *descriptions* of states of affairs employ linguistic conventions. He points out that "what we are discussing is not what is and is not going on in the world, but how what is going on is to be *described.*"[14] The problem with the principle of phenomenological meaning, claims Feyerabend, is that this distinction ultimately collapses. Feyerabend points out that when I utter 'I am in pain' I must have in mind a conventional notion or description (i.e. a 'theory') of what pain is; for example: pain is not something inanimate objects have, it is not contagious, and it can disappear through the use of morphine. Other people share this description of what pain is. However, the principle of phenomenological meaning would have it that only the *immediately given sensation* of pain determines the meaning of the sentence 'I am in pain'; the above *descriptions* associated with pain would then have no role in determining the meaning of the pain statement. The result of meaning determined solely by sense-data is that, in the case of my saying 'I am in pain',

> I may utter it on the occasion of pain (in the normal sense) [i.e. as defined in the description of pain]; I may also utter it in a dream with no pain present ... or I may have been taught ... to utter it when I have pleasant feelings and therefore utter it on these occasions. Clearly, all these usages are legitimate, and all of them describe the 'immediately given pain'.[15]

---

[12] Feyerabend (1962), pp. 38-9.

[13] Feyerabend (1962), p. 39. This may seem an odd tactic to employ, given Feyerabend's adherence to the claim that theories determine the meanings of statements of facts). But this claim mentions theories and *statements* of fact, not theories and facts. This distinction comes out in the presentation of Feyerabend's argument.

[14] Feyerabend (1965a), p. 196.

[15] Feyerabend (1965a), pp. 195 – 6.

Without inferring a description of pain (from other people's behaviour) I would use the word 'pain' irregularly. The claim that *only* the facts or sense-data of psychological phenomena impart interpretations to observation sentences disregards the conventional nature of linguistic meaning.

In response to Feyerabend's first argument in set 2, it may be retorted that the distinction between fact and convention *need* not come under threat of collapse because of the principle of phenomenological meaning. The causal theory of reference (CTR), for example, adheres to something like the principle of phenomenological meaning without collapsing the distinction between fact and convention. The CTR will be considered in Chapter 3.

Feyerabend's second objection in set 2 is that "it is not true that *every* assertion about sensations excludes doubt."[16] He cites Berkeley's example of a hot sensation becoming a pain sensation, with a 'grey area' as to when the heat becomes pain (or 'heat' becomes 'pain'). He also describes a medical nerve-test on a patient who has had temporary paralysis: the patient may be unsure whether he has had the sensation of a blunt or a sharp instrument pressed against his skin. It is therefore doubtful that a stimulation of nerve-endings renders the interpretation of the statement 'I am in pain' fixed.

In later writing on this topic of psychological states or sensations, Feyerabend shifts the emphasis of the above criticism from being *uncertain* about the sensation to *misinterpreting* the sensation. Feyerabend debated this issue with Herbert Feigl and gives an example from his own personal experience of how a sensation such as pain may be misinterpreted:

> Feigl believed in incorrigible statements. He said ... that being in pain he knew directly and with certainty that he was in pain. I didn't believe him but only had general objections to offer. One night, however, I dreamed that I had a rather pleasant sensation in my right leg. The sensation increased in intensity, and I began to wake up. It grew even more intense. I woke up more fully and discovered that it had been a severe pain all the time. *The sensation itself told me* that it had been a sensation of immense pain, which I had mistaken for a sensation of pleasure.[17]

[16] Feyerabend (1960c), in PP3, p. 24.
[17] Feyerabend (1995), pp. 116 – 7.

Here, Feyerabend seems to take the phenomenon of pain as a 'given', a phenomenon the existence of which is certain, but the meaning or significance of which may be variously construed and misconstrued. As we will see, this shift in emphasis from uncertainty to misinterpretation will answer some problems raised by Feyerabend's second area of objection to the principle of phenomenological meaning.

In response to this second argument of set 2, Elie Zahar retorts that the issue here is not one of uncertainty, but one of inadequate description: we simply do not have the vocabulary to describe the miriad sensations on the pleasure-pain kline:

> Our sensations form a potentially infinite set and have infinitely many nuances; it is no wonder that they cannot all be captured by a finite number of adjectives.[18]

Even though Zahar's point is a fair challenge to Feyerabend's argument from uncertainty, it does not offer any resistance to Feyerabend's argument that observational terms' meanings are *variant* because we can misinterpret (or re-interpret) sense-data statements. What might be considered unsatisfactory, however, are Feyerabend's examples of such misinterpretation; for they all concern states (of mind or body) which are either not fully conscious (sleep) or pathological (a patient with poorly functioning nerve endings). While his argument would be strengthened if it did not depend on such exotic examples, we cannot dismiss his argument simply on the grounds of their rum character.

Feyerabend's third argument in set 2 comes from the conclusions of the other two arguments in the set. That is, if the meaning of the term 'pain' *were* determined by the private sensation of pain, and 'pain' *were* used to make an assertion without regard to any conventional use or description of pain, then the term 'pain' would mean different things at different times to different people. Consequently, I would not know for sure what you mean when you claim to be in pain. This consequence of supporting a sense-datum theory of meaning with respect to introspective statements is what Elie Zahar calls the 'intransmissibility thesis'. The intransmissibility thesis is "the ... view that meaning cannot be infallibly communicated from one person to another"[19]. Lack of a common description of a sensation (such as pain) would be sufficient for what the intransmissibility thesis proposes. When lack of any common description is combined with the claim of

[18] Zahar (1982), p. 399.

[19] Zahar (1982), p. 401.

Feyerabend's second argument, that the same stimulation may cause me to assert at one time that I am in pain, and at another time that I experience pleasure, then the result is not merely intransmissibility between persons: an *individual observer* would be unable to *re*identify the same stimulation with the appropriate interpretation. Sense-data are not generally regarded as *interpreted*; but Feyerabend's point is that sense-data *statements* are interpreted, and that statements of facts and sense-data have no fixed interpretation provided by facts or sense-data.

Zahar is in agreement with the substance of the third argument: "Feyerabend successfully shows that the intransmissibility thesis applies to all phenomenological concepts."[20] Zahar also admits that, in raising the problems of intransmissibility and reidentification, Feyerabend has laid bare 'the conjectural assumption of meaning invariance' by showing that

> a proposition S about sense-data is incorrigible for at most one person, namely the observation reporter Q, and only at the instant t at which Q utters S. Another person Q*, who might want to rely on the absolute truth of S, can never be sure that he and Q attach identical meanings to the descriptive symbols occurring in S.[21]

Feyerabend's argument showing intransmissibility is sound, according to Zahar. But Zahar accepts the intransmissibility thesis, and does not regard it as a reason for rejecting sense data-ism. Zahar's holding the intransmissibility thesis (including his claim, to be quoted shortly, that intuition adequately explains the relation between a statement and a phenomenon) surely suggests that the meaning of an observation statement is or may be unstable in case that the meaning is determined by phenomena. And this is what Feyerabend is proposing in opposing the stability thesis.

Moving on to set 3, the final set of arguments against the principle of phenomenological meaning, Feyerabend's begins with an infinite regress claim. Let:

S = an observation statement

P = a phenomenon which is said to determine the meaning of S

M = the relation between P and S

---

[20] Zahar (1982), p. 401.

[21] Zahar (1982), p. 400.

The view that "phenomena must speak for themselves"[22] would have it that the acceptance or rejection of S is determined by P. Yet the reason why S is accepted is not *simply* because of P, it is because of M: we accept S because there is a mapping, M, of P to S and we reject S when there is no such mapping. For example, if S is the statement 'I am in pain', then the following series of steps would (under the principle of phenomenological meaning) be involved in our determining the meaning of S:

(1) I form or utter S because I know *that* the phenomenon P is mapped to the statement S.

(2) To know *that* P is mapped to S, I must know the mapping, M.

(3) To know the mapping M, I must have a thought in the form of a statement S'.

(4) To form S', I must know *that* M is mapped to S'.

(5) To know *that* M is mapped to S', I must know the mapping M'.

(6) To know the mapping M', I must have a thought in the form of a statement S"

(7) To form S", I must know *that* M' is mapped to S"

(8) To know *that* M' is mapped to S", I must know the mapping M".

etc. (Repeat ad infinitum from (3) to (5) or (6) to (8) with appropriate substitutions.)

The corollary of Feyerabend's infinite regress argument against the principle of phenomenological meaning is that an "observer must perform infinitely many acts of observation before he can determine the meaning of a single observation statement."[23] In short, phenomena (in this example, psychological phenomena) cannot determine the meanings of observation statements.

Feyerabend's infinite regress argument is criticised by Zahar on the grounds that it

> clearly contains a gratuitous assumption, namely that [M], in order to exist at all, must be adequately described by some S'. It is as if intuition were impossible without some accompanying linguistic, or quasi-linguistic, entity.[24]

But Feyerabend's argument is surely not about whether M exists, but what it is to know that M; for to know that M would be to know that the meaning of S is determined by P (according to the principle of phenomenological meaning). As John Preston points out, "[t]he question is whether we can know that S 'fits' P without having any means to express this knowledge."[25] Zahar's suggestion that M is intuited runs into a series of objections:

---

[22] Feyerabend (1965a), p. 204.

[23] Feyerabend (1965a), pp. 204 – 5. See also Feyerabend (1958), p. 26.

[24] Zahar (1982), p. 398.

[25] Preston (1997), p. 35.

> It is hard to see how a meaningful sentence expressing a relationship could have its meaning completely determined by a relationship which cannot be adequately described, and is inexpressible. Such a sentence would be one whose meaning was humanly unstable: its meaning could therefore be conveyed only in an appropriate situation in which the audience was confronted with the relationship. But how, in such a situation, would the communicator direct the audience's attention to, or know whether the audience had focused on, the correct relationship?[26]

Feyerabend uses this infinite regress in a *reductio ad absurdum*[27], thereby constituting the main argument of set 3. Since the principle of phenomenological meaning has it that phenomena are capable of determining statements, then M, if it is a phenomenon itself, must be capable of determining a statement S'. Since M is never expressed by a statement S' (because of the regression) then M "cannot be immediately given in the sense in which P is immediately given, i.e. it cannot be a phenomenon."[28] In the case where M is not a phenomenon, the principle of phenomenological meaning would regard M as of no significance in the formation of observation statements; since no significance is attached to the relation between P and S, the principle falsifies itself! If M *were* a phenomenon, then the *principle* of the principle of phenomenological meaning is that M should determine the meaning of a statement S'; yet M never meets this requirement, so the principle is false.

Rather than dealing directly with the relation between a statement and a phenomenon, Richard T. Hull suggests that Feyerabend's traps be avoided by considering the relation between a state of affairs and a thought about the state of affairs. Hull points out that statements obtain their connexion to facts by way of conventional laws (in the language community):

> I utter something to my doctor and he slaps me on the back and tells me it's great I'm feeling so well. This is good evidence that I have not expressed the thought that-I-am-in-pain.[29]

So facts do not *directly* determine statements, rather facts determine thoughts and they can do so in a way that does not lead to an infinite regress. Let:

H = the fact that I am in pain (it so happens that I am in pain)

T = the thought that-I-am-in-pain

M* = the relation between H and T

---

[26] Preston (1997), p. 35. Preston adds that Zahar's assertion of ineffability concerning the statement/phenomenon relation "injects mysticism into the foundations of empirical science." This would be an unacceptable consequence for many holders of the principle of phenomenological meaning.

[27] Feyerabend regards the *reductio* as 'the argument' and the infinite regress merely as a part of the *reductio*. See Feyerabend (1958), in PP1, p. 25.

[28] Feyerabend (1958), in PP1, p. 25.

Hull argues that M* is an analytic relation:

> for I already know that the thought that-I-am-in-pain means that I am in pain before a particular situation of my being in pain arises. That the thought that-I-am-in-pain could fail to intend the fact that I am in pain is either false or unintelligible.[30]

This analytic relation between H and T is due to intentionality: if I am thinking that-I-am-in-pain then that in itself is sufficient for my being in pain, without any further appeal to facts: "the fact need not obtain in order for the intentional nexus to connect the thought with the fact"[31] If I think that-I am-in-pain, then that can mean only one thing – my thought is necessarily about the *fact* that I am in pain. No infinite regress is called upon to express the relation between the thought and the fact; so the meaning of the statement 'I am in pain' is fixed by the state of affairs indirectly. Hull makes an interesting case where facts and thoughts match nicely. However, Hull's explanation is not in accord with the principle of phenomenological meaning. One of his premises is that the meaning of 'I am in pain' is sufficiently constrained by the behaviour of the members of the language community, and this is inconsistent with the principle of phenomenological meaning. Furthermore, the analytic relation which Hull proposes has synthetic import. Synthetic *a priori* statements would not be acceptable to Positivists, and Feyerabend is primarily concerned with attacking Positivist supporters of the stability thesis. Such supporters ought not to go along with Hull's reply to Feyerabend's criticism. And Feyerabend would point out that no empiricist should accept synthetic statements which are in principle unfalsifiable.

Feyerabend's final argument in set 3 considers the weaker claim of the principle of phenomenological meaning that, "given a class of *interpreted* sentences, the relation of phenomenological adequacy might allow us to select those sentences that correctly describe P."[32] The view under consideration is that, from a class of *statements*, we would be able to know, simply by acquaintance with a phenomenon, which statement to use as our observation statement of that phenomenon. Many statements can be "obtained from"[33] the single statement 'there is a table', such as 'there seems to be a table', 'a table is located in the place I am indicating', 'a table exists before us', and so on. A spectral statement is each member of the class of

---

[29] Hull (1972), p. 383.
[30] Hull (1972), p. 381.
[31] Hull (1972), p. 381.
[32] Feyerabend (1965a), p. 205.
[33] Feyerabend (1965a), p. 206.

'statements which are [said to be] phenomenologically adequate' to the selected and interpreted sentence 'there is a table'; and the class they make up is called, appropriately, "the *spectrum* associated with the phenomenon in question."[34] From this spectrum we cannot, *merely* by acquaintance with a phenomenon, select our observation statement. So Feyerabend concludes "that phenomena alone cannot even select interpretations, but that additional considerations are needed."[35]

I confess that I find this argument of Feyerabend's the most opaque of all those so far considered. My first difficulty with it is that I do not see the need to have a preference for one of the statements if all of the statements are phenomenologically adequate. If, as Feyerabend seems to have allowed for the purposes of his argument, the phenomenon has determined the meaning of all of the spectral statements, then surely any one of them will do as a satisfactory observation statement. Feyerabend would probably reply that, since there are many different acceptable statements, then the phenomenon does not provide invariant meaning (so the principle of phenomenological meaning is untenable, QED). But my problem with this reply is that many different statements need not entail many different meanings. This brings out what I see as the second difficulty with Feyerabend's argument, the spectrum of statements.

Feyerabend proposes the spectrum of statements because, "given a phenomenon, there are always many different statements (expressed by the same sentence) that will be found to fit the phenomenon."[36] As the foundational sentence, Feyerabend takes the example "there is a table". This sentence is said, then, to *express* the *statements* "there is a table", "there seems to be a table" and many other statements. An unhappy feature of Feyerabend's argument is that the spectrum which Feyerabend proposes could imply that spectral statements *do* have a common element of meaning and that this common element is the denotation of the spectrum. As the chapter on reference will show, just such arguments *are* used to support the stability thesis. However, Feyerabend would probably not be fazed, for he would maintain that any given spectrum has infinitely many possible statements, none of which is selected as 'the meaning' or privileged interpretation or best description by the phenomenon in question. To make his argument convincing,

[34] Feyerabend (1965a), p. 206.
[35] Feyerabend (1965a), p. 206.

though, I think Feyerabend would need to improve upon and better explain the notion of the spectrum which is so central to his argument.

With the three sets of arguments considered here, Feyerabend concludes that "the meaning of an observation term and the phenomenon leading to its application are two entirely different things."[37] The first set of arguments (about knowledge by acquaintance) against sense-datum theories of meaning was judged not convincing. The arguments in set 2 have shown that sense-data are not *sufficient* to determine the meaning of observation sentences, (or that, if sense-data are, observational meanings will be variant); so set 2, while generally successful in dismissing the principle of phenomenological meaning, does not adequately support Feyerabend's conclusion (stated at the beginning of this paragraph) that the principle of phenomenological meaning does not in part determine the meanings of observation statements. The *reductio* argument in set 3 is surely Feyerabend's most effective support for ruling out sense-data as the source of invariant meaning for observation statements. But that sense-data do offer *some* constraint on the meaning of observation statements is an option against which Feyerabend has not offered compelling evidence in these arguments.

Part 2: Use Theories

The second source of support for the stability thesis is the principle of pragmatic meaning, the principle that "the interpretation of an expression is determined by its 'use' "[38]. Feyerabend gives a formal description of the use of an observation sentence and he calls this formal description the 'characteristic' of an observation language: "The characteristic of an observation language completely determines the 'use' of each of its atomic sentences."[39] He defines the characteristic as follows:[40]

> C = a class of observers using a language
> A = class of atomic sentences (each a physical event) of the language considered
> S = class of observed situations
> F = the function correlating members of A with C (the associating function)
> R = the function correlating S with acceptance or rejection of any member of A (the causal relevance function)

[36] Feyerabend (1965a), p. 206.
[37] Feyerabend (1965a), p. 206.
[38] Feyerabend (1958), in PP1, p. 21.
[39] Feyerabend (1958), in PP1, p. 18.
[40] Feyerabend (1958), in PP1, p. 18.

The set {C, A, S, F, R} fully characterises "[t]he pragmatic properties of a given observation language"[41]. Feyerabend uses the characteristic to substantiate his claim that the act, and even conditions, of sentence utterance do not make "any stipulation as to what those sentences are supposed to assert"[42]. To understand the sentences as statements, further conditions must be added, namely an interpretation. Feyerabend illustrates this claim with the following example.

A language, L, describes the colours of self-luminescent objects using predicates P_i (i = 1, 2, 3 ...). The 'characteristic' of L is defined and determined for everyday situations (moderate velocities and masses) and the predicates of L are regarded as designating properties of objects, irrespective of whether the objects are being observed. A new theory is formulated which states that the wavelength of light is dependent upon the relative velocities of the observer and the light source. Using this theory and L to describe the colour of a self-luminescent object *a*, the expression '$P_i(a)$' has now become "no longer complete and unambiguous"[43], for in addition its former meaning, we may also interpret '$P_i(a)$' as '$P_i(a,p)$', where *p* represents "the relative velocity of *a* and the co-ordinate system of the observer – which may or may not be observable"[44]. The 'characteristic' of L restricts the use of '$P_i(a)$' to the everyday level, and we can continue to use it without any formal alteration on the everyday level. While the characteristic of L may remain unchanged with regard to the use of '$P_i(a)$', '$P_i(a)$' has a different meaning to what it had before the positing of the Doppler effect, for now it designates the relation '$P_i(a,p)$'. In this formal way, Feyerabend is saying that no change has occurred in the observation *sentence's* place in L, but there is a different observation *statement*. The use of the language has not altered, nor have the relevant phenomena, but the interpretation has.

From this example, Feyerabend hopes to have shown that the principle of pragmatic meaning is untenable. Commenting on the colour predicate example, John Preston argues that no serious challenge has been mounted against the principle of pragmatic meaning, for:

---

[41] Feyerabend (1958), in PP1, p. 18.
[42] Feyerabend (1958), in PP1, p.18.
[43] Feyerabend (1958), in PP1, p. 32.
[44] Feyerabend (1958), in PP1, p. 30.

(f) the example fails to provide "support for the idea that a change in meaning can occur in the absence of any change in application."[45]

If there has been no change in the characteristic, no change in the use of the sentences of L, then "what ... could make us think that the meaning ('interpretation') of that sentence has changed?"[46] Indeed, without any change in the use of the observation sentence '$P_i(a)$', argues Preston, a speaker of L would explain '$P_i(a)$' in the same old non-Doppler effect terms. In which case, if there were a change in meaning "it must be one which transcends the speaker's *knowledge* of meaning".[47] If a change of theory *is* to effect a change of meaning in an observation statement, then there will need to be some change in the use of that or related observation statements.

Since it would be possible to narrow a domain to the extent where the *displayed* linguistic behaviour of two speakers (one a Doppler fan, the other with no such specialist knowledge) was identical, Preston's objection needs to include as regularities of use, the case where sentences are *privately* accepted (i.e. "uttered assertively to oneself"[48]). Indeed, Preston allows for such a notion of use when he speaks of "changes in the *(possible, if not actual)* correct use"[49] of A; for the very sentences of the new theory would, though they were never *publicly* uttered, constitute private changes to the application of A. Then, it is claimed, Feyerabend's example does not show that use (given in the characteristic) is held constant while meaning has changed; consequently (f) is vindicated.

Such a response, however, seems to suggest that a sentence will have as many different meanings are there are ideolects, for then a (private or possible) change of use would *always* entail a change of meaning. The principle of pragmatic meaning would then be required to defend itself by differentiating "between those properties which comprise [...] 'the use' and those which do not"[50], and this is not a task which Preston undertakes. Paul Horwich, a supporter of a use theory of meaning, makes such a distinction on the following basis:

---

[45] Preston (1997), p. 37.
[46] Preston (1997), p. 37.
[47] Preston (1997), p. 37.
[48] Horwich (1998), p. 94.
[49] Preston (1997), p. 38. My italics.
[50] Horwich (1998), p. 60.

> the way to pick out the particular use property of a word that comprises what we call 'the use' is to find the use property which provides the best explanation of all the others.[51]

For example, the discovery of a tenth planet would lead to new uses of 'planets', such as 'There are ten planets', yet this is not considered sufficient to change the meaning of 'planets'. If the meaning of 'planets' is altered, it will be done so in such uses as 'Planets do not orbit stars', for such uses alter a more explanatorily basic meaning of 'planets'. Of such a procedure for ascertaining meaning-determining use, Horwich admits:

> The outcome [...] may no doubt be indeterminate. There will sometimes be equally good ways of finding a simple regularity in the use of a word [...] But a distinction with unclear boundaries is a distinction none the less – one that puts us in a position to say of certain novel deployments of a word that they definitely do not amount to changes in its use.[52]

Horwich also admits that the procedure may rely on distinguishing analytic from synthetic statements. So it would seem that vindicating (f) is a project which is beyond the scope of this chapter. As regards Feyerabend's criticism of the principle of pragmatic meaning, it can at least be said that the issues it raises are legitimate.

In the Doppler effect example, Feyerabend had attempted to show that meaning can differ when use is identical. He also tried to illustrate the converse: two observation languages with different characteristics can be "jointly interpreted by one and the same theory"[53]. For example, "Maxwell's electrodynamics plays this role with respect to the phenomena of light and electricity."[54] Sentences which had been used in the description of light, and sentences used in the description of electricity, endured no alteration of use in their respective domains by the introduction of Maxwell's theory; yet before Maxwell, no light-sentence had the same interpretation as an electricity-sentence.

Finally, Feyerabend objects to the principle of pragmatic meaning on the grounds that that there are phonological regularities which are meaningless: "the fact that in certain situations [...one] (consistently) produces a certain noise, does not allow us to infer what this noise means."[55] Snoring would be an example of such behaviour.

---

[51] Horwich (1998), p. 60.
[52] Horwich (1998), p. 60.
[53] Feyerabend (1958), in PP1, p. 32.
[54] Feyerabend (1958), in PP1, p. 32.
[55] Feyerabend (1958), in PP1, p. 22.

To such a criticism, Horwich gives two replies. First, "the use of a [...] term must cohere with the regularities that constitute the meanings of the other words."[56] Snoring expressions clearly do not meet this condition. Second, it is important to distinguish the making of a sound[57] from *accepting* a sentence. The propositional attitude of a snorer, or of one who listens to snoring, is not that of holding as true or false the sound made. The mere regular production of sounds is not sufficient for those sounds to be used to mean something.

What Feyerabend's arguments against the principle of pragmatic meaning have shown is that, while the meanings of observation terms may be *influenced* by their use, "the logic of the observational terms is not *exhausted* by the procedures which are connected with their application 'on the basis of observation.'"[58]

Regarding Feyerabend's arguments against the sense-data and use theories of meaning, John Preston concludes:

> Observation terms do have (relatively) stable meanings, not because their meaning is fixed by invariant phenomenological features, but because it is fixed by their *use*, which is relatively impermeable to theoretical considerations.[59]

While there may be a use theory of meaning which adequately supports this conclusion, Feyerabend has shown that *naive* sense-data or use theories of meaning will not do.

Feyerabend describes as 'semantic theories of observation' the views expressed in the principle of phenomenological meaning, and principle of pragmatic meaning. These views account for observation statements in terms of their meaning, claiming that observation statements derive their meaning from either their use or from sense data. The PTO, by contrast, attempts to show that "[o]bservational statements are distinguished from other statements not by their meaning, but by the circumstances of their production."[60]

---

[56] Horwich (1998), p. 94.

[57] These responses to Feyerabend's criticism are good for marks as well as sounds.

[58] Feyerabend (1963b), p. 16. My italics.

[59] Preston (1997), p. 101.

[60] Feyerabend (1965a), p. 212.

Part 3: The PTO And Quinean Considerations

Feyerabend described himself as "a Quinean"[61] because he thought that the PTO implied the indeterminacy of translation thesis:

> I have now discovered that I said everything Quine is famous for, such as radical translation, much more briefly and with much better arguments in 1958, in my Aristotelian Society paper.[62]

John Preston dismisses such claims, maintaining that Feyerabend "failed to understand Quine."[63] The task of Part 3 is to present the PTO and weigh Feyerabend's claim against Preston's dismissal.

According to the PTO, observation sentences are uttered by a conditioned observer. Such an observer, O, has the ability to observe a situation, *s*, if O "is able to distinguish between *s* and other situations."[64] O is able to demonstrate this ability when "[O] can be conditioned such that [O...] produces a specific reaction *r* whenever *s* is present, and does not produce *r* when *s* is absent."[65] When the conditioned observer is an average human being, *r* can be an utterance. Feyerabend's notion of an observation sentence, then, is that speakers of the same language are trained to respond (or be disposed to respond) consistently to a physical stimulus by uttering a particular observation sentence. This appears to be what Quine thinks too: "an occasion sentence is observational to the extent that all speakers assent to it in response to the same stimulations."[66]

The process of conditioning the human observer presupposes that he has some minimal level of knowledge (or theory).[67] Feyerabend acknowledges as much when he states that "behavior that is not connected with any theoretical element [...] is impossible."[68] Indeed the human observer is born with some minimal theory, for "[k]nowledge can *enter* our brain without touching our senses. And some knowledge *resides* in the individual brain without ever having entered it."[69] Such theory is so

---

[61] Feyerabend (1999), p. 237, referring to the (1958) paper in which he presented the PTO.

[62] Feyerabend (1999), p. 362.

[63] Preston (1997), p. 220, n. 20.

[64] Feyerabend (1958), in PP1, p. 19.

[65] Feyerabend (1958), in PP1, p. 19.

[66] Hookway (1988), p. 132.

[67] Townsend (1971), p. 207 thinks that this presents a problem for the PTO. But Feyerabend does not deny that some minimal theory is involved if an observer is to be conditioned; he denies that such minimal theory adequately constrains the meanings of the observation sentences.

[68] Feyerabend (1962), p. 95.

[69] Feyerabend (1969a), in PP1, p. 134. I take it that this is not one of Feyerabend's Quinean views.

minimal as to be a negligible constraint on the meaning of any conditioned response, *r*.

Since an observation sentence is a conditioned response, tokens of its production are a matter of conformity with convention and of physically stimulated causal regularity; but since the beliefs of the observer are bound to determine what he means when he utters the observation sentence, the taught convention and the stimulation (or circumstances of utterance) do not sufficiently constrain the meaning of the observation sentence. That is the point of Feyerabend's contentious analogy between a human observer and an instrument:

> [H]owever well behaved and useful a physical instrument may be, the fact that in certain situations it consistently reacts in a well-defined way does not allow us to infer (logically) what those reactions mean.[70]

Having made the point and the analogy, Feyerabend admits that the analogy with the instrument is limited (and presumably the point is too); for what makes an instrument different to the human observer is that he "also interpret[s] the indications of these instruments [...] or the observational sentence uttered"[71]. Interpretation is an "additional act"[72], but, for a normal human being, *is inseparable from the action of uttering a sentence.* Feyerabend makes this point in a number of ways. One way is in his description of a *language* as a characteristic *plus* an interpretation (given in Part 2). Observation sentences, when spoken by normal human beings, are spoken as sentences of a language, and *ex hypothesi* they have an interpretation (i.e. they are statements, in Feyerabend's terminology). Feyerabend also points out that the average human being has a level of theory well beyond the minimum required to become a conditioned observer:

> [E]liminate part of the theoretical knowledge of a sensing subject and you have a person who is completely disoriented, incapable of carrying out the simplest action. Eliminate further knowledge and his sensory world (his 'observation language') will start disintegrating ; even colours and other simple sensations will disappear until he is in a stage even more primitive than a small child.[73]

So the PTO is *not* proposing that a human (i.e. language-speaking) observer is merely a maker of noises. Such an observer, if he is a well-functioning human being, has an extensive knowledge (not necessarily current, scientific, theoretical

---

[70] Feyerabend (1958), in PP1, p. 22.
[71] Feyerabend (1958), in PP1, p. 19.
[72] Feyerabend (1958), in PP1, p. 19.
[73] Feyerabend (1969a), in PP1, p. 133.

knowledge), so that his observation utterances will be meaningful statements - expressions of that knowledge. Quine shares this view.[74]

The distinction which Feyerabend makes between observation sentences and statements is not odd, but his choice of terminology is. (It is also odd that speakers of the same language are assumed to have the same beliefs.) Observers who speak the same language are conditioned to give a particular response to a particular stimulus. The wider theoretical knowledge (or beliefs) of the observer are not *necessary* for the making of such responses (the conditioned observer *need not* be a natural language speaker), and in this respect the responses are (rather misleadingly) termed observation *sentences*. Very little meaning content or informative output is necessary for *r* to be such a sentence; all *r* necessarily conveys is that the conditioned observer has had a particular stimulation. The observer's understanding of the stimulation is (again misleadingly) what constitutes his 'experience', according to Feyerabend. Stimulation is to sentence what experience is to statement. An experience is what an observer *describes* (i.e. interprets) the stimulation as, and this can vary from observer to observer; *that* a stimulation has occurred is indicated by the production of an observation sentence, and this response *cannot* vary from observer to observer when they have undergone the same conditioning. While Feyerabend's *terminology* aids confusion, the *distinction* between experience and stimulation is Quinean. Quine held that *stimulation* is "the uninterpreted impact of external things upon our cognitive apparatus."[75] For Quine the *experience* (in Feyerabend's sense) would be something like the observer's beliefs about the stimulation. A Quinean distinction akin to that of Feyerabend's sentence and statement is not so easy to discern, partly because Quine's observers are all healthy, human, language speakers. They do not, therefore, merely utter sentences (in Feyerabend's sense). But Quine and Feyerabend would both agree that *all* healthy human language speakers are more than conditioned noise-makers:

(g) For all such observers, their observation sentences (in Feyerabend's sense) *express* observation statements.

The motivation for Feyerabend's distinction between sentence and statement is to point out that observation sentences can be stimulus synonymous, but have different

[74] As his final sentence in 'Two Dogmas of Empiricism' makes clear: "Each man is given a scientific heritage plus a continuing barrage of sensory stimulation". 'Two Dogmas of Empiricism' at http://my.dreamwiz.com/reality/data/philosophy_information2_quine.htm

[75] Hookway (1988), p. 191.

interpretations. And Quine would agree with that point, for it is essential to the indeterminacy of translation thesis.

Feyerabend summarises Quine's views on radical translation thus:

> The argument seems to consist of two parts: (1) given some body of evidence you can always have many different theories which fit the evidence, and (2) solipsism in its linguistic form i.e. I can never know that when you say 'Jensen is progressive', you do not really mean: 'Popper is a donkey.'[76]

While (1) and (2) are consistent with Quine's views, they are weaker claims than those of the indeterminacy of translation thesis. (1) and (2) look like epistemological claims, when in fact:

> Quine's point is not primarily epistemological: his claim is not that correct translation is *under*determined by available evidence, but rather that it is not *determined* by the facts.[77]

So in (1), 'some body of' should read 'all' and both tokens of 'evidence' should be replaced with 'facts'. A similar point applies to (2). It is not just that I can never *know* the meaning of what you say, but that "there is no objective fact of the matter what we are talking about"[78] (beyond what is expressed by behaviour). The remaining discussion questions whether Feyerabend agrees with the metaphysical claims of the indeterminacy of translation thesis.

In distinguishing between observation sentence and statement, the PTO differentiates

> between the causes of the production of a certain observational sentence, or the features of the process of production, on the one side, and the meaning of the sentence produced in this manner on the other.[79]

Feyerabend gives two reasons for making this distinction:

(h) "First, because the existence of a certain observational ability [...] is compatible with the most diverse interpretations of the things observed;"[80]

and

(i) "secondly, because no set of observations is ever sufficient for us to infer (logically) any one of those interpretations."[81]

---

[76] Feyerabend (1999), p. 339.

[77] Hookway (1988), p. 137. My italics.

[78] Hookway (1988), pp. 142-3.

[79] Feyerabend (1962), p. 94.

[80] Feyerabend (1958), in PP1, p. 22.

[81] Feyerabend (1958), in PP1, p. 22.

These two reasons seem to be related in the following way. The second reason (i), Feyerabend tells us, is the problem of induction. The claim that it is never rational to accept a general hypothesis which has been inductively inferred implies that it is never rational to accept a scientific theory which has been similarly inferred. According to Feyerabend (*thesis I*), a theory determines the interpretations of the theory's observation sentences. It follows, then, that it is not rational to accept that an observation sentence has one, and only one, interpretation (i.e. that there is a one-to-one correspondence between observation sentences and observation statements). Some explanation of (h) has already been given in the preceding paragraphs where the view that that observation sentences can be stimulus synonymous but have different interpretations was presented. What remains to be explained of (h) and (i) is how, from the diverse interpretations, one is chosen.

In *thesis I*, Feyerabend proposes that the meanings of observation sentences are determined by the theory held by the speaker. The theory is chosen according to the criterion of predictive success. A theory is judged according to:

> [T]he way in which the prediction sentences are ordered by it and by the agreement of this *physical* order with the *natural* order of observation sentences as uttered by human observers.[82]

The physical order (i.e. the syntax) of the sentences which a theory predicts should match as closely as possible the syntax[83] of the conditioned observer's observation sentences. Then the theory will indirectly predict "the natural order of sensations."[84] *Why* this is so is clear: a conditioned observer's linguistic response is correlated with a particular stimulation; if a theory can predict the response, it indirectly predicts (a better word might be 'indicates') the stimulation.

The criterion of predictive success strikes me as making Quinean claims, namely:

> Our knowledge of the external world is mediated through 'stimulations' at the surfaces of our perceptual organs, and our framework of sentences is tied down to reality only insofar as it enables us to anticipate these stimulations.[85]

The framework or relation of sentences (not just observation sentences) is given by a theory or conceptual scheme. According to Feyerabend, the right relation of such sentences is provided by a *theory* which would enable one observer to predict, for

---

[82] Feyerabend (1965a), p. 215. More details have been given in Part 3.

[83] Confusingly referred to as the 'natural order of observation sentences' – 'natural' for all of the observers so conditioned.

[84] Feyerabend (1965a), p. 215.

[85] Hookway (1988), p. 216.

any given circumstances, the observation sentences of a *differently* conditioned observer:
such a theory would then enable one observer to speak another observer's *observation* language. For Quine, the right relation of observation sentences is provided not by a theory but by a translation manual. On the grounds of this difference, I will go about arguing that Feyerabend's PTO does not propose Quine's indeterminacy of translation thesis. Next, I describe briefly Quine's two arguments for the indeterminacy of translation thesis.

In the Quinean scheme of things, physical evidence underdetermines the theory of nature. Quine's underdetermination claim has a Feyerabendian resonance:

> Physical theories can be at odds with each other and yet compatible with all possible data even in the broadest sense. In a word they can be logically incompatible and empirically equivalent.[86]

Quine's epistemological holism holds that the truth or falsity of an individual sentence cannot be tested against the physical evidence, but that only a whole theory can. Such holism has semantic consequences when the meaning of a sentence is considered to be "the difference its truth would make to possible experience"[87]. Since such a test for truth can only be applied to theories as wholes, and experience makes no difference between a number of incompatible theories, the meaning of a sentence will be indeterminate, not simply underdetermined. The hop from epistemological underdetermination to semantic indeterminacy comes from linking empirical adequacy with truth. Dagfinn Føllesdal remarks that this argument for indeterminacy of translation "proceeds via holism and a verificationist theory of meaning."[88] In this respect, Quine's argument is similar to Feyerabend's reason (i) (the problem of induction) – the reason for his distinction between observation sentences and statements. Feyerabend's scepticism about induction would suggest that he does not make the jump from empirical adequacy to truth, and the corresponding jump from empirically underdetermined theories to ontologically indeterminate observation statements. However, I think that it is at least arguable that Feyerabend *does* make a leap from epistemological claims about theories to ontological claims. I should add that this leap is not apparent in his PTO, so I am *not* claiming that the PTO presents the indeterminacy of translation thesis. In other

[86] Quine, quoted in Føllesdal (1973), pp. 292-3.
[87] Quine, quoted by Føllesdal (1973), p. 290.
[88] Føllesdal (1973), p. 290.

parts of his writings, Feyerabend seems to combine epistemological holism with ontological claims as follows.

For Feyerabend, observation statements of empirical facts have a peculiar feature: they can be mutually inconsistent, yet empirically adequate in a common domain. For example, we can state that motion is uniform and that motion is nonuniform, and both observation statements are empirically adequate within a common domain; as such, they are both statements of facts in Feyerabend's view. Feyerabend claims that the facts can be "made inaccessible"[89] or even "eliminated"[90] by different theories. Facts, unlike energy, can be created and destroyed by empirically adequate theories.[91] So different general theories express different ontological commitments:

> A comprehensive theory, after all, is supposed to contain also an ontology that determines what exists and thus delimits the domain of possible facts and possible questions.[92]

The semantic consequence is that the observation term 'motion', for example, will mean different things in different theories. If the motion of a body is described as uniform by one theory and nonuniform by another, then the observation term 'motion' has changed its meaning. The meaning of 'motion' is not simply underdetermined, for statements containing it have factual, and so ontological, import. The indeterminacy of the meaning of observation terms comes from Feyerabend's views about *theories*, not from his PTO.

Quine's second argument for the indeterminacy of translation thesis concerns the use of translation manuals. Such manuals are constructed as follows:

> [A]n observation sentence in one language/theory should be correlated with an observation sentence in the other if and only if any stimulation that prompts asset [sic] to the one, prompts assent to the other.[93]

It has previously been made clear that, for both Feyerabend and Quine, observation sentences uttered by human observers are more than noises – they are meaningful and therefore already part of a language. Since Quine holds that "language and theory are inseparable"[94], observation sentences are, upon utterance, part of a

---

[89] Feyerabend (1975), p. 42.

[90] Feyerabend (1975), p. 42.

[91] Feyerabend toys with the options: either "different languages [or theories] will posit different facts under the same physical circumstances in the same physical world, or [...] they will arrange similar facts in different ways." Feyerabend (1975), p. 286.

[92] Feyerabend (1975), pp. 176-7.

[93] Føllesdal (1973), p. 294.

[94] Føllesdal (1973), p. 291.

language *and* a theory. So in forming a translation manual "we are just correlating two comprehensive language/theories concerning all there is."[95] The distinction between a translation manual and an empirical theory is what motivates Quine's second argument for the indeterminacy of translation thesis.

"A translation manual," points out Christopher Hookway, "is simply a mapping from expressions to expressions [... it] simply tells us which pairs of expressions have the same meaning; it does not tell us what they mean."[96] From the stimulus synonymy of sentences it is possible to infer which words (or sentence parts) are synonymous too, and this allows the construction of a manual. In positing such translation manuals, the question is: 'How far will it take the radical translator?' – "how much of language can be made sense of in terms of its stimulus conditions"[97]?

The radical translator's dependence on stimulus conditions encounters two problems. The first problem is that many different English observation sentences have the same stimulus meaning as one alien observation sentence. So 'Gavagai' could be translated as 'Behold, a rabbit', 'Behold, an undetached rabbit part', or even 'Behold, an instantiation of universal rabbithood'. Even when the translator has decided on which theory of nature to attribute to the alien, the "choice of translation manual is still open."[98]

The second problem is that the stimulated assents and dissents "do not reflect semantic properties of individual sentences."[99] For example, from the alien *dissenting* to 'Gavagai', it does not follow that 'Gavagai' does not mean 'Behold a rabbit'. The translator may not distinguish between rabbits and hares and so may himself utter 'Behold a rabbit' in the presence of a hare when the more discriminating alien, knowing that a hare is not a rabbit, will refrain from uttering 'Gavagai'. Yet 'Gavagai' could still be translated as 'Behold a rabbit', even in the case when the alien *dissents* to 'Gavagai' and the translator *assents* to 'Behold a rabbit'. This case hinges on a difference in the knowledge or beliefs of the two parties. It

[95] Føllesdal (1973), p. 295.
[96] Hookway (1988), pp. 151-2.
[97] Hookway (1988), p. 129.
[98] Hookway (1988), p. 137.
[99] Hookway (1988), p. 134.

shows that a translator can retain the translation manual pairing 'Gavagai' with 'Behold a rabbit' if he makes adjustments in other parts of the manual, namely, in the translations of standing sentences (for example, the alien believes that rabbits are under a certain size, or can run under a certain speed.) And this can be done for all the possible manuals which arose in the first problem.

Since appeal to stimulus synonymy cannot determine a unique translation manual; and since, because of Quine's physicalism and his view that linguistic facts are necessarily public, there are no other facts to appeal to, Quine decides that choice of translation manual is indeterminate with respect to the facts.

Underdetermination is a notion which applies to theories of nature and indeterminacy applies to translation manuals. To distinguish underdetermination from indeterminacy, it will therefore be helpful to distinguish a theory of nature from a translation manual. This is not so easy to do, for as Dagfinn Føllesdal observes, "the view that translation manuals are just a species of empirical theories is deeply rooted"[100].

When I construct a translation manual, "the only entities I am justified in assuming are those that are appealed to in the simplest theory that accounts for all the evidence."[101] Since theory is underdetermined by the physical evidence, the final theory choice is made on pragmatic grounds such as simplicity. Now, since Quine maintains "that such theories make claims about the world"[102], and since our theory preference is determined by pragmatic features, it follows that we use such pragmatic features as "guides to truth."[103] Epistemologically underdetermined as the theories are, only one of them can be the true theory (and in fact neither may be). Pragmatic features such as simplicity and personal quirks such as my own laziness also play a role in determining my choice of manual, *yet here they play no role as guides to truth*: there is no true translation manual because "[i]n translation we are not describing a further realm of reality"[104]. Hence Quine's remark:

[100] Føllesdal (1973), p. 296.
[101] Føllesdal (1973), p. 295.
[102] Føllesdal (1973), p. 293.
[103] Føllesdal (1973), p. 293.
[104] Føllesdal (1973), p. 295.

> [W]hen I say there is no fact of the matter as regards, say, the two rival manuals of translation, what I mean is that both manuals are compatible with all the same distributions of states and relations over elementary particles.[105]

The different ontological commitments of different translation manuals are consistent with the one chosen theory of nature, and for this reason "[s]tatements about the ontological commitments of theories will be relative to translation manuals."[106] From this claim comes Quine's notion of inscrutability of reference, the idea that even as I hold a theory of nature, different sets of ontological commitments are open to me: "I can systematically reinterpret my own utterances and conclude that 'rabbit' in my mouth is true of rabbit parts or stages."[107]

In Quine's second argument, indeterminacy of translation comes about because, given a theory of nature, many different translation manuals, and so many different ontological commitments, are possible. In the PTO, by contrast, Feyerabend makes no use of translation manuals, and this rules out any significant similarity between the PTO and Quine's second argument. Since Feyerabend's wider view is that ontological commitments are given by *theories*, Quine's second argument, locating indeterminacy in translation manuals, would suggest that Feyerabend's wider view can only assert that ontological commitments (and meanings) will be *underdetermined*, not indeterminate. So Quine's second argument for the indeterminacy of translation thesis cannot be marshalled to support Feyerabend's claim that he is a Quinean. If Feyerabend's claim that the meanings of observation terms are indeterminate is tenable, it will be because of similarity between his general views on theories and Quine's first argument.

Feyerabend is of the opinion that two people who are disposed to utter the same sentences under the same sensory stimulations may express completely different beliefs or meanings. But this is not what Quine means by translational indeterminacy. In his letters in the early 'seventies, Feyerabend claimed that his PTO anticipated the main ideas behind Quinean radical interpretation.[108] In the early 'nineties, in the third edition of *Against Method*, Feyerabend offers a much weaker and more measured comparison: "Quine [...] also used a criterion of observability

---

[105] Quine, quoted in Hookway (1988), p. 137.
[106] Hookway (1988), p. 141.
[107] Hookway (1988), p. 142.
[108] Feyerabend also maintains that Carnap had been a proponent of the PTO (see, for example, Feyerabend (1965a), p. 212). This claim is challenged by Thomas Oberdan (1990).

that is rather similar to mine."[109] I have argued that there are *similarities* between the PTO and Quine, but that the PTO does not imply that the meanings of observation terms are indeterminate.

Part 4: The PTO: Criticism And Conclusion

The general consensus[110] is that the PTO is untenable. While I think this is true, not all the criticisms levelled at the PTO are justified in my view. Here in Part 4 I will consider six.

First, John Preston criticises the PTO for assuming that the utterance of an observation sentence "will occur independently of the interpretation [...the observer] may connect with the statement"[111]. Preston believes that this assumption "implies, falsely, that scientific observers are not concerned with the meaning of the observation-statements they produce."[112] Dudley Shapere puts the matter vividly when he remarks:

(j) "Feyerabend's observation-sentences, being mere uninterpreted noises, are no more 'linguistic' than a burp."[113]

Preston is surely right to claim that I utter an observation sentence partly because of what I believe and partly because of what it means. But it also seems to me that Preston and Shapere are being a little unfair when they attribute the likes of (j) to the PTO. In Part 3 I argued that when the observer is a healthy human being, Feyerabend's view of an observation sentence is the same as Quine's inasmuch as both Feyerabend and Quine hold that when people say something they also mean something (claim (g)). Claim (j) attributes to the PTO the view that people speak first and think afterwards. But I have argued that the PTO claims that the meanings of observation utterances can *change* because observation sentences can be stimulus synonymous and yet have different interpretations. From such a view it does not follow, and nor does Feyerabend assert, that observation utterances are 'verbal knee-jerk reactions' and meaningless *when made*. Such an interpretation of the PTO overplays Feyerabend's analogy of observers with measuring instruments.

---

[109] Feyerabend (1993), p. 212.

[110] For example: Shapere (1966); Butts (1966); Townsend (1971); Hull (1972); Hacking (1975), p. 128 observes the consensus; Suppe (1977), p. 638; Suppe (1991); Preston (1997), pp. 45-54.

[111] Feyerabend (1965a), p. 198.

[112] Preston (1997), p. 48.

[113] Shapere (1966), p. 60.

The second criticism, related to the first, is that Feyerabend's notion of a conditioned observer is unsatisfactory. John Preston maintains that the PTO "conflate[s] the nomic regularity of causation with the normative regularity of rule-governedness."[114] At first glance, this criticism may seem unfair. After all, the PTO does *distinguish* causally determined behaviour dispositions from behaviour dispositions determined by social conventions and linguistic rules. Considering first the nomic relation, the PTO maintains, rightly or wrongly, that the disposition to make a verbal response to a given stimulation is uniform among the members of a language group. Here, the physical causal relation is that between stimulation and disposition to respond. The verbal character of the response one is disposed to give, however, is conventional insofar as it is taught. The symbols and syntax are chosen by convention and their manipulation will conform to conventional rules. A defender of the PTO might then assert that the above causal and conventional qualities of observation sentences allocate *distinct* nomic and normative roles to speech dispositions.

The problem with this defence of the PTO is that it (perhaps unwittingly) leaves open the possibility that the normative rules governing syntax and symbol manipulation could *conflict* with the nomic or causal rules governing utterance disposition. This conflict would then be evidence of conflation. The conflation which criticism two alleges is found in the PTO's claims that one is disposed, for purely nomic reasons, to make a particular *utterance*, but that the *form* of that utterance is purely conventionally determined. Since a disposition to make an utterance is a disposition to make an utterance of a certain form, the nomically determined disposition gets mixed up with the normatively determined form.

The third criticism is that, in the PTO, the disposition to assent to an observation sentence is not formed on the basis of what the sentence means. Indeed, observation sentences can mean just about anything, depending on the theory held. Consistent with this view, the criterion of predictive success has the consequence that "one is caused to accept observation sentences which, as the observer interprets them, may be true or false."[115] Feyerabend admits that the acceptance of an observation

[114] Preston (1997), p. 54.
[115] Suppe (1977), p. 638.

sentence "has nothing to do with the truth of the theory [or beliefs which the sentence expresses]."[116] The third criticism is that it does not make sense to separate like this the truth and meaning of an observation sentence from acceptance or rejection of that sentence. I agree. It seems to me that the PTO does not clearly explain the relation between, on the one hand, syntax production/simulation and, on the other, the empirical theory which constrains the meaning of the observation syntax.

The fourth criticism, made by Dudley Shapere in 1966, criticises the criterion of predictive success because, while the criterion aims to reproduce syntax which indicates the order of stimulations, scientific theories often edit information about stimulations:

> [S]cientific theories often, as a matter of fact, alter that order rather than imitate it; and in many cases some of the elements of experience are declared irrelevant. So 'interpretation', rather than 'imitation', takes place even with regard to the alleged 'order' of experience or sensations.[117]

Feyerabend anticipates this problem (though Shapere does not seem to notice), for in 1965 he writes:

> Not every interpretation of the sentences uttered will be such that the theory furnishing the interpretation predicts it in the form in which it has emerged from the observational situation.[118]

But this seems to undermine the criterion of syntax prediction which the PTO proposes; at the very least, Feyerabend's remark only serves to make even less clear the relation between syntax prediction and the theory of nature used to interpret the syntax.

The fifth criticism also addresses the lack of clarity in the relation between syntax prediction and a theory of nature. It sounds like the criterion of predictive success allows for an optimum theory – the one which can simulate all the observation sentences. Since "observational statements are not meaningful unless they have been connected with theories"[119]; and since one theory is chosen by the criterion of predictive success; then Feyerabend's reason (h), claiming that any observation sentence has many different interpretations, is contradicted. The alternative is that many different but empirically adequate theories of nature will be able to meet the

[116] Feyerabend (1965a), p. 216.
[117] Shapere (1966), p. 61.
[118] Feyerabend (1965a), p. 214.

syntactic criterion of predictive success; but then the criterion cannot interestingly be used as Feyerabend proposed - as a means of testing theories,[120] for many theories will pass the test.

The sixth criticism points out that, in the PTO, the theory which best predicts observational syntax will be the preferred theory of nature:

> an acceptable theory ... has an inbuilt syntactical machinery that imitates (but does not describe) certain features of our experience. This is the only way in which experience judges a general cosmological point of view.[121]

But there are an infinite number of observation sentences in a natural language, and an infinite number of observations. Translation manuals got round this problem because they were formed for words, not sentences. Since there is no hint of recursivity in the criterion of predictive success, it is difficult to see how it would work for all observation sentences.

In my opinion, the previous five of the six criticisms considered point out grave problems and inconsistencies in the PTO. The untenability of the PTO has consequences for the semantic proposals of the IT. According to Feyerabend:

> The most important consequence of the transition to the pragmatic theory of observation is the reversal that takes place in the relation between theory and observation.[122]

So the PTO was supposed to justify *thesis I* (statement (b)); and *thesis I* motivated the logical independence claim (c). With the dismissal of the PTO, the semantic IT looks bereft of support; but Feyerabend made a number of moves to argue that there was yet life in the IT in general and even in the semantic IT. For one, he mustered other arguments from anthropology, psychology and sociology to support his scepticism about the meanings of observation terms.[123] He also made stronger appeals to the history of science rather than to abstract arguments in the philosophy of language. Such moves suggested that the IT concerns a much wider range of phenomena than merely the semantic relations between the statements of scientific theories. This suggestion is borne out by recent publications[124] about the IT, for they

---

[119] Feyerabend (1965a), p. 213.

[120] "[W]e accept the theory whose observation sentences most successfully mimick our own behavior." Feyerabend (1965a), p. 217.

[121] Feyerabend (1965a), p. 214-5.

[122] Feyerabend (1965a), p. 213.

[123] For example, the linguistic relativity principle of B.L. Whorff (see Feyerabend (1975), pp. 286-7); and Piaget's writings on perception (see Feyerabend (1975), p. 227).

[124] For example, Chang (1999), Hoyningen-Huene & Sankey (eds.) (2001).

deal not only with semantic issues, but also ontology, value theory, rationality and multiculturalism.

What Chapter 2 has tried to show is that Feyerabend's *early* presentation of the IT as a semantic thesis, that is, the MVT, is not convincingly supported by Feyerabend's semantic theory of observation terms. What remains unclear about the PTO is the relation between phenomenally caused stimulations and observation sentences, on the one hand, and experiences, theories, and knowledge, on the other hand. Without an explanation of this relation, the PTO, and the manner in which *thesis I* 'rides shotgun' with it, are very obscure proposals.

If *thesis I* and the logical independence claim (c) are going to survive it will be without the help (or hindrance) of the PTO. The next chapter examines an argument (actually a group of semantic arguments) which, if sound, would show that *thesis I* and (c) are *not* going to survive because they are false.

Chapter 3

The Causal Theory of Reference and The Meaning Variance Thesis

[E]ither the theory of reference is called upon to underwrite the success of contemporary science, or else it is simply a decision about how to write the history of science (rather than the provision of a 'philosophical foundation' for such historiography). The one task seems too much to ask, and the other too slight to merit the title of 'theory'.

Richard Rorty (1980), *Philosophy and the Mirror of Nature*, pp. 287-8.

Introduction

The quixotic PTO was intended to motivate[1] *thesis I*:

(a) *thesis I*: "the interpretation of an observation language is determined by the theories which we use to explain what we observe, and it changes as soon as the theories change."[2]

I have argued that the PTO does not adequately support or explain *thesis I*. This presents a problem for the semantic claims of the IT, for Feyerabend remains committed to semantic holism. Furthermore, *thesis I* is Feyerabend's main justification for the claim that:

(b) in the common domain, the truth of T_1 may be largely independent of the truth of T_2

With *thesis I* largely unsupported, it would seem that (b) is too; and so is Feyerabend's early view that:

(c) *all* statements of T_1 and T_2 may be logically independent in the common domain

With claims (a), (b), and (c) already on shaky ground due to the inadequacy of the PTO, the aim of this chapter is to decide if they are completely without foundation.

The *extent* of meaning change initially proposed by *thesis I* is that an alteration in a general theory would change the meanings of *all* the terms in the theory:

> the change of rules accompanying the transition [...T_1 to T_2] is a fundamental change, and the meanings of *all* the descriptive terms of the two theories, primitive as well as defined terms, will be different: [...T_1 to T_2] are incommensurable theories.[3]

Perhaps never before in the field of Semantics was so much meaning change owed to so many theories with so few adjustments. Such extreme semantic holism takes a very narrow view of what constitutes sameness of meaning and a very wide view of what constitutes difference in meaning. Such extremity makes the notion of 'meaning' of little interest or utility, as early critics[4] pointed out.

In response to the early criticism, Feyerabend modifies his views. First, he claims that a change of theory will not always incur a change in meaning:

---

[1] In that it led Feyerabend to "tentatively put forward" *thesis I*. (See PP1, p. 31).

[2] Feyerabend (1958), in PP1, p. 31. Italics removed.

[3] Feyerabend (1965c), p. 231. My italics.

[4] For example, Shapere (1966), pp. 54-7; Achinstein (1964), p. 504.

> the transition from T to [T*] may not involve a change of meaning because there is no change in the kinds of entities being posited, only in the quantitative values.[5]

The importance of Feyerabend's comment at this point is that he foregoes the extreme holistic claim that "the slightest alteration of theoretical context alters the meaning of every term in the context."[6] The second revision which Feyerabend proposes is that when meaning change does occur as a consequence of the transition between two incommensurable theories, not all terms are (or need be) affected:

> [I]f we consider two contexts with basic principles that either contradict each other or lead to inconsistent consequences in certain domains, it is to be expected that *some* terms of the first context will not occur in the second with exactly the same meaning.[7]

Claim (c), the proposal of *radical* meaning variance, can therefore be disregarded.[8] The rest of this chapter will examine arguments against claims (a) and (b).

Chapter 3 asks if some causal theories of reference convincingly show that there is continuity of reference between the terms of T_1 and T_2. If causal theories of reference succeed in showing such continuity, then the IT's claim (b), that the truth of statements of T_1 may be independent of the truth of statements of T_2, will be judged unconvincing. Causal theories of reference require that there are external components to reference (which causal theorists call 'meaning'); consequently claim (a) is insufficient as an account of the meanings of scientific terms. There are two general reasons why I do not find these arguments compelling. First, it seems to me that the causal theories of reference here considered do not adequately describe how reference is fixed. Second, the notion of reference advanced by causal theories does not adequately address the concerns raised by the IT.

Part 1 shows how the issue of reference arose in the early responses to the IT and advances the hypothesis that the IT relies on some description theory of reference. Part 2 presents six problems generally ascribed to description theories of reference (from now on, generically referred to as the 'Description Theory'). Part 3 presents Hilary Putnam's causal theory of reference (CTR). Part 4 addresses criticisms of

---

[5] Feyerabend (1965b), p. 267.

[6] Shapere (1966), p. 54.

[7] Feyerabend (1965a), p. 180. My italics.

[8] Shapere (1966), pp. 54-5 notes that Feyerabend "introduces at various points, qualifications which appear to contradict" the claim that theory change entails meaning change and claim (c). Yet Newton-Smith (1981), pp. 155-6 ascribes both these claims to Feyerabend; and Suppe (1991), p 303 maintains that "Feyerabend is committed to the view that any change in a (global) theory changes all the meanings of its terms".

Putnam's CTR and concludes that Putnam's description of how reference is fixed is unsatisfactory. Part 5 considers Michael Devitt's modifications to the Putnam view. Part 6 looks at the notion of partial reference, a notion which Devitt, among many others, employs in his CTR. Part 7 considers causal descriptive theories of reference and concludes that *none* of the theories considered in Chapter 3 has given an adequate account of how the reference of scientific terms is determined. This conclusion implies that causal theories of reference do not convincingly undermine claims (a) and (b) which comprise the meaning variance thesis. Part 8 advances a further argument for why claim (b) is not under threat from causal theories of reference. Part 8 also draws some general conclusions about theories of reference and the IT. Part 9 suggests along what lines a kind of semantic IT might be developed.

Part 1: The Relevance of Reference

Peter Achinstein points out that, for Feyerabend, the meaning of a term is constrained by many elements:

> Feyerabend, e.g., in a discussion of the term 'absolute temperature' in thermodynamics, alludes to the definition, derivation and range of application of this expression, as well as to various characteristics [i.e. properties] of temperature determined by the laws of this theory, and suggests that all are involved in understanding its meaning.[9]

Once a change occurs in any of the above aspects of meaning, the meaning changes. Achinstein admits that some of these aspects of meaning "might be deemed relevant for understanding a scientific term and hence for knowing its meaning"[10]; but he is not sure what, if any, aspect is necessary or sufficient for giving the meaning of a scientific term. Arthur Fine criticises Achinstein on the grounds that "Achinstein's analysis does not provide adequate tools for deciding about whether the meaning of a term has changed."[11] If a semantic theory which successfully proposes necessary and sufficient conditions for meaning change can be found, then Feyerabend's claim that at least some terms change meaning in the transition from T_1 to T_2 can be adequately addressed.

---

[9] Achinstein (1964), p. 502, n. 9.
[10] Achinstein (1964), p. 502.
[11] Fine (1967), p. 236.

Fine gives two criteria for meaning change which introduce "what is in effect a notion of 'same extension' as a guarantee of substitutivity."[12] In the same year that Fine proposes his two criteria, Israel Scheffler points out:

> [D]eduction within scientific systems [...] requires stability of meaning only in the sense of stability of reference. That is to say, alterations of meaning in a valid deduction that leave the referential values of constants intact are irrelevant to its truth-preserving character.[13]

For the purposes of theory comparison, as long as the terms of T_1 are co-referential with those of T_2, statements of T_1 are not logically independent of those of T_2. A theory of reference which shows that the terms of T_1 and T_2 are co-referential would defeat Feyerabend's claim (b) for logical independence. However, neither Fine nor Scheffler offer (in the works quoted above) such a theory of reference.

Hilary Putnam does propose a theory of reference which purports to show a massive degree of co-reference between the terms of T_1 and T_2. Using as an example the term 'temperature', Feyerabend asserts:

> Galileo's thermoscope was initially supposed to measure an intrinsic property of a heated body; however, with the discovery of the influence of atmospheric pressure, of the expansion of the substance of the thermoscope (which, of course, was known beforehand), and of other effects (nonideal character of the thermoscopic fluid), it was recognized that the property measured by the instrument was a very complicated function of such an intrinsic property, of the atmospheric pressure, of the properties of the particular enclosure used, of its shape, and so on.[14]

Because Galileo had a general theory of what he was measuring which was different to our theory, "we do not mean by the word 'temperature' what Galileo meant (i.e. what Galileo meant by the synonymous Italian word)."[15] So it is false that:

> the meanings of observation statements as obtained with the help of measuring instruments remain invariant with respect to the change and progress of knowledge.[16]

Putnam rejects Feyerabend's claims about meaning change, asserting that "[i]t is evident that Feyerabend is misusing the term 'meaning'."[17] As far as the Galileo example is concerned, says Putnam:

> Galileo was measuring and theorizing about the magnitude we call 'temperature' in English, but [...] we have somewhat different beliefs concerning it than he did.[18]

---

[12] Hesse (1968), p. 50.
[13] Scheffler, quoted in Sankey (1994), p. 39.
[14] Feyerabend (1962), p. 37.
[15] Putnam (1975), p. 121.
[16] Feyerabend (1962), p. 37.
[17] Putnam (1975), p. 122.
[18] Putnam (1975), p. 129.

What Putnam goes on to do is to offer a semantic theory which shows how reference can remain constant while beliefs about the referent vary. This chapter is concerned with Putnam's and others' efforts to that end. In the various causal theories of reference considered here in Chapter 3, the common idea is:

> as long as we continue to use the word 'temperature' to refer to the same physical magnitude, we will not say that the meaning of the word has changed, even if we revise our beliefs many times about the exact laws obeyed by that magnitude, and no matter how sophisticated our instruments for measuring temperature may become.[19]

Such a theory of reference would drastically limit the occasions of meaning change in theory transitions, and would deny that theory change need ever imply meaning change; *thesis I* (claim (a)) would then look irrelevant in talk about meaning (i.e. reference) change. The logical independence claim (claim (b)) would also be confounded by such a theory of reference; for if, in the common domain, a term in T_1 does refer (and it is difficult to see how the term could have no reference in an empirically adequate theory), then the reference of the term will continue, one way or another, in T_2.

Continuity of reference between the terms of successive general theories can occur in a number of ways. Israel Scheffler's proposal concerned "two theories which share the same predicates but where these predicates have different senses."[20] Michael Martin points out that *some* referential variance is compatible with the claim that statements of T_1 are semantically comparable with statements of T_2. Martin argues that the case where predicates of T_1 and T_2 have overlapping extensions is sufficient for those predicates to be mutually inconsistent. It is not necessary that a predicate of T_1 have an *identical* extension to a predicate of T_2, or that a predicate of T_1 be a *proper subset* of T_2 in order that the predicates co-refer. For co-reference, all that is required is that predicates of T_1 and T_2 have an *intersecting* extension.[21] Martin's point seems to be a fair one. More problematic is the notion of partial reference presented by Hartry Field. The idea behind partial reference is that a predicate of T_1 intersects with two predicates of T_2, but that the aforementioned predicates of T_2 do not themselves intersect. The resulting claim is that the predicate of T_1 does not fully refer to anything, but does partially refer to two things. Field's comments will be looked at later.

---

[19] Putnam (1975), p. 128.
[20] Martin (1971), p. 23.
[21] See Martin (1971), p. 26.

Chapter 1 showed that Feyerabend rejected (in the case of general theories) the Received View's use of bridge hypotheses as a means of achieving theory reduction. He spurned the claim that impetus and momentum are materially equivalent *on the grounds that* their descriptions are mutually inconsistent.[22] Description theories of reference hold that a term denotes an object (or kind of object) because the object satisfies the definite description associated with the term. Impetus satisfies the impetus theory's description of impetus, and this description is inconsistent with the Newtonian description of momentum. Feyerabend's conclusion that 'impetus' and 'momentum' are not co-extensive is based on impetus and momentum satisfying inconsistent descriptions. It would therefore appear that Feyerabend's claims of meaning variance assume a Description Theory of reference. As Howard Sankey puts it, "Feyerabend clearly assumes that considerations between concepts are capable of deciding questions of co-reference."[23] Whether or not the IT *actually* employs the Description Theory of reference is a matter which I will come back to at the end of the chapter.

Jerzy Giedymin has criticised Feyerabend's brand of semantic holism on the grounds that *thesis I* implies:

> the denotations of the primitive terms of a theory are determined by all axioms of the theory [...and] only those assignments of denotations to primitives are permitted under which the axioms remain true.[24]

In which case, "this would make the theory true *a priori*."[25] The internalism of this view, that the meanings of statements are constrained solely by their embedding theory, is further incentive to regard the IT as being susceptible to criticisms to which the Description Theory is also susceptible. For example, it is not clear how non-synonymous terms co-refer; indeed the Description Theory and Giedymin's above observations explain claim (b), the claim that there may be radical discontinuity of reference between incommensurable theories. If the IT does indeed rely on the Description Theory of reference, then successfully challenging the Description Theory will be sufficient to undermine the IT.

---

[22] See Chapter 1, Part 4.

[23] Sankey (1991), p. 227. Sankey points out that attributing the Description Theory to Feyerabend is not a straightforward matter. I return to this at the end of Chapter 3.

[24] Giedymin (1970), p. 259.

[25] Giedymin (1970), p. 259. Suppe (1977), p. 640 makes the same point.

Part 2: Putnam's Objections To The Description Theory

Putnam rejects the Description Theory's account of how we refer to objects. In particular, Putnam claims that the Description Theory inadequately describes how we refer to natural kinds (NKs). NKs are "classes whose normal distinguishing characteristics are 'held together' or even explained by deep-lying mechanisms."[26] Natural kinds include lemons, water, gold and cod. Putnam shows that no conjunction of properties is both necessary and sufficient to pick out a NK. NKs cannot be designated by property ascriptions because of the existence of abnormal members: some lemon will not be yellow, but green or brown, and some will not taste bitter (if grown under certain conditions). Exceptions do not prove the descriptive rule of what a kind is, but undermine it. In this way, Putnam shows that no description is sufficient to designate a NK. Where Putnam is more startling is in his claim that no definition of a NK term is analytic. Here, Putnam distinguishes a term like 'bachelor' from a NK term like 'lemon'. For the definition 'unmarried man' is true of 'bachelor' in such a way that bachelors could not be otherwise; but it could happen (however unlikely) that scientists discover that lemons are not citrus fruits. So the known material make-up of lemons and the classification of that make-up is a matter of scientific investigation and, as such, is always open to revision. Putnam rejects the Description Theory because descriptions of NK terms are neither sufficient nor (analytically) necessary to refer. It is especially his support for the latter which forms the bedrock for his own theory of reference and which will be returned to later in this essay.

Putnam further breaks the bond between descriptions and referents by pointing out that I do not need to know or understand a description in order to designate a NK; so I can refer to a lemon without needing to know its DNA structure. This point is also developed by Donnellan[27] in his distinction between using an expression referentially and attributively. Take the old Peter Sellers sketch:

> Person A: "Does your dog bite?"
> Person B: "No."
> *Person A goes to stroke the dog and the dog bites Person A.*
> Person A: "I thought you said your dog doesn't bite!"
> Person B: "That is not my dog."

---

[26] Putnam, 'Is Semantics Possible', in Schwartz (1977), p. 102.
[27] 'Reference and Definite Descriptions', in Schwartz (1977).

Person A is using 'your dog' to refer to a particular dog he has in mind, (the dog standing next to person B). This is Donnellan's *referential* use of an expression and it shows that to refer to a particular individual we have no need of a true description of it. Donnellan's *attributive* use of an expression involves reference not to an *intended* individual but to *some unrecognised* individual. The joke is funny because everybody believes that Person A is using 'your dog' referentially – everybody, that is, except Person B who regards Person A's use of 'your dog' as attributive. All this serves to illustrate Putnam's point that in referring to a NK, the description is not of itself important. I can be ignorant of the correct description and still refer (by referential use of a definite description); and I can understand the correct description and still not grasp the referent (by attributive use of a definite description).

A third area in which Putnam has objections to the Description Theory is in its constructivist leanings. Sankey[28], for example, has claimed that the Description Theory gives support to the incommensurability thesis' (IT's) radical discontinuity of reference; but Putnam's point is one of meaning stability when he holds that lemons are lemons (or whatever they are) independently of what I conceive them as. The Description Theory has it that the essential property of a NK is determined by whatever theory I hold at a given time, so that "whether something is a lemon or not ... is a matter of the best conceptual scheme, the best theory".[29] The danger of the Description Theory is that, rather than the referent giving rise to its description, it is the description which is regarded as prior to the referent. Putnam gives a little illustration as to why the 'stuff' of the referent precedes any description of it:

> Even if cats turn out to be remotely controlled from Mars we will still call them 'cats' ... Not only will we still call them 'cats', they are cats.[30]

If all cats turned out to be robots then, no matter whether we *thought* of them as animals or as robots, they are, and always have been, robots; what they are *called* (or described as) does not *make* them what they *are*.

The objection just outlined was that the meaning of a term is its reference and not the description or list of concepts we have of the term. However, 'meaning' is a very

---

[28] Sankey (1994), p. 76.
[29] Putnam, 'Is Semantics Possible', in Schwartz (1977), p. 104.
[30] ibid. p. 107

broad term, and clearly it does encompass linguistic, conceptual and mental elements; it is worth pointing out that Putnam *does* consider these aspects of 'meaning', it is just that he wishes to ground the meaning of NK terms in the external world and not in the formulation of words or thoughts. As he himself says: "Linguistic competence and understanding are not just *knowledge.*"[31] Putnam's comment draws out the fourth criticism of the Description Theory of reference: the Description theory implies that there is little more to the meaning of a term than the mental content I associate with the term.

The way in which the Description Theory deals with non-synonymous co-referring expressions provides a fifth area of the Description Theory with which Putnam disagrees. Putnam employs Kripke's notion of necessary truth to solve the problem of contingent identity statements. Putnam and Kripke would say that *if* the morning star is the evening star in actual fact, then the statement 'the morning star is the evening star' is necessarily true. The Description theory, on the other hand, holds that 'the morning star is the evening star' is true only contingently. Whether the Putnam/Kripke notion of necessary truth is a satisfactory alternative to that employed by the Description Theory is an issue that will be considered later.

Proponents of the Description Theory must struggle to account for how non-synonymous expressions co-refer. What nobody considered, says Putnam, tendering the sixth criticism of the Description Theory, is the possibility that expressions could be (regarded as) *synonymous* and yet *not* co-refer. The reason why supporters of the Description theory tend to overlook such a possibility is because they regarded property ascriptions as providing necessary and sufficient conditions of a term's extension; so two terms co-refer *because* their descriptions were synonymous. Consider the statements:

(1) I live in Paris

(2) I live in the capital of France

If we take all the descriptions of 'Paris' and all those of 'the capital of France' we would find that they are identical; since (according to the Description Theory of reference) the descriptions are the necessary and sufficient constraints on the reference of each expression, the referents are identical too (at least contingently so). Putnam sets the cat among the pigeons by giving an example where terms are

[31] 'Explanation and Reference', in Putnam (1975), p. 199.

synonymous but *not* co-referential. On Putnam's famous Twin Earth (TE), aluminum is called 'molybdenum' and is very rare, whereas molybdenum is called 'aluminum' and is very common. On (American English-speaking) Earth (E), of course, aluminum is called 'aluminum' and molybdenum is called 'molybdenum'; and aluminum is the common metal while molybdenum is rare. When Putnam compares a speaker from E and a speaker from TE, he finds that "there may be no difference in their psychological states when they use the word 'aluminum'"[32] and, correspondingly, they would ascribe identical properties to 'aluminum'. However, the extension of 'aluminum' as used by the TE speaker is different to the extension of 'aluminum' as used by the E speaker. Same term, same descriptions, same concepts, but different referents.

Putnam's direct challenge to the Description Theory is that descriptions and concepts do not fix the reference of NK terms. The Causal Theory of Reference (CTR) which he proposes instead has therefore to deal with the question: 'How is reference fixed?' and, along with that, 'What description is appropriate to NK terms?' Putnam has advanced his main objection to the Description Theory from a number of directions and in various ways, and it is only to be expected that he will advance his CTR by similar manifold means.

Part 3: How Reference Is Fixed In Putnam's Causal Theory Of Reference

In the CTR, the reference of a NK term is fixed without describing any of its properties. We simply point to the NK and name it; from that moment on a NK term has, potentially, entered the language. Put baldly like this, the CTR seems highly implausible for it sounds like Putnam is saying that NK terms refer by a point and grunt mechanism. Putnam is aware that his theory must appeal not just to Neanderthals and those under the age of two; yet he is also aware that he must take the description of properties out of reference fixing. It is Putnam's attempt to square this circle that makes for much of the beauty of his CTR, for what he does is integrate ostensive definition into a broad theory of linguistic competence. In doing this he employs a range of notions, namely: indexicality, introducing event, causal

[32] Putnam, 'Meaning and Reference', in Schwartz (1977), p. 123-4.

chain, causal physical magnitude, division of linguistic labour, stereotype, semantic and syntactic markers. Each of these elements will now be sketched.

The reference of a NK term is fixed by pointing at a NK and giving it a name. This is the (rather unpromising) gist to the CTR:

> Our theory can be summarized as saying that words like 'water' have an unnoticed indexical component: 'water' is the stuff that bears a certain similarity relation to the water *round here*.[33]

Here, we can see that a NK term is defined locally and that the NK must be immediately physically present to the one who first gives it a name. So for Putnam the word 'defined' or 'definition' does not involve a description such as dictionaries provide, for no description is forthcoming. Rather, the definition of a NK term is given by its initial reference. Schwartz puts it this way:

> One would come closer to the position of Kripke and Putnam if one simply said that 'water' has no definition at all, at least in the traditional sense, and is a proper name of a specific substance.[34]

The "certain similarity relation" of which Putnam speaks really just means 'same type of stuff as', and that 'stuff' is described and conceived by us as chemical or atomic. The genius of Putnam's 'same stuff' relation is that chemical and atomic systems of classification may come and go, for they are ways of describing; but *what* they describe will not change, for water will always be the stuff it is.

The first attaching of a name to a NK by use of an indexically given paradigm example is called the 'introducing event' and from then on the name and the stuff are forever wed. Once the introducing event has happened, examples of the NK can be referred to by using the appropriate NK term, even if you are quite ignorant of what it is you are referring to. Putnam removes the ambiguity highlighted by Donnellan by insisting that in all future uses of a NK term, the reference of the term is given by the initial baptism (or referential use, as Donnellan would call it). This phenomenon of rigid designation does not only apply to NK terms. In fact, the term 'rigid designator' originally came from Kripke who applied it to proper names. Kripke gives as an example a person who believes that Quine was a Roman emperor. In saying that Quine was in charge of things when Jesus was born, the speaker is *actually* referring to the contemporary Harvard logician. This is because Willard van

---

[33] Putnam, 'Meaning and Reference', in Schwartz (1977), p. 131.

[34] Schwartz (1977), p. 30.

Orman Quine was once given his name in a baptismal introducing event. That attachment of term to referent is thereafter passed on to other members of the community who use the name 'Willard van Orman Quine'. All uses of this name form a 'causal chain' whose links could, theoretically, be traced back to that initial baptism which sanctions the correct referent. Putnam and Kripke thus use the ideas of 'introducing event' and 'causal chain' to construct a theory of linguistic meaning from a non-conceptual base: ostensive definition. What was a solitary act of 'pointing and grunting' determines linguistic currency; herein lies the importance of causal chains:

> what is important about Kripke's theory is not that the use of proper names is causal [...] but that the use of proper names is *collective*.[35]

A given community uses 'Quine' to refer to Quine in virtue of their usage being causally connected to Quine's baptism; and this reference relation holds even though one or two individuals in the community think that Quine was a Roman emperor.

The CTR has so far been presented as a theory of meaning for proper names and NK terms. In both cases, the entity, when it is first named, is pointed at and therefore is immediately physically present to the namer. However, to point at a single hydrogen atom or be immediately physically present at a black hole would be problematic; yet Putnam's CTR is good for theoretical NK terms as well as physical magnitude terms. The referents of such terms "are invariably discovered through their effects".[36] *Whatever* causes those huge accelerations of matter is a black hole, and *whatever* causes a frog's leg to spasm and a lightbulb to glow is electricity. An indexical component is preserved in the naming of theoretical entities because, whereas we cannot point to the theoretical entity itself, we can point to its effect (and sometimes that effect will be the effect produced on a measuring instrument such as an ammeter or radio telescope). Putnam clearly asserts that physical magnitude terms are introduced by causal descriptions; once the theoretical entity has been named in this way, the theoretical NK term or the physical magnitude term is disseminated along causal chains throughout the language community.

Putnam gives three conditions which he says must be met for a person to use a physical magnitude term successfully:

---

[35] 'Explanation and Reference', in Putnam (1975), p. 203.
[36] 'Explanation and Reference', in Putnam (1975), p. 202.

(d) the user knows that the term is a physical magnitude term
(e) the user's use of the term is connected causally to the introducing event where a causal description of the referent was given
(f) the referent exists

As regards the first condition, a physical magnitude is that which is able to be 'more or less' in quantity and to have a location. Putnam considers it important that the user of a physical magnitude term knows that the referent has these two properties. So if I am to use the term 'electric current' successfully, I need to understand that electric current can be 'more or less' (strong or weak) and that it can be found somewhere ('all along this wire', for example). Likewise, an electron can be one or many and it has a location ('in this 'cloud' or even just 'somewhere', for example). The second condition for successful use of a physical magnitude term is the user's inclusion in the causal chain and of the introducing event involving a 'causal description'. Here, it seems that Putnam is succumbing to the Description Theory's technique of property descriptions, so that 'makes a frog's leg spasm' is a causal property of electricity. But the causal description could also be 'makes my leg spasm', or 'makes my arm spasm' or 'caused by rain clouds in certain storms'. Putnam would say that if that which makes a frog's leg spasm is the same magnitude as that which is caused by rain clouds in certain storms, then the former is necessarily the same magnitude as the latter. No causal description of electricity is necessary to refer to electricity, but if the description does refer to electricity, it is necessary and sufficient to do so. In this way Putnam continues to eschew the Description Theory's approach. The third condition for successful use of a physical magnitude term is that the physical magnitude exists. This means that those who use the terms 'phlogiston' or 'ether' do not use them successfully for there are no such things.

I can acquire and use NK terms and physical magnitude terms without any particular knowledge of their referents because the term has been passed down to me by causal chains with the referent already attached. Hence I can go to a jeweller and ask for a gold chain without needing to know how to test if it really is gold. Even the jeweller may only have some rule of thumb tests such as weighing it and scraping it. The only way to be sure, in as much that we can be sure, that it is gold is to test further chemical and possibly even atomic properties, and not many people know how to do that; yet we all manage to refer to the NK gold. Putnam makes the point that the devising of crucial experiments to test for a NK such as gold, or a physical

magnitude such as radiation, is the job of experts. He calls this the 'division of linguistic labour': "The division of linguistic labor rests upon and presupposes the division of nonlinguistic labor".[37] Scientists, jewellers and Country and Western singers may be experts in their own fields, but clearly scientists are the experts when it comes to investigating NKs and physical magnitudes. It may be objected that whatever criterion the scientific expert has for gold, such as having atomic number 79, amounts to a description which is necessarily and sufficiently a description of gold; consequently the CTR capitulates to the Description Theory of reference. But Putnam says no: what the test is does not make the stuff what it is; better tests, more accurate ones may be found to better determine if X is the same stuff that we call 'gold'; describing and understanding elements in terms of the periodic table may also be abandoned one day. Clearly, though, when it comes to 'introducing events', and to an understanding of the physical nature of the world, scientific experts have an important role to play in the language community. Hence Putnam calls experts "a special case of ... being causally connected to an introducing event."[38]

Is Putnam saying that, in the division of linguistic labour, experts define what a NK is? Here, the answer would have to be "No". When a NK term, like 'water' or 'gold' has already had its introducing event (which they both have), then the NK *will* always be the kind of stuff it *was*. The role of the experts is to determine if a particular entity, such as the ring on my hand, is a member of the NK class gold. The class of stuff is fixed by a paradigm example ostensively defined at an introducing event; the experts decide if what is on my finger is part of that class of stuff or not. Members of a class are called the 'extension' of that class. This explains Putnam's comment:

> When a term is the subject of linguistic labor, the 'average' speaker who acquires it does not acquire anything that fixes its extension.[39]

In the case of the term 'gold', the reference is already given (the NK gold) and this the experts cannot change. The 'reference fixing' done by experts is often really 'extension fixing' (telling me if my ring is gold, telling me if the tree in my garden is an ash); unless, that is, the expert is the introducer of a NK or physical magnitude term.

---

[37] Putnam, 'Meaning and Reference', in Schwartz (1977), p. 124.
[38] 'Explanation and Reference' in Putnam (1975), p. 205.

In the CTR the reference of a NK term is fixed by an act of ostension and naming at an introducing event. The reference of a theoretical NK term or physical magnitude term is fixed by a naming of and (implicit) act of ostension towards something causally related to the physical magnitude at an introducing event. Causal chains spread the word and its fixed reference throughout the community. Experts in various fields test for alleged tokens of the NK and physical magnitude but, in the division of linguistic labour, most people get on with referring to the rings on their fingers as 'gold' in the hope or belief that their jeweller is not unreliable. People can do this because they know that gold is yellow, shiny, metallic, heavy, and not brittle. In other words, they have a 'stereotype' of gold. In spite of all this talk of reference fixing, Putnam makes it clear that having linguistic competence of a NK term is "more than just having the right extension or reference".[40] It also involves "associating the right stereotype"[41] with the term.

Stereotypes are the normal, everyday descriptions we use of NKs. So the stereotype of 'dog' can include any or all of the following: has four legs, has a tail, is covered with hair, has a snout, barks. Now not all dogs have all these properties (some dogs which have survived car accidents have only three legs). Such differences do not matter because the stereotype is merely a description we associate with the NK; "it is not a necessary and sufficient condition for membership of the corresponding class."[42] It is not even a necessary condition of being a dog that it be an animal, for if cats could turn out to be robots, dogs could likewise be 'anti-cat devices' (sent from Pluto). In other words, stereotypes do not determine reference, but they are the quick and easy way for us to grasp whether a given object falls within the extension of a term.

Putnam likens a stereotype to "an oversimplified theory"[43] and, as such, its terms are theory laden. The NK term 'dog' is theory laden with the stereotype of being an animal; but were we to find that dogs were robots we could say 'dogs are robots' without any internal contradiction thereby avoiding any paradox:

> I can refer to a natural kind which is 'loaded' with a theory which is known not to be any longer true of that natural kind, just because it will be clear to

---

[39] Putnam, 'Meaning and Reference', in Schwartz (1977), pp. 126-7.
[40] Putnam, in Schwartz (1977),'Is Semantics Possible', pp. 177-8.
[41] ibid., p. 178.
[42] 'Explanation and Reference', in Putnam (1975), p. 205.
[43] Putnam, in Schwartz (1977), 'Is Semantics Possible?', p. 113.

> everyone that what I intend is to refer to that kind, and not to assert the theory.[44]

Of course it takes time for the new theory or stereotype to spread in the language community; but once it has spread, the semantic marker of the word 'dog' will include 'robotic device' and not 'animal' any longer. I can find no hard and fast distinction between a semantic marker and a stereotype except perhaps that the notion of 'semantic marker' includes the class NKs (i.e. the class of the class 'natural kinds'). Semantic markers are fundamental stereotypes such as 'animal' and 'liquid' as opposed to other stereotypes like 'hairy' and 'transparent'. There also may be an implication that stereotypes are more ideolectal then semantic markers, for the latter form community-wide componential analyses which may be altered by the theories (stereotypes) emanating from individuals.

The final element needed for linguistic competence, according to Putnam, is the syntactic marker. This is the knowledge of 'well-formedness': is the term countable? does it take a singular or plural form of the verb? Knowledge of syntactic markers lets us be clear about, and rightly express, the difference between:

(3)Two of my hairs were removed by the beautician.

(4) All of my hair was removed by the beautician.

To use the word 'lemon' competently, then, Putnam claims that we need to know four things. First, that it is a NK with an essential *sine qua non* (for example, a particular DNA structure). Second, its stereotype (for example, yellow colour and tart taste). Third, its semantic markers (for example, organic matter, name of a fruit). Fourth, its syntactic markers (for example, countable noun).

Putnam clearly tells us that reference is fixed and stuck fast by one ostensive naming. But in answer to the question 'What description is appropriate to NK terms?', Putnam goes beyond mere reference fixing to linguistic competence. This broader conception of language entails that meaning does indeed involve knowledge which distinguishes Putnam's position somewhat from Kripke's. Casting our minds back to Kripke's example about 'Quine' referring to a Roman emperor regardless of whether the user of the term was aware of its true referent, Putnam takes a slightly different position to Kripke:

---

[44] ibid.

> [U]nless one has some beliefs about the bearer of the name which are true or approximately true, then it is at best idle to consider that the name refers to that bearer in one's ideolect.[45]

Putnam does not shift from Kripke's view that the user's knowledge does not determine the reference of a term; but an act of *articulation* does require knowledge on the part of a speaker, so that:

> [I]f you had wrong linguistic ideas about the name 'Quine' – for example, if you thought 'Quine' was a female name (not just that Quine was a woman, but that the name was restricted to females) then there would be a difference in meaning.[46]

The other linguistic features he clearly has in mind here which would change are stereotypes, semantic markers and syntactic markers (so that 'Quine' would be syntactically associated with the pronouns 'she' and 'her' instead of 'he' and 'him', for example).

Part 4: Criticism Of Putnam's Causal Theory Of Reference

Before deciding if Putnam's CTR is a panacea to the six ills of the Description Theory of Reference (outlined at the beginning), I will consider some problems with what Putnam has to offer. Some of these problems relate to the historical aspects of the theory (introducing event, causal chains), and others to the very existence and nature of NKs themselves; for clearly a theory of the *reference* of NK terms, if it is to be satisfactory, must also give a satisfactory account of what a NK term is and of what a NK is.

At Putnam's introducing event a term and its referent become attached. The question then arises: 'How can Putnams's CTR accommodate the fact that the meanings of words change, and that words can have more than one meaning?' In the history of English language terms become unattached from their referents and sometimes acquire new referents. A 'broadcast', for example, was the motion of sewing seed by throwing it, not the sending of electromagnetic waves or the television and radio programmes the waves carry; now it means all these things. A further problem with the introducing event is that we can rarely know when it

---

[45] 'Explanation and Reference', in Putnam (1975), p. 203.
[46] ibid.

occurred, so we can never be sure that the stuff referred to by the term bears the 'same stuff' relation to what we refer to by the term. These problems concerning rigid designation and the introducing event have been worked on by Michael Devitt who, as we shall see in a later section, modifies the CTR, proposing that a term has a number of referents and that the introducing event "is only one of many confrontations between a term and the world."[47]

The introducing events of theoretical NK and physical magnitude terms involve a slightly different set of problems. We have already met with Putnam's claim that competent use of a physical magnitude term requires certain conditions to be met, including the user being aware that he is using a term which indicates a referent capable of quantity and location. With regard to electricity, Putnam has this to say:

> I cannot, however, think of anything that every user of the term 'electricity' has to know except that electricity is (associated with the knowledge of being) a physical magnitude of some sort and, possibly, that 'electricity' ... is capable of flow or motion.[48]

William K. Goosens objects that having the knowledge of flow or motion was not necessary to the introduction of the term 'electricity', for perhaps the referent at the introducing event was static electricity. Goosens maintains that the knowledge of 'magnitude' would also not be necessary. His reason for doing so is that knowledge of the referent is contingent and empirically given: "With electricity present, we discover it is capable of flow and is a quantity."[49] However, Goosens point, like Putnam's, is concerned with linguistic competence and not with reference fixing – a distinction which Goosens does not seem to take account of. So, whereas it is not necessary to have location and magnitude in mind in order to fix reference, it is necessary to have those two features in mind when using the physical magnitude term competently. The introducing event of a physical magnitude term would therefore have a rather anomalous nature in that reference could be fixed while at the same time the new physical magnitude term could be used without full linguistic competence. Such linguistic anomaly should not be counted against the CTR because it may indeed be a feature of the early use of some theoretical NK and physical magnitude terms such as 'electrons' and 'quanta'.

---

[47] Devitt (1979), quoted by Sankey (1994), p. 57.

[48] 'Explanation and Reference', in Putnam (1975), p. 199.

[49] Goosens (1977), p. 144.

A more significant problem for the CTR is the conflict between reference fixing at the introducing event and subsequent expert-determined extensions in the division of linguistic labour. This is not simply the problem of overly-rigid designation which Devitt claims to have solved using multiple groundings (more later); for even when a NK term has a 'basket of referents', there is still the relation 'same stuff as', except it is applied to several rather than one referent. Or, to put it another way, this is not the problem of a term being fixed to one referent or to several, this is the problem of the reidentification of *any* NK after the introducing event. Shapere (1982), Zemach (1977) and Mellor (1996) are all concerned with this matter. Shapere's objection to reidentification of a NK is founded on his objections to essentialism and this will be looked at shortly; Zemach's presentation of the reidentification problem is done within the context of TE and this will also be considered later. Mellor's comments on the reidentification of a NK will be considered now.

Mellor believes that, in the division of linguistic labour, the experts' criteria for NK reidentification are actually "causally downwind of the usage they are supposed to constrain".[50] As an example Mellor takes chlorine. The term 'chlorine' was first introduced into the language by Sir Humphry Davy who also demonstrated that the referent of 'chlorine' was an element: a clear case of NK term and NK. Subsequent experts have demonstrated that chlorine has an isotope, and exists as Cl-35 and Cl-37. It is then problematic to say that the initial referent of 'chlorine' is 'the same stuff as' Cl-35 *and* 'the same stuff as' Cl-37: surely 'the same stuff' relation can apply to that which is only and exactly the same stuff? This then is Mellor's first point: "some natural kinds have the wrong archetypes"[51], for Davy's chlorine is not 'the same stuff as' today's Cl-35 and Cl-37 if we follow Putnam's CTR to the letter. The standard response to the problem posed by Mellor is to invoke Devitt and say that when Davy used the term 'chlorine' he was referring to a 'basket of stuff' as was shown by subsequent confrontations between the term 'chlorine' and its referents. Devitt's solution makes it clear that a NK term can have more than one referent, but it does not here give a full enough account of the 'same$_x$ as' relation; in particular it does not show how there are "sound inferences from individual essences to kind essences."[52] I will come back to Mellor's criticism shortly.

[50] Mellor, in Pessin & Goldberg (eds.) (1996), p. 74.
[51] ibid. I take it that an 'archetype' is the stuff initially baptised.

One way used to demonstrate the existence of a kind essence is to reduce each member of the extension of a NK term to its atomic features and then to see what microfeature each member has in common. Shapere criticises the compositional approach by saying that it "is by no means an *a priori* or necessary truth"[53] that the identification of a NK is a function of the content or arrangement of its atomic parts; for "the notion of an independent particle may go"[54]. And if it should be that there are no independent particles, what then *is* a NK? In spite of this problem, reduction is a key aspect to Putnam's theory. How Putnam manages to retain it in his view of the reference of NK terms will be explained in the next paragraph. Before doing so I would point out that the reduction of NKs to microconstituents in turn attaches great importance to theoretical NKs and physical magnitudes because the reference of *all* NK terms would be predicated on the reference of theoretical and physical magnitude terms. The CTR, then, is not in the first instance concerned with common or garden NK terms, but with theoretical NK and physical magnitude terms, and their referents' causal natures.

This is the point where the causal element of the CTR becomes of great importance. The celebrated aspect to the CTR is not its historical theory of causal chains, for more recent adaptations of the CTR have downplayed its historical side. The 'pride and joy' of the CTR is the connexion it makes between a microstructure and the causal properties of that microstructure[55]. Causal properties of water include: under normal atmospheric pressure it forms a solid at 0°C and a vapour at 100°C; it attains its maximum density at 4°C; it requires 4200 joules to raise the temperature of 1 kg of water by 1°C. It is because water is H_2O that it behaves the way it does. The CTR's notion of necessary causal properties is very different to the Description Theory's linguistically necessary properties. This different type of necessity is described by Stathis Psillos:

> This is not a matter of logical necessity, but it is a matter of *nomological* necessity. Had the laws of nature been different, water would have different properties. But those properties being what they are, water has the kind-constitutive properties it does.[56]

---

[52] ibid., p. 70.

[53] Shapere (1982), p. 11.

[54] ibid., p. 14

[55] Putnam's CTR draws on the notion of metaphysical necessity when Putnam makes claims like: "[If water is H_2O, then] it isn't logically possible that water isn't H_2O." (Putnam (1975), p. 233, italics removed). Yet Putnam's CTR also draws on the notion of physical necessity. He discusses these two matters in his retrospective paper 'Is Water Necessarily H_2O', in Putnam (1990), pp. 54-79.

[56] Psillos (1999), p. 288.

For Putnam, the microstructure of a NK accounts for the causal properties it *actually* has. The CTR can withstand Shapere's objection (given in the previous paragraph) because our current scientific atomism may radically change, but the causal properties of NKs will never change (in the actual world).

So continuity of reference has been expressed as a problem of reidentification. A NK could be reidentified by examining its microfeatures; and the microfeatures are inferred from, or expressed as a function of, causal relations. If water is H_2O then it is so necessarily: water will retain its causal properties in counterfactual situations, *and so* will retain its microstructure. If water is H_2O then it is so necessarily for no other reason than the way it behaves.

In considering Putnam's claim that "[a] statement can be (metaphysically) necessary and epistemically contingent"[57], consider how the statement 'X is H_2O' operates within the CTR:

(g) If an individual X is H_2O then necessarily X has causal features f1 and f2.

(h) If an individual X has causal features f1 and f2 then necessarily X is H_2O.

Dudley Shapere criticises the CTR on the grounds that "it seems impossible to show how, on the Kripke-Putnam view, scientists could ever come to the conclusion that they were mistaken"[58]. This criticism presupposes that the CTR asserts something like statement (g). The problem with (g) is that it is false only when X does not have causal features f1 and f2. Whether X is H_2O or not makes no difference to the truth of (g); so whatever entity is *posited* will be the entity *referred to*, as long as it has causal features f1 and f2. If (g) is a claim of the CTR, then the CTR runs into the problem that the rigid designation it proposes is too rigid ((g) supports the reference of 'phlogiston', for example). Putnam tries to side-step this problem by insisting (as I pointed out earlier) that NK and physical magnitude terms must denote only entities which exist. So statement (g) would be modified to say that *if* H_2O exists *and* X is H_2O then X necessarily has causal features f1 and f2. But this merely brings us back to the CTR's claim that microstructural description is the description of the essence of a NK. As Putnam puts it:

> I pointed out that difference in microstructure invariably (in the actual world) result in differences in lawful behavior [...] Since there is a standard description of microstructure, and the microstructure is what determines physical behavior

[57] Putnam (1977b), p. 130.

[58] Shapere (1982), p. 8.

(laws of behavior), it seemed to me that the only natural choice for a criterion of substance-identity was the microstructural criterion.[59]

I will return to the microstructural criterion for substance identity in a moment. Looking at statement (h), according to the CTR, the theoretical elements of the water molecule are inferred from causal relations and this order of inference is expressed in statement (h). However, statement (h) is not true, for to move from causal properties (f1 and f2) to constituent properties (H_2O) is not a valid inference: other constituent properties may have the same causal features; or other constituent properties may have the same causal features (f1 and f2) plus an as yet unknown causal feature f3. The scientific inference within a causal theory is therefore better expressed:

(i) If an individual X has causal features f1 and f2 then necessarily X is H_2O or some other thing(s).

My understanding of Putnam's CTR is that it asserts (g) and (i). (g) gives rise to the problem of too rigid designation and the microstructural criterion for substance identity. (i) implies a very weak constraint on the reference of 'water'. If the CTR is to succeed in describing adequately the reference of NK terms it will need to deflect criticisms of the two problems associated with (g).

Twin Earth (TE) provides the litmus test for the continuity of reference of NK terms by means of rigid designation. Putnam is adamant that 'water' refers to the substance H_2O and only H_2O (if water is H_2O), and that TE people, who call the substance XYZ 'water', are using the term incorrectly. For this to be the case, Putnam would have to assume that the term 'water' received its introducing event on Earth (a fact which he does not seem to state explicitly). Putnam then resets the year to 1750, before modern chemistry (and the periodic table) developed, and still maintains that 'water' referred only to that substance which we now call 'H_2O'; even in 1750, visitors from Earth to TE would have been using the term 'water' incorrectly (not using correct Earth English, that is). Zemach disagrees and says that in 1750, 'water' had the extension (H_2O or XYZ) on Earth and TE, and that the reference of 'water' had changed to H_2O (and only H_2O) by 1850 (and the arrival of modern chemistry). Zemach does not go into any detail about how 'water' referred to more than one substance; but given Devitt's approach, it seems likely that Zemach's view could be accommodated with a less rigid causal theory.

---

[59] Putnam (1990), p. 69.

Mellor, on the other hand, takes the view that the referent of 'water' need not have any particular microstructure:

> There was water on both planets alike, and there still is. We simply discovered that not all water has the same microstructure; why should it? Because its microstructure is an essential property of water? Well, that's what's in question.[60]

Mellor is not saying that physical entities are without microstructural properties (and hence, causal relations); rather, he doubts that those properties are *essential* kind-constitutive properties.

Shapere agrees that Putnam's very notion of a NK is problematic: "Nothing satisfactory is said about how we are to decide what it is to count as an essential property."[61] If the causal features follow necessarily from the essential kind-constitutive properties (as in statement (g)), then a NK is the subject of a closed definition, such that:

> we discover, from an examination of things of that kind in our spatiotemporal region, what the essence is, and from then on refuse to consider anything to be that kind unless it has that property.[62]

As an illustration of why the above closed view of kind-constitutive properties is wrong, Shapere describes an alternative region to TE, where a field melds the particles of the nucleus of gold atoms, but the stuff regains the normal nuclear characteristics of gold when removed from the region. Within the region described we would still refer to the stuff as 'gold', he says, even though it would not have gold kind-constitutive properties. Shapere then concludes that no common microstructural essence, underwritten by common causal properties, is necessary for a given entity to be called 'the same stuff as' another given entity. If I were to travel to Shapere's region, the only evidence that the ring on my finger is gold would be that the same ring was gold before I arrived there. The different natural laws of Shapere's region would make my ring unrecognisable as gold as far as the CTR is concerned, for there would be no trans-regional causal properties with which to form the relation '$same_x$ as' at the microconstituent level.

---

[60] Mellor (1996), p. 72.
[61] Shapere (1982), p. 4.
[62] ibid., p. 5.

Putnam now appears to accept much of the force of Shapere's example:

> I do not think that a criterion of substance-identity that handles Twin Earth cases will extend handily to 'possible worlds'. In particular, what if a hypothetical 'world' obeys different laws? [...] It is clear that we would call a (hypothetical) substance with quite different behavior water in these circumstances. I now think that the question, 'What is the necessary and sufficient condition for being water in all possible worlds?' makes no sense at all. And this means that I now reject 'metaphysical necessity'.[63]

Shapere's example shows that the notion of possible worlds does not fit well with the CTR; but Shapere does *not* show that the essentialist views of the CTR are untenable in the actual world - the universe with the physical laws it actually has.

It seems to me that, removing possible worlds from the picture, the CTR withstands Shapere's criticism; but the CTR's microstructural criterion of substance identity still begs the question in the way Mellor has pointed out. Furthermore, the microstructural criterion for substance identity of Putnam's CTR has the undesirable consequence of making reference too fixed (as was seen from statement (g)); however Michael Devitt's CTR will address this problem of rigid designation (in the next part).

I conclude, then, that Putnam's views on the rigid designation of NK (and physical magnitude) terms have gone awry. On the one hand, the criterion of substance identity makes reference too fixed (statement (g)); on the other hand, the CTR, in claiming statement (i) (that identity of causal features does not entail that water has certain constituents, it only makes it *possible* that they do), makes too weak a claim about NK essences to ensure the reidentification of NKs. Consequently, the reference of NK terms is not adequately fixed.

To claim that the essence of a NK is a particular microstructure meets with the objection that "an essential property need not be a fundamental one".[64] So gold could be defined in deeper quantum-mechanical terms than its atomic number, suggests Shapere. Shapere's point is not so much which is the right microstructural description, but which is the *essential* one: how deep must we dig to find the essence of a NK? Shapere also criticises Putnam for thinking that "there are well-

---

[63] Putnam (1990), pp. 69-70.

[64] Shapere (1982), pp. 4-5.

circumscribed boundaries between substances or kinds, and well-defined sets of essential properties for them."[65] Stathis Psillos agrees that "there are borderline cases, or untypical cases [...] especially when it comes to biological kinds".[66] But he continues:

> But the very possibility of untypical, or borderline, cases requires that there are typical and clear-cut cases of belonging to the extension of a kind.[67]

However, it will be recalled that Putnam's argument from abnormal members was precisely one of his *criticisms* of the Description Theory of reference. Now it would seem that what is sauce for the Description Theory goose is *not* sauce for the CTR gander! Stathis Psillos appeals to the relation of nomological necessity (as has already been quoted) to defend NKs when he states:

> Had the laws of nature been different, water would have different properties. But those laws being what they are, water has the kind-constitutive properties it does.[68]

Since such uses of the word 'water' create in us a natural predisposition towards NKs (the very issue in question), let us talk about some individual sample of water. This individual sample is indeed the stuff it is, and if that is all Psillos is saying, then that is not an argument for the existence of the *NK* water. The nomological necessity is that the individual sample of water is H_2O *or other components with identical causal features*, and this is not a very interesting constraint on the reference of NK terms.

Distinguishing NKs from nominal kinds is also a problem for Putnam. Schwartz is surely right to point out that there are important differences between gold, water and tigers on the one hand, and bachelors, lawyers and boats on the other. Putnam, however, wishes to claim indexicality and rigid designation for artifact terms in addition to NK terms:

> It follows that 'pencil' is not synonymous with any description – not even loosely synonymous with a loose description. When we use the word 'pencil', we intend to refer to whatever has the same nature as the normal examples of the local pencils in the actual world.[69]

Schwartz makes the interesting remark that NK terms are well served by the CTR but that nominal kind terms are better suited to a Description Theory of reference.

---

[65] ibid., p.15.
[66] Psillos (1999), p. 288.
[67] ibid., pp. 288-9.
[68] ibid., p. 288.
[69] Putnam, 'The Meaning of Meaning', in Pessin, Andrew & Sandford Goldberg (eds.) (1996), p. 26.

Schwartz's views have more *prima facie* appeal than the procrustean bed which Putnam recommends.

One final problem for the CTR is the *qua* problem. Ostension and causal contact are not enough to fix reference: by listing causes and effects associated with a magnitude, and even by pointing, it is not clear if I am referring to a particular object, a group of objects, or a representative sample of objects. The need for a categorial term is addressed in the rest of the theories of reference in this chapter.

With regard to the Description Theory Of Reference we started with six problems. The first was that there is no definite description of a NK (or physical magnitude) which is necessary and sufficient to refer. Putnam instead suggested that NK terms were rigid designators, a move which will need to be modified and which we will examine in the next section; Putnam's own solution to the first problem has not been completely adequate, but merits development.

The CTR has met with more success in the second problem of explaining how terms refer even when their associated descriptions are unknown or inaccurate.

The CTR avoids the linkage of meaning with a theory or conceptual scheme (the third problem) and instead uses indexicality to fix reference; but this method does so at the cost of there not being practically any theory at the introduction of a term, merely "the assumption of a something-I-know-not-what".[70] Subsequent extensions of a term depend on a 'same$_x$ as' relation which *is* theory-based. All use and extensions of the term after the introducing event are therefore subject to the influence of theories/conceptual schemes. Putnam has offered strong reasons for why meaning, that is *reference*, is not a question of beliefs, by keeping conceptual elements outside the reference fixing event; yet the *communication* of that reference does require certain beliefs.

[70] Shapere (1982), p.21.

The CTR deals with the Description Theory's problem that identity statements involving non-synonymous co-referring expressions are merely contingent (the fifth problem) by saying that they are epistemically contingent but metaphysically necessary; but to do this, the CTR has used a very different type of necessity to the one used in the Description Theory. Putnam has offered us a certainty based on natural causes rather than linguistic ones, surely a more assured base on which to start. That the co-referents of certain non-synonymous terms necessarily enter into identical causal relations is a more satisfying statement than saying that they 'happen to have' identical descriptions. Using the same notion of necessity, the CTR also shows how apparently synonymous terms may not co-refer (the sixth problem).

Part 5: Devitt Does Designation

While Putnam's CTR has addressed issues of reference which the Description Theory found problematic, the CTR has also created some problems of its own. The *qua* problem arose because, even in the initial act of pointing and naming, a role for description became apparent. The microstructural criterion of substance identity proved problematic because the argument from causal properties to essences was not convincing. For Putnam, the introducing event of a NK term ostensively defines not merely an individual, but a class. The reidentification of members of the class depends on the 'same$_x$ as' relation which is theoretical and microstructural in import. Putnam admits as much, but he downplays it by asserting that *whatever* your theory is, the 'same$_x$ as' relation will always reidentify *that* stuff. The problematic consequence of such naïve causalism is:

> any abandoned term will refer, no matter how mistaken and misguided are the descriptions associated with it, given that some thing or other was present in the grounding of the term.[71]

Yet not everything that is named, even for causal reasons, exists. Causal constituents *have* to be inferred for the purposes of science and these constituents will be described by a theory. If the theory is discredited then the identity, and even the existence, of the causal constituents come into question. This issue will be brought out more fully when phlogiston theory is considered in Parts 7 and 8. Up to now, I have argued that Putnam's aim "to get away from the picture of the meaning of a word as something like a *list of concepts*"[72] leads him to underplay the role of theoretical content in referring. I have also argued that Putnam's distinction

---

[71] Psillos (1997), p. 270.

[72] Putnam, 'Is Semantics Possible?', in Schwartz (1977), p. 111.

between, on the one hand, the introducing event as a reference fixing event independent of mental content, and on the other hand, subsequent uses of the term which assume linguistic competence and mental content, is too tidy.

In what follows, I will look at how Michael Devitt tries to remedy Putnam's problems. The apparatus of Devitt's theory is first given mostly with regard to proper names. I conclude that Devitt's CTR inadequately accounts for the reference of proper names; and as an account of the reference of NK terms, Devitt's CTR has all the problems it had with proper names, plus some more. My criticism of Devitt is that there are flaws in the particulars of what he proposes: the Devitt is in the detail, as it were.

Devitt regards Putnam's conception of causal chains emanating from a single introducing event as "an idealized picture"[73]. Rather then one introducing event there can be many *groundings* of a term in more than one object, or many times in the same object. A grounding occurs under certain conditions. First, a person perceives an object, "preferably face-to-face"[74]. Second, the person's belief that the object belongs to a very general category is true. Third, the person acquires a new ability to use a term or has on old ability reinforced.

The first condition, requiring a physical encounter, is implicit in Putnam's theory, where ostensive definition is stipulated at an introducing event. Devitt agrees with Putnam that theoretical entities "cannot be grounded by perception"[75] directly and so he advocates the description of causal properties. Devitt takes the notion of 'quasi-perception' perhaps slightly further than Putnam would like when he claims that "certain sorts of representations of the object"[76] can be used to ground a term. Here, Devitt is thinking of non-linguistic representation, so that "a film or painting of an object can serve as well to ground a name in the object as perceiving the object."[77]

---

[73] Devitt (1981), p. 27.
[74] Devitt (1981), p. 133.
[75] Devitt (1981), p. 199.
[76] Devitt (1981), p. 59.
[77] Devitt (1981), p. 59.

It is the second condition for grounding which appears most at odds with the CTR, for the belief condition is one which is associated with the Description Theory. However, it will be recalled that Putnam thought it necessary to have "some beliefs about the bearer of the name which are true or approximately true"[78]. I regarded Putnam as saying that such beliefs were only necessary for linguistic competence and not for reference-fixing.[79] But for Devitt, belief has a role both in reference-fixing and linguistic competence. Devitt wants to incorporate mental representations into a CTR and does so by pointing out that to take part in a grounding (and thereby acquire the ability to designate an object), we must have some belief which is *caused by that object.* He explains this addition to the CTR:

> The central idea of a causal theory was that present uses of a name are causally linked to first uses. I claim now that first uses are causally linked to the object.[80]

Devitt's point is that "To perceive something is to be causally affected by it."[81] It is not possible to perceive an object without having a mental representation (or thought) which is of that object and caused by that object - such is the nature of perception. It may look as if Devitt is claiming merely that a grounding involves simple intentionality, and that there is no belief about the grounding object which is necessary for a successful grounding to occur; and this claim is surely no significant addition to Putnam. But I will shortly argue that Devitt assigns a much greater role to mental content in reference grounding.

Peter Sellers' sketch is again called upon, this time to illustrate the role of beliefs in groundings. From Devitt's perspective, the problem situation can have arisen out of two groundings of the description 'your dog' (and its corresponding 'my dog'). Person A referred to (Devitt would say 'designated') the dog physically present: he perceived the dog, he had a mental representation of that dog, and that mental representation involved the belief that the dog present belonged to the man present. This belief was therefore caused by the dog. This is *not* to say that the dog is *responsible for the truth or falsity* of the man's belief; it is merely to say that the belief concerns the dog and that without that dog there would be no such belief which could be true or false. Person A then has the *new* ability to designate the dog present with the words 'your dog', and this meets the third condition of a grounding:

---

[78] Putnam, 'Explanation and Reference', in Putnam (1975), p. 203.

[79] Though Putnam is not entirely clear on this matter, as the fact that there are different interpretations of him suggests, e.g. Devitt (1981), p. 197.

[80] Devitt (1981), p. 28.

a new ability to use a term (or terms). As far as groundings are concerned, the problem highlighted by the sketch is that Person B's use of the corresponding definite description ('my dog') has a different grounding to Person A's 'your dog'. Person B's use of 'my dog' is grounded in a different dog, a dog which (let us assume) he has perceived and of which he is thinking when he uses the description 'my dog'; so Person B's ability to refer to the dog which actually belongs to him started at a different grounding. Each person's bringing his respective object to mind in tandem with the respective definite description ('your/my dog') constitutes each person's ability to refer to each dog. Devitt sums up his whole approach thus:

> My strategy is to tie an ability to an object and a term in virtue of their role in bringing about the relative mental representations.[82]

A question which will arise again and again about Devitt's strategy is 'What are *relevant* mental representations?'

For a grounding to take place there must be a significant connexion between my mental representation and the object I am designating. If I conceptualise the dog lying on the floor in my kitchen as an electric kettle, then I have simply not adequately perceived the dog. I can *call* the dog ' the electric kettle' or any appellation I wish, and successfully refer to it if I have that *dog* in mind; but if I think of the dog as something that I can fill with water, and as something which heats water (but only to dog's body temperature, of course), then I have not perceived the dog sufficiently to have 'grounding thoughts' in it. According to Devitt, a "successful grounding will be in an object that fits a category determined by the mental states of some person"[83]. I will call this condition the sortal predicate requirement for a grounding. At a grounding, an object gives rise to its term *and* a belief or beliefs about the object. This then begs the question of what beliefs are necessary for a grounding to happen.

In considering what beliefs are necessary for a grounding to occur, the two sides of the matter are, on the one hand, that the belief contain an accurate sortal predicate, such that "the cause must be an object of the sort [...the grounder] has in mind"[84]; and on the other hand, the rejection of any belief requirement, instead stipulating

---

[81] Devitt (1981), p. 27.
[82] Devitt (1981), p. 130.
[83] Devitt (1981), p. 63.
[84] Devitt (1981), p. 62.

"that there be *something external to the mind* immediately responsible for the experiences in question."[85] The problems associated with the former side are some of those associated with the Description Theory of reference; and the problems with the latter are some of those associated with Putnam's CTR. Devitt wants to find a middle way with his CTR, for "we have to draw a line somewhere saying that some sort of error invalidates reference. Reference failure is possible"[86]. Considering such reference failure may help clarify what beliefs are necessary for a grounding to occur.

A name introduced for an entity which is thought to exist but does not is called a 'failed name'. When the term 'Vulcan' was introduced to refer to the planet between Mercury and the Sun, the intended naming event failed because "the singular term used to pick out the object for naming, for example, 'that planet', is empty."[87] Devitt is of the opinion that a grounding fails to take place "if there is nothing there of the appropriate category to be named." However, there are examples which run counter to Devitt's. Sir William Herschel observed through his telescope what he thought was a comet and he named it 'Georgium Sidus'. What he actually saw was the planet Uranus; but, in spite of Herschel's possessing the wrong sortal predicate, and in spite of it being Lexell[88] who suggested that Georgium Sidus was probably a planet, Herschel's grounding of 'Georgium Sidus' is regarded as successful, for he is lauded as the discoverer of Uranus. 'Georgium Sidus' is not a failed name.

Devitt believes that "What object the network is grounded in depends, in part, on the mental processes of the person involved in the grounding"[89], and he has strong reasons for believing so. For example, the *qua* problem which confronted Putnam strongly suggests "that the only difference between naming a cat and a time slice of a cat is in the intentions"[90]. Yet Devitt's sortal predicate requirement seems too vague a requirement to be of use in overcoming the *qua* problem. Devitt tries to wriggle out of the difficulties posed here by insisting "only that the object be in the same very general category as it is taken to be."[91] Even then, he weakly admits that "[t]here is

[85] Devitt (1981), p. 62.
[86] Devitt (1981), p. 62. I have removed the italics.
[87] Devitt (1981), p. 176.
[88] According to Kuhn (1996), p. 115.
[89] Devitt (1981), p. 62.
[90] Devitt (1981), p. 61.
[91] Devitt (1981), p. 63.

an element of arbitrariness in our determination of these categories."[92] I have expressed concern about the vagueness of the belief condition on a grounding, but I will now press on with further details, and two further criticisms, of Devitt's CTR.

Groundings are only the first link in what Devitt calls *d-chains*. The name 'd-chain' is short for 'designating chain' and Devitt wants to make a clear distinction between his use of 'designation' and 'denotation' (while 'reference' remains the general term). The distinction exactly parallels Donnellan's referential and attributive kinds of reference (see Part 2). After generating at a grounding the ability to designate an individual by a term, that ability is passed on to others who have not been present at the grounding. Reference borrowing is acquiring the *ability* to designate that which was designated at a grounding. A d-chain may be summed up as a grounding, an ability, and reference borrowing(s). Devitt's views on designation (and therefore on d-chains) will be the focus of attention, but his views on denotation will be used to fill out his picture of reference.

Using his cat, Nana, as an example, Devitt shows how d-chains operate. When the Devitts first got their cat, Mrs Devitt said, "Let's call her Nana." Here, a grounding had taken place and those present (the Devitt couple) had grounding thoughts and the ability to designate that cat with that name. The ability to designate Nana can be passed on, even in the absence of Nana. For example, at his place of work, Devitt might say to a colleague who has never met Nana, "Our cat is called 'Nana'." Then the colleague has borrowed the ability to designate Nana and can say things like, "How long have you had Nana?". Such comments on reference borrowing contain no surprises so far, but later I will present what I think is a glitch in Devitt's account of d-chains.

One benefit of Devitt's notion of d-chains is how they account for identity statements and non-synonymous co-referring expressions. Devitt and other causal theorists are keen to explain – or rather explain away - Frege's notion of 'sense'. In the identity statements 'Muhammed Ali is Muhammed Ali' and 'Muhammed Ali is Cassius Clay', Devitt agrees with Frege that they are 'the same but different'. Expressing these statements in general terms as 'a = a' and 'a = b', Devitt comments:

[92] Devitt (1981), p. 63.

> Frege rightly saw that the solution to the difficulty lay in the different 'mode of presentation' of the object associated with 'a' from that associated with 'b'. Frege's mistake was to embody these modes within 'senses'. For me the modes are types of d-chain exemplified in the networks.[93]

When a person says, 'Muhammed Ali is Muhammed Ali' he is using the same ability twice: each designation comes from the same d-chain.[94] Of even more interest is how Devitt's d-chain approach can be used to adjudicate on the reference of contentious co-referring non-synonymous expressions such as 'dephlogisticated air' and 'oxygen'; but I will leave this until I come to look at Kitcher and Psillos in Part 7. For now, Devitt's relevant proposal is that "the way a name is treated conceptually appears in the account of d-chains in a theory of reference"[95].

Donnellan's distinctions (between the referential and attributive uses of a term) and Devitt's corresponding terms ('designation' and 'denotation') are useful for describing what happens when error enters into the act of referring. Continuing with Devitt's cat examples: I think that Nana, whom I see regularly, is the neighbour's cat when in fact she is my lodger's cat. The neighbour does have a cat, called 'Jemima', so there are two cats, but I have only ever seen and designated Nana. One day I say, "Our neighbour's cat has disappeared." From this simple situation, a number of semantic questions arise.

The first question is, when I make the statement, 'Our neighbour's cat has disappeared', am I *referring to* Nana or Jemima? The Devitt view is that "my description is linked to *both* cats, though the links are of a different kind."[96] I designated Nana – the cat I had in mind. I denoted Jemima – the one who satisfied the description. A tracing of the d-chain would show when it happened that I falsely grounded the definite description 'the neighbour's cat' in Nana[97]; and so a d-chain would explain why I had Nana in mind when I used that definite description.

---

[93] Devitt (1981), p. 153.

[94] When 'Muhammed Ali' has groundings in two different objects (i.e. people), then a person with the two corresponding abilities to refer might say something like, "Oh, you mean *that* Muhammed Ali!"

[95] Devitt (1981), p. 156.

[96] Devitt (1981), p. 49.

[97] Perhaps I never knew that the lodger had a cat, but I did know that the neighbour had one and then I grounded the definite description in Nana; or perhaps the lodger lied to me by purposely creating a situation where I borrowed the wrong reference.

A second question is whether my statement 'Our neighbour's cat has disappeared' is *true*. Devitt's answer is that the truth value of a statement depends on what it means - on whether we consider its referring expression as designating or denoting:

> The object that bears on the truth value of a statement containing a designational token is the object it designates. On the other hand, the object that bears on the truth value of a statement containing an attributive token is the object it denotes.[98]

Definite descriptions can be ambiguous because "the truth conditions of statements containing them vary according as the description is referential or attributive."[99] The statement 'Our neighbour's cat has disappeared' (in the aforementioned example) is one of referential ambiguity: I clearly and determinately *designate* Nana with the description token 'the neighbour's cat'; and I clearly and determinately *denote* Jemima with that same definite description.

The same type of ambiguity can also apply to proper names such as the names of authors, Devitt claims. 'Shakespeare' can *designate* the playwright and poet from Stratford who was perceived at groundings by a number of people who were the first links of d-chains leading up to my use of his name now. 'Shakespeare' can also *denote* (in Devitt's terminology) *whoever* it was who wrote Hamlet (possibly Francis Bacon? Ben Jonson?). Devitt concludes that "the truth value of many statements containing 'Shakespeare' will depend on whether the name is designational or attributive."[100]

Another type of ambiguity which proper names have is the purely designational ambiguity which arises from the fact that more than one person is called 'John'. Similarly, there is more than one bearer of the name 'Nana' (the bespectacled Greek singer Nana Maskouri, for one). Devitt acknowledges the situation: "I am likely to [be able to] designate several objects with the sound type 'Nana'."[101] He then gives how the sound type is disambiguated: "It is only the thoughts that are *about our cat* that are relevant to the ability in question."[102] It is not simply what the speaker has

---

[98] Devitt (1981), p. 53.

[99] Devitt & Sterelny (1987), p. 82.

[100] Devitt (1981), p. 158. Devitt admits that in the name examples he gives, including the Shakespeare one, "we have not clearly introduced an attributive use but rather have started to run the attributive and designational". ibid.

[101] Devitt (1981), p. 130.

[102] Devitt (1981), p. 130.

in mind which determines the designatum; more precisely, "it is the ability exercised"[103], an ability which forms part of a d-chain:

> The reference of a speaker's token of [...'Nana'], who he 'has in mind', is determined by his psychological states *together with* the way those states are causally embedded in the environment. For the token refers to the object which grounds the ability exercised in producing the token.[104]

If Devitt's causalism had a motto, it would surely be: 'Designating thoughts don't come from nowhere'; they are caused by the designatum (in the case of reference borrowing, the borrower uses a term which can be traced back to designating thoughts). A term token refers to (designates) what caused it, that is, to that which gave rise to the ability to use the term token to refer.

In addition to the designational and denotational *ambiguities* of definite descriptions and proper names, Devitt also discusses their referential indeterminacy. He remarks that "there may be nothing in reality to determine whether some name tokens are attributive or designational"[105]:

> In such a case we must say that the token partially designates the object to which it is linked by a d-chain and partially denotes the object picked out by the identifying expression.[106]

Further details of this interplay will be considered in the notion of a false grounding. A name is falsely grounded when the wrong name or a false definite description is attached to an individual under grounding conditions.

Designational indeterminacy may be found in statements which use terms (and their corresponding abilities) where the grounding has gone wrong. Devitt gives the following example: I say to you, 'This is Nana' (the name of my cat) while indicating Jemima (the neighbour's cat). You accept the grounding because you have not seen either cat before, though you have designated Nana before through reference borrowing. *Jemima is black, Nana is not, and neither cat is Persian.* Now consider the following statements you might make (in the presence of Jemima) after this false grounding:

(5) That cat is Nana.

(6) Nana is a cat.

(7) Nana is a Persian.

---

[103] Devitt & Sterelny (1987), p. 59.

[104] Devitt & Sterelny (1987), p. 60.

[105] Devitt (1981), p. 160.

(8) That cat is black.

(9) Nana is black.[107]

Devitt claims that statement (5) is false and statement (8) is true for the reason that the demonstrative 'that cat' is deictic, *Jemima* is the cat being pointed out (there are no others present), and, perhaps most importantly, no token of 'Nana' is employed in (8)[108]. I see no problem in Devitt's claim that (8) is true, but will argue in the next paragraph that Devitt's views on grounding are insufficient to explain why (5) is false. Devitt points out that in (6), (7) and (9), 'Nana' takes on an indeterminate hue because you have two abilities to designate with 'Nana': one grounded in Nana (which you had previously borrowed from me) and one *falsely* grounded in Jemima (as described above); so each 'Nana' token is partially based on both abilities. The upshot, claims Devitt, is that (6) is true[109] (for both Nana and Jemima are in fact cats); (7) is false[110] (for neither Nana nor Jemima is a Persian); and (9)'s truth value is partially true (for Jemima is black and Nana is not). Each of (6), (7) and (9) employ partial reference, for their 'Nana' tokens refer indeterminately and in a limited way to both cats; for it is not clear which designational ability is being exercised. (The details of the notion of partial reference will be addressed in Part 6. Here, I address the details of Devitt's notion of d-chains).

What I do not understand is *why,* in the above account, 'Nana' in (5) is not designationally indeterminate ; for the speaker has two abilities to designate with 'Nana' (one ability borrowed previously and one more recently obtained in the false grounding). So why is (5) not just *partially* true/false? Devitt's reason for denying the indeterminate truth value of (5) is that it is "an identity belief of the sort that *passes on* the benefit of a grounding"[111], so that "any thought associated with 'Nana' resulting from this identification will contain a token grounded in the object designated by 'that cat'"[112]. I can accept that 'that cat' is univocal, but the speaker's use of 'Nana' is surely indeterminate for the reasons stated earlier[113]. I also accept that the grounding *passed on* by (5) is univocal, but what has happened to the

---

[106] Devitt (1981), p. 160

[107] Devitt (1981), p. 143.

[108] Strangely, Devitt does not propose the third reason.

[109] For how a statement can, according to some, be determinately true yet contain a partially referring term, see Hartry Field's position in Part 6.

[110] For how a statement can, according to some, be determinately false yet contain a partially referring term, see Hartry Field's position in Part 6.

[111] Devitt (1981), p. 143. My italics.

[112] Devitt (1981), p. 143.

[113] Namely, that the speaker had two abilities to refer using 'Nana', one borrowed and one obtained in the false grounding, and there seems to be no fact of the matter which ability is being used.

borrowed ability and why is it completely discounted? I agree with Devitt that our intuitions about (5) are that it is simply false, but it seems to me that the machinery of his theory does not adequately explain *why* it is false.

There seems to be a further inconsistency which arises in Devitt's theory of reference. To alter the previous example slightly, I say 'Nana is our cat' (instead of 'This is Nana') when Nana is absent but Jemima is present. You take 'our cat' to refer to the cat present (Jemima). Let it also be the case that Jemima is Siamese and Nana is not. You then say:

(10) 'Nana is Siamese'.

Devitt again states (just as he did of (5)) that your statement is simply false, this time because:

> We must disallow that the groundings of one term can be transmitted to another in this way. Although an identity belief involving a nondemonstrative representation can be used to *introduce* the designational use of a term, it cannot *reinforce* that use.[114]

So 'Nana is our cat' can (in the above circumstances) be a grounding sentence, but the false grounding of 'Nana' in Jemima, coupled with a previously borrowed reference (of 'Nana') which was rightly grounded (in Nana), does not give rise to any referential indeterminacy in (10), according to Devitt. This claim strikes me as inconsistent with Devitt's earlier point about partial reference: in example (9), Devitt asserts that it is *partially true* that 'Nana is black' (when she is actually not black); but in example (10) Devitt maintains that is *simply false* that 'Nana is Siamese' (when she is actually not Siamese).

I have pointed out what I take to be three problems with Devitt's proposals. First, the sortal predicate requirement is too imprecise. Second, in certain circumstances, an ability to designate disappears, apparently without adequate explanation. Third, the general point Devitt is making in examples (10) (and (5) and (8)) is:

> [I]f a person says [...] '*a* is F' because *b* is F and he has come mistakenly to believe that *a* = *b*, he has said something simply true or simply false.[115]

Yet Devitt appears to break his own rule in claiming that (9) is *partially* true, where 'Nana' partially designates Nana and partially Jemima. In the next Part, I explain in

---

[114] Devitt (1981), p. 149.
[115] Devitt (1981), p. 149.

more detail the notion of partial reference, a notion which Devitt makes frequent use of in his account of how multiple groundings (and their d-chains) fix reference.

Devitt's notion of 'groundings' has differed from Putnam's notion of 'baptisms' in that groundings may recur, but it is not clear that Putnam's baptisms do so; grounding the same term in different entities may occur, but Putnam's Twin Earth example suggests that 'double-baptisms' are a form of semantic error; and a grounding can be false, but the first Putnamian baptism of a term probably cannot be. Devitt's CTR has not presented any adequate solutions to most of the problems highlighted in Putnam's CTR, but he has tried to describe how the reference relation between term and object changes, something which Putnam avoided. In a way not apparent in Putnam's CTR, Devitt's CTR explicitly allows for referential variance between the terms of successive scientific theories; yet allowing such meaning change does not stop Devitt opposing the semantic incomparability of successive theories. To secure comparability of the statements of successive theories, Devitt makes much use of the notion of partial reference:

> The theories we want to compare are ones 'in the same domain'. What sense can we make of being in the same domain? I suggest that we can make sense of it only in terms of shared partial referents.[116]

However, in Part 6, I argue that partial reference is not shown to be up to the task which Devitt sets it .

Part 6: Partial Reference

Michael Devitt allows the notion of partial reference a substantial role in his CTR. He admits that "reference may often be an idealization of partial reference"[117] and couches reference as a special case of partial reference: "for a term to have a full referent is for it to have only one partial referent."[118] Devitt takes the notion of partial reference from Hartry Field, and I will first briefly consider what Field has to say about partial reference. Then I will question if Devitt is wise to employ partial reference in his CTR; and I will cast doubt on the ability of Devitt's CTR to quash the semantic incomparability claim (claim (b)) of the IT.

---

[116] Devitt (1979), p. 45.
[117] Devitt (1981), p. 123.

Field illustrates his idea of partial reference with Newtonian examples. Newton made experimental claims like:

(11) "The mass of Object A is between 1.21 and 1.22 kilograms [said after putting Object A onto a pan balance and accurately weighing it]"[119]

If we take it that Newton was referring to proper mass *or* relativistic mass, statement (11) is true (under the conditions given) in either case. The same may be said of theoretical claims like:

(12) "To accelerate a body uniformly between any pair of different velocities, more force is required if the mass of the body is greater."[120]

Statements (11) and (12) were as true on Newton's lips as on Einstein's, but, claims Field, Newton was *not* simply referring to proper mass *or* relativistic mass. What Field wants to argue is:

> [T]here are sentences with perfectly determinate truth values which contain referentially indeterminate names and predicates, so that it makes perfectly good sense to ask whether the sentence is true or false even though it doesn't make sense to ask what the name really denotes or what the real extension of the predicate is.[121]

I will side with David Papineau in presenting one reason[122] why Field does not succeed.

As used nowadays, the term 'mass' is ambiguous and refers to proper mass and relativistic mass. Which, then, was Newton referring to? The answer Field gives is that there is no way to decide; for Newton's 'mass' did not completely refer either to proper mass or to relativistic mass. Nor did Newton's 'mass' refer ambiguously to both in the manner of our modern term 'mass'. For example, when Newton states that

(13) 'Momentum = (mass)v',

we are tempted to think that he is referring to relativistic mass. However, Newton would have maintained that the mass referred to in equation (13) is *in*variant. So Newton combined (13), an equation associated with relativistic mass, with a notion associated with proper mass. Likewise, when Newton states that

(14) mass is invariant,

---

[118] Devitt (1979), p. 43.

[119] Field (1973), p. 468.

[120] Field (1973), p. 470.

[121] Field (1973), p. 463.

[122] There are other reasons. In a short but detailed response to Field, John Earman and Arthur Fine (1977) argue that Newton referred to proper mass. They also challenge Field's claim that the term 'mass' is ambiguous in modern physics, maintaining

it might seem that he is referring to proper mass; but he combines (14) with the notion that the product of mass and velocity gives *momentum* (i.e. (13)), and this latter is a notion of *relativistic* mass. As causal theorists everywhere propound: that Newton had some false beliefs about relativistic mass and proper mass does not preclude him from referring to either of them. Field goes along with this view to some extent, but then gives it a little twist by claiming that "there is no basis for choosing between"[123] whether Newton meant proper mass or relativistic mass in any of his utterances. Why Field asserts this is as follows. The conjunction of statements (13) and (14) "was objectively false"[124], but the conjunction is false in the case that 'mass' refers to relativistic mass *or* in the case that it refers to proper mass. And "there are many of Newton's utterances containing the word 'mass' that we want to regard as true"[125], such as (11) and (12), but they are true in the case that 'mass' refers to relativistic mass *or* in the case that it refers to proper mass. So Field concludes that, for any of Newton's utterances containing 'mass' there is *a limited referential indeterminacy* - *"there is no fact of the matter* as to *which* of these quantities he was referring to."[126]

Field expresses such limited referential indeterminacy using the phrases 'partial reference' and 'partial denotation'[127] in relation to term tokens:

> I want to say that Newton's word 'mass' partially denoted proper mass and partially denoted relativistic mass; since it *partially* denoted *each* of them, it didn't *fully* (or determinately) denote *either*.[128]

To appreciate what Field means by the partial denotation of Newtonian 'mass' it is worth contrasting referential indeterminacy with ambiguity. A term is ambiguous if different tokens of it *fully* denote one of two different referents. So modern users of the term type 'mass' denote proper mass or relativistic mass, for modern tokens of 'mass' will refer to one of these two magnitudes. In this sense, the modern term *type* 'mass' is ambiguous: it partially denotes one of two *determinate* referents. As Field observes, ambiguity "does not demonstrate the existence of indeterminacy"[129]. Indeterminacy arises when "each *token* of 'mass' partially denote[s] two different

---

instead that, 'mass' refers only to proper mass. Earman and Fine did not go as far as to dismiss the "conceptual possibility" (Earman & Fine (1977), p. 536.) of partial reference, only that it did not apply to Newton's term 'mass'.

[123] Field (1973), p. 467. I have removed the italics.

[124] Field (1973), p. 468.

[125] Field (1973), p. 473.

[126] Field (1973), p. 467. I have added the italics to "which".

[127] Field does not use 'denotation' with Devitt's special nuance.

[128] Field (1973), p. 474.

[129] Field (1973), p. 475, n. 12.

quantities."[130] Newtonian tokens of 'mass' referred partially to both relativistic mass and proper mass, but not completely to one or the other (or both). Statements containing an ambiguous term are determinately true or false, depending on which meaning is taken; but statements like (13) or (14)) will be only partially true or partially false because of a term's referential indeterminacy.

Field's principle point about partial reference is not in the first instance an epistemological one. It is accepted that we use our best current scientific theory to *judge* what, if anything, Newton's 'mass' referred to. But the semantic and ontological point of partial reference addresses "not what scientists theory-dependently take to be the references of scientific terms [...], but what those references are."[131] Field's main point with partial reference is to describe in what *manner* scientific terms refer to those entities or magnitudes which are *in fact* there (from an external perspective).

Field's claim that

(j) statements (11) and (12) are determinately true, even as uttered by Newton, and even though Newtonian 'mass' only partially refers

appears to contradict other claims he makes about the development of scientific terms. That is, he maintains:

(k) "many of our *current* scientific terms are referentially indeterminate"[132]

and that science often progresses by a process of extensional refinement:

> the set of things that [... a scientific term] partially denoted after [a major change of theory] is a proper subset of the set of things it partially denoted before.[133]

So Field takes it that (11) and (12) are true in the case that 'mass' refers to proper or relativistic mass[134]. Yet it is precisely the references of our current scientific terms of 'proper' and 'relativistic mass' which are used by Field to assert the *determinate* truth of Newtonian statements such as (11) and (12), *even though Field also maintains that current scientific terms are themselves likely referentially indeterminate.* This ongoing shift in reference suggests to David Papineau that Field

---

[130] Field (1973), p. 475, n. 12.
[131] Pappineau (1979), p. 151.
[132] Field (1973), p. 480. My italics.
[133] Field (1973), p. 479.
[134] Field calls this an example of 'double refinement', since modern Physics has given two subsets of Newtonian 'mass'.

asserts, largely without grounds, that some of Newton's utterances containing 'mass' come out determinately true or false:

> [E]ven if we allow the idea of partial references, there is no reason to suppose that any statements will have anything other than indeterminate truth values, that is, be true according to some partial references and false according to others.[135]

The thrust of Papineau's criticism is not epistemological: Papineau is not criticising Field for saying that we *judge* the truth of Newton's statements according to the tenets of our own theory. Rather, Papineau criticises two of Field's *semantic* claims, namely (j) and (k); for if (k) alludes to the current scientific terms 'proper mass' and 'relativistic mass' (which it almost certainly does) then (j) and (k) are mutually inconsistent (under Field's own conditions).

I will make two criticisms of Devitt's use of partial reference. First, Devitt holds that many current scientific terms will undergo more refinement, so there are many previously and currently held scientific terms which only partially refer[136]. Devitt's holding these views then opens him up to the same criticism meted out to Field; for, like Field, Devitt also believes that there are scientific statements which are determinately or completely true, even when their terms only partially refer. Indeed, this claim is intended to support Devitt's argument against the semantic incomparability of theories:

> Given our present theory of reality we can (in principle) explain and justify [...] our intuitive judgement of the truth value of any past or present statement irrespective of any difference in meaning or reference between it and the statements of our present theory.[137]

Once again, then, Field's problem arises, that there is a dearth of current determinately referring terms with which to support the determinate truth of scientific statements like (11) and (12) , which we intuitively regard as determinately or true.

My second criticism is that the role Devitt affords to partial reference in his CTR is so great that a large number of scientific statements would have indeterminate truth values. Consequently, his CTR would not form a convincing argument against the IT's claim[138] that some, or even many, statements of successive theories are

---

[135] Pappineau (1979), p. 153.

[136] "It seems to me plausible that the terms of a scientific theory typically do partially refer." Devitt (1979), p. 44.

[137] Devitt (1979), p. 45.

[138] Against which I have already argued in Chapter 1. There, my simple argument does not draw on any theory of reference.

semantically incomparable. The nature of multiple groundings and d-chains, as Devitt relates them, is such that there is likely to be a preponderance of referential indeterminacy. He tells us that "a term partially refers to two different objects (sorts of objects) if the network underlying it is causally grounded in both."[139] In Devitt's theory, such mixed d-chains happen very easily.

Devitt's use of partial reference is not adequate for demolishing the semantic incomparability claim of the IT. He admits that "we lack a worked out theory of partial reference"[140], but I have not denied that reference can be partial. What Devitt lacks, it seems to me, is a theory of partial reference consistent with his other beliefs about the semantic relations between the terms and sentences of past and current successive scientific theories. A consequence of his account of partial reference and d-chains, (an account which he admits with disarming candour is not satisfactorily 'worked out') is that it does the opposite of what Devitt says it does: it lends support to the semantic *in*comparability claims of the IT (in that partial reference/referential indeterminacy of the terms of past *and* current theories entails that the statements of successive theories will be semantically incomparable).

Part 7: Cause and Description

While causal constraints are necessary, they are clearly not sufficient for an adequate CTR. A description associated with a NK term has a role in determining the reference of the term. Those CTRs which take Putnam's dictum that 'meanings just *ain't* in the head' and recast it as 'meanings ain't *just* in the head' are sometimes called 'causal-descriptive theories of reference' (CDTRs). What such theories try to do is explain how descriptions associated with a term help determine its reference. Here, I present an early version of Philip Kitcher's CDTR and consider some criticisms offered by Stathis Psillos. I conclude that not all of Psillos' criticisms of Kitcher are valid, but that one his criticisms of Kitcher's use of the principle of humanity at least raises concerns about Kitcher's CDTR. In the rest of Part 7, I address the brand of CDTR towards which Stathis Psillos is sympathetic and conclude that still no adequate theory of the reference of (theoretical) NK terms has been forthcoming.

---

[139] Devitt (1979), p. 44.

Like Putnam and Devitt, Philip Kitcher has it that a baptismal event, or set of dubbing events, is associated with a term type. When I utter a token of a term type, my utterance is "normally initiated by an event"[141] which is associated with the pertinent term type. This set of dubbing events helps make up what Kitcher calls the 'reference potential' of an expression type. Addressing some of the previously-stated objections to Putnam's CTR, Kitcher grabs the bull by the horns when he tells us that "terms whose reference potential contains two or more different initiating events [...] may reasonably be called *theory-laden.*"[142] He explains that the use of a term with such a reference potential

> depends upon hypotheses to the effect that the same entity is involved in the appropriate way in the different events which belong to the same reference potential.[143]

If one of these hypotheses begins to look doubtful, then "the use of the term which depends on it would have to be revised"[144]. So a reference potential with two or more initiating events will have two modes of reference, one designating (to retain Devitt's terminology) the causal agent which was present when named, and the other describing (or denoting, in Devitt's terminology) that causal agent.

Unlike Putnam, Kitcher would explicitly allow that *on some occasions* travellers to Twin Earth successfully refer to the water-like substance by using the term 'water', and on other occasions they refer to water by using 'water'. In allowing that 'water' can be a referentially ambiguous NK term, Kitcher's view is similar to that of Devitt's multiple groundings; so, in the case of Twin Earth, 'water' can be grounded in and designate H_2O *or* XYZ. Both Kitcher and Devitt also allow for referential and attributive modes of reference; but where they differ is in how they handle false groundings and false definite descriptions. That is, Devitt *disallows* that such descriptions as 'the H_2O in our glasses' can, in *general* use, refer to the water-like stuff on Twin Earth.[145] It was shown that Devitt[146] tries to legislate that denotation (attributive reference) is a mode of reference which, when the description is false, may not legitimately be passed on: at a face-to-face grounding, a false definite

---

[140] Devitt (1979), p. 44.
[141] Kitcher (1978), p. 540.
[142] Kitcher (1978), p. 540.
[143] Kitcher (1978), p. 540.
[144] Kitcher (1978), p. 540.
[145] As I have understood and presented him (in Part 5), Devitt allows the multiple grounding of the NK term 'water' (its designating H_2O and XYZ), but denies that false descriptions of NKs, no matter what their causal nexus with an entity, confer an ability to refer which can be passed along causal-historical chains in the language community.

description may *designate* the object present, but others who come to use the false description will fail to refer (in either sense of 'refer') to the object. I have argued that this move of Devitt's has an *ad hoc* quality.

Kitcher, on the other hand, recommends that we discern the causal-historical chains which link a token utterance *and* its description or content, to an initial baptism event. We replace, where necessary, the content of another's utterance token with our own content in accordance with the principle of humanity (see shortly). Having decided, say, that the term token does denote (refer attributively) to something (as opposed to nothing), we must then decide, from the context of utterance, what 'thing' caused (along causal-historical chains) the token utterance; that is, we must determine what the token designates (c.f. its referential use). On Twin Earth, as far as Kitcher is concerned, false definite descriptions *can* be passed on such that they still designate what they have failed to denote. Kitcher wants to say that there may be occasions (many more than Devitt allows) when not just 'water' but definite descriptions containing 'H_2O', though false, *do* refer to XYZ; for the principle of humanity is such that we may *understand* 'the H_2O in my glass' as meaning 'the XYZ in my glass'.

To determine the reference of any expression token, Kitcher considers the context in which the token is used and employs the principle of humanity. The principle of humanity is a hermeneutic device whereby we

> impute to the speaker whom we are trying to translate a "pattern of relations among beliefs, desires and the world ... as similar to ours as possible."[147]

To see how Kitcher's ideas work we will look at two of his examples; the first is a fictional narrative about a millionairess and the second is concerned with the reference of NK terms in phlogiston theory.

Eustacia Evergreen is a well known millionairess who wants to withdraw from the glare of public attention, so she employs an impersonator. The impersonator moves into a neighbourhood posing as Eustacia Evergreen while the millionairess Eustacia Evergreen leads a quiet life. Over time, neighbours get to know impersonator

---

[146] See Devitt (1981), pp. 148-9.

[147] Kitcher (1978), p. 534, quoting Richard Grandy (1973), 'Reference, Meaning and Belief', *Journal of Philosophy*, 70, pp. 439 – 452.

Eustacia and are led to believe that she is millionairess Eustacia. A neighbour promises his friend that he will take him to meet Eustacia Evergreen; but in so doing, is the neighbour referring to the impostor or the millionairess? For he *intends* to introduce his friend to a millionairess and yet *the person* he intends to introduce his friend to is not a millionairess. Kitcher's approach to this problem is to

> specify a set of entities (the pair set of the milionairess and the impostor) such that each token of 'Eustacia Evergreen' refers to one member of the set, even if, in the case of some referents, we are unable to decide which member is the referent.[148]

Kitcher concludes that the neighbour's *dominant intention* is to introduce his friend to a particular celebrity millionairess, not to an employee of that millionairess; so the neighbour was referring to the millionairess in the above use of the token[149]. However, the neighbour may use other tokens of 'Eustacia Evergreen' which refer instead to the impostor. So a neighbour might say, 'Yesterday, our rich and famous neighbour, Eustacia Evergreen, invited us to her next cocktail party.' Here, the neighbour's dominant intention is to refer to the woman who is actually the impostor. In each case, Kitcher advises me to use the principle of humanity to determine who I would have referred to were *I* the neighbour of Eustacia Evergreen's impostor.

In the Eustacia Evergreen story, there were two referents of 'Eustacia Evergreen' which actually existed, namely the millionairess and the impostor. The case of phlogiston theory differs from the previous example because there are not, in fact, two NKs of 'dephlogisticated air': at most there is only one,[150] and some tokens of 'dephlogisticated air' may lack a referent altogether. Kitcher therefore remarks that 'dephlogisticated air' cannot *always* be translated as 'oxygen' "because of a false presupposition, the idea that something is emitted in combustion, infects most of the terminology."[151] Since we cannot translate 'dephlogisticated air', Kitcher suggests that we instead 'disentangle' it. By this he means simply that we should *sometimes* translate 'dephlogisticated air' as 'oxygen', and sometimes not, depending on context.

---

[148] Kitcher (1978), p. 527. In spite of this passing allusion to partial reference, none of Kitcher's examples is said to demonstrate limited referential indeterminacy. He writes: "I have been unable to find a convincing example from the history of science which would demand the use of Field's apparatus of partial reference." Kitcher (1978), p. 546, n. 33. However Stanford and Kitcher (2000), p. 119, do come up with just such an example.

[149] Kitcher (1978), p. 527-8.

[150] In fact, 'dephlogisticated air' may on occasions have had referred to NKs other than oxygen; but my point is that, in Kitcher's view, at least one of the tokens of 'dephlogisticated air' fails to refer, and in this respect the example of 'dephlogisticated air' differs from the Eustacia Evergreen example.

Georg Stahl had first named as 'phlogiston' the substance which is given off into the air during combustion. For example, when wood burns in a sealed container, the common air in the container becomes '*phlogisticated* air'. Common air therefore has the capacity to take up or absorb a certain amount of phlogiston during the process of combustion. The process of smelting, turning some metal oxides to metals by heating them, was explained by Stahl as the emission of phlogiston from the charcoal to the ore. Ore was said to lack phlogiston, and when the ore's dearth of phlogiston is remedied by the hot charcoal, the ore becomes a metal. When Joseph Priestly, mirroring the smelting process, heated mercuric oxide with a *lens*, he obtained a pure metal, mercury, and a different gas (or 'air'). Priestly concluded that, in its conversion to mercury, the warmed mercury calx, being phlogistonless, had absorbed phlogiston from the air, leaving the air *dephlogisticated.* Priestly found confirmation that such air was *de*phlogisticated because objects burned more fiercely (more readily discharged phlogiston) when placed in it. The air (or gas) obtained in Priestley's experiment, the air described as 'dephlogisticated', was in fact oxygen. This ability to use 'dephlogisticated air' to refer to the gas obtained by heating the red calx of mercury was passed on throughout the community of phlogistonists and beyond. But tokens of 'dephlogisticated air' do not always refer to oxygen, Kitcher asserts, because:

> His [i.e. Priestley's] later utterances could be initiated either by the event in which Stahl fixed the referent of 'phlogiston' or by events of quite a different sort, to wit, encounters with oxygen.[152]

So the reference potential of the term *type* 'dephlogisticated air' consisted of a set of at least two ordered pairs: the Stahlian {dephlogisticated air_1, the air which has phlogiston removed from it}, and Priestley's introduction {dephlogisticated air_2, the air obtained by heating mercuric oxide}. To determine the reference of any given *token* of 'dephlogisticated air', Kitcher recommends the employment of the principle of humanity.

For Kitcher, the principle of humanity is an essential presupposition underlying successful attempts at cross-theory communication. When a term type is common to two different theories, there will likely be a difference between the respective reference potentials. In the effort to understand each other, persons from each theoretical viewpoint "will endeavor to formulate hypotheses about the referents of

[151] Kitcher (1978), p. 531.
[152] Kitcher (1978), p. 537.

their rivals' tokens which will explain their rivals' linguistic behavior."[153] Doing this, claims Kitcher, it becomes quite clear when Priestley is referring to oxygen, and when 'dephlogisticated air' fails to refer at all.

Stathis Psillos objects to Kitcher's use of the principle of humanity because it splits Priestley's intentions from his beliefs. Priestley believed that dephlogisticated air_1 = dephlogisticated air_2 and his intention was to refer to the object which satisfied this identity. The principle of humanity distorts Priestley's situation:

> The principle of humanity establishes an incoherence between the subject's beliefs and intentions (that is, an incoherence in the subject's own perception of the situation he was in) in order to maximise coherence in our judgements of what our subject was doing in light of our knowledge of the situation he was in.[154]

The principle of humanity wants to make Priestley's beliefs and assertions as like a modern chemist's as possible. For the modern chemist employing Kitcher's theory of reference, 'dephlogisticated air' is an ambiguous expression type referring either to oxygen or the empty set. But as far as Priestley was concerned, 'dephlogisticated air' was *not* an ambiguous term! Priestley would have been happy to assert:

(15) 'Dephlogisticated air exists and oxygen does not'.

But under some token interpretations, the Kitcherian modern chemist will take a token of 'dephlogisticated air' to mean oxygen, and this would be to interpret Priestley's happy assertion (15) as inconsistent. The problem, then, is that under *Priestley's* interpretation, (15) is *not* inconsistent. Inconsistency is not one of the major flaws of phlogiston theory; so to say that Priestley's phlogiston theory implies inconsistent statements such as (15) is to misrepresent Priestley's phlogiston theory.

An undesirable general consequence of Kitcher's combining the principle of humanity with that of reference potential is that, according to Psillos:

> the principle of humanity makes referential continuity too easily available: arguably all past abandoned expression-types end up having referential tokens [... so that] no abandoned concept has failed to characterise some natural kind we now posit.[155]

However, it seems to me that the problem here is not the principle of humanity but the problem common to so many CTRs, namely, that reference comes cheap when it comes causally. What the principle of humanity provides is content, but it is the

---

[153] ibid., p. 541.

[154] Psillos (1997), p. 265.

causal nature of Kitcher's CTR which will allow the reference a NK term token to be traced back to some causal agent; the principle of humanity, on the other hand, will provide the current view about what that causal agent could reasonably be described as being. Applying Kitcher's theory of reference to *Stahl's* phlogiston theory, we might say that Stahl was one of the first to describe systematically the reversible relationship between some metals and their calces (or oxides). The principle of humanity, combined with a CTR, warrants the inference that whatever would cause Stahl to say, 'Heat mercury in air and it gives off phlogiston, thereby becoming a calx' is what would cause a modern chemist to say, 'Heat mercury in air and the mercury oxidises'. Ergo, for certain metals 'to give off phlogiston' means (or entails) that they will absorb oxygen. What must be removed from iron to obtain iron oxide? Stahl's answer: phlogiston; but the Modern chemist's answer: nothing. As far as I can see, then, Kitcher's use of the principle of humanity is a narrower constraint than the purely causal one, and leads to the conclusion that some, if not many, of Stahl's (and, for that matter, Priestley's) tokens of 'phlogiston' failed to refer.[156] So I do not think that this particular criticism which Psillos levels at Kitcher's theory is convincing.

Stathis Psillos is prepared to accept that the expression type 'dephlogisticated air' can refer to oxygen, yet he maintains that virtually none of Priestley's tokens of 'dephlogisticated air' did so. Priestley did refer to oxygen, maintains Psillos, but only by using demonstrative pronouns and deictics, "saying the likes of '*This* air (or this stuff) makes me feel so light.'"[157] For Psillos, "[t]he shift from 'this air' or 'this stuff' to 'dephlogisticated air' is crucial. The fact that the former may refer does not entail that the latter refers too."[158] Here, Psillos is placing restrictions on multiple or false groundings which are even narrower than Michael Devitt's. When Priestley made utterances like:

(16) I call the stuff which makes me feel so light 'dephlogisticated air'

(17) I call the stuff which makes me feel so light 'air with phlogiston removed'

Devitt would probably[159] say that in utterances like (16) and (17), Priestley, having already grounded in the air obtained by heating mercury oxide the expressions 'dephlogisticated air' and 'air with phlogiston removed', may use those expressions

---

[155] Psillos (1997), p. 269.

[156] Matters are different for tokens of 'dephlogisticated air' because of the different reference potentials.

[157] Psillos (1997), p. 268.

[158] Psillos (1997), p. 268.

[159] Priestly does seem to have met the three conditions which Devitt lays down for a grounding to occur (see Part 5).

in sentences such as (16) and (17) to refer to oxygen. Devitt would also hold that the referential *ability* first made in the grounding of the expression 'dephlogisticated air' in oxygen can be passed on, but that it can only passed on to those who would *not* assent to (17) at the same time as (16). (Since Priestley himself would have assented to (16) and (17), Devitt would describe Priestley's grounding of 'dephlogisticated air' in oxygen as a *false* grounding.) Psillos maintains that Priestley probably *never* grounds 'dephlogisticated air' in oxygen (and so never passed on the reference ability). So Psillos is stricter than Devitt when it comes to giving adequacy conditions for a grounding to take place, though Psillos and Devitt have similar views about the passing on of the benefits of such a grounding. Kitcher is the least strict of the three, concerning groundings and the passing-on of abilities to designate. The positions of Psillos, Devitt and Kitcher concerning Priestley's utterance tokens containing 'dephlogisticated air' are:

(f) None of Priestley's tokens of 'dephlogisticated air' referred to oxygen. (Psillos)

(g) Some of Priestley's tokens of 'dephlogisticated air' referred to oxygen, but Priestley's occasional ability to refer to oxygen using tokens of 'dephlogisticated air' could not be passed on to other phlogistonists in d-chains. (Devitt)

(h) Some of Priestley's tokens of 'dephlogisticated air' referred to oxygen, and Priestley's grounding of 'dephlogisticated air' in oxygen was the grounding back to which can be traced some other phlogistonists' ability to refer to the air obtained by heating mercury oxide (oxygen) using tokens of 'dephlogisticated air'. (Kitcher)

What motivates Stathis Psillos' position (f) is the view that merely being in causal contact with a member of a NK is not sufficient to be able to refer to it (other than deictically); one must also, as an effect of such causal contact, be able to give a core causal description of the (putative) entity referred to (or be causally-historically related to one who can). The core causal description associated with NK term will include a description of the NK's kind-constitutive properties and a description of the causal roles of the kind-constitutive properties. If the core causal description is false (if there is nothing which satisfies it) then its associated term fails to refer. So Psillos' causal-descriptivist view of the term 'phlogiston' used by phlogistonists is that:

> [N]one of the properties of oxygen were the causal origin of the information they [the phlogistonists] had associated with phlogiston. And nothing in nature could possibly be the causal origin of such information. What it is correct to say is that 'phlogiston' refers to nothing.[160]

[160] Psillos (1999), p. 291.

And Psillos concludes that since 'phlogiston' refers to nothing, then 'dephlogisticated air', inasmuch as it refers to common air with phlogiston removed, also refers to nothing. I will next offer three criticisms of these views.

The first is that Psillos' view is surely too strict and entails a counter-intuitive view of the relation between the statements of phlogiston and oxygen theories. Psillos maintains:

> We can still understand and explain Priestley's assertions that dephlogisticated air supports combustion better than ordinary air and that dephlogisticated air is phlogiston-free, even if we admit that both of them are false (as I think we should).[161]

However, it seems remarkable to me that *all* Priestley's statements containing 'dephlogisticated air' were false. If the expression *type* 'dephlogisticated air' includes oxygen in its reference potential, as Psillos maintains, then it must have been grounded in oxygen at some stage – a fact also acknowledged by Psillos. If we go along with Stathis Psillos' general argument, then we will conclude that the person who first had the ability to refer to oxygen using the term token 'dephlogisticated air' could *not* have been a phlogistonist; for only such a person has the possibility of being free of the erroneous identity belief that dephlogisticated air_1 = dephlogisticated air_2 (or of holding (16) and (17)). Psillos' position seems to be that oxygenists can refer to oxygen using 'dephlogisticated air' as much as they like: phlogistonists can't at all.[162] But does it not seem odd – and counter to our intuitions about the relation between the languages of scientific theories - that a phlogistonist can give what is, to a modern chemist, a recognisable (and systematic though, admittedly, false) account of the relation between a metal and its oxide without *ever* referring to oxygen?

A second criticism concerns the appeal to kind-constitutive properties in the core causal description. "The appeal to kind-constitutive properties is essential" Psillos tells us, "because it is *these* properties which, ultimately, fix the reference of the term."[163] These properties pertain to the 'internal structure' of members of the NK:

> [P]ositing a natural kind with a certain internal structure should be tied to a description of its properties – a description, that is, of what this internal

[161] Psillos (1997), p. 267.

[162] Psillos concedes that phlogistonists may on occasions have referred to oxygen using tokens of 'dephlogisticated air', but that they would have done so "only accidentally" (Psillos (1997), p. 268).

[163] Psillos (1999), p. 295.

> structure is – in such a way that if there is no kind which has these properties, then we may have to just admit that a word which was taken to refer to this kind does not, after all, refer.[164]

In Part 4, criticism was made of the microstructural criterion of substance identity, and it seems to me that Psillos' view of NKs, and of the co-reference of NK terms, is open to the same criticism.[165]

Finally, the third criticism - or set of criticisms - is directed at the claim that the core causal description determines the reference of NK terms. Advocates of the CDTR have replies to these criticisms, but I go on to say why I do not find the replies completely convincing.

At first glance, it may seem that the CDTR may be criticised on the same grounds as the Description Theory of reference, i.e. that one cannot give a description of properties which are necessary *and* sufficient for an entity to be member of NK, and that without such a core causal description, the reference of the associated term cannot be determined. The CDTR side-steps this criticism by denying that the core causal description is or need be a description of the properties necessary and sufficient to be a member of a NK. The details are as follows. In the early stages of scientific enquiry, a detailed description of a postulated entity's constituents and causal role is neither available nor necessary for the enquiry to proceed. In due course, descriptions may be *added* to the core causal description, but if the core causal description is falsified, then the associated term will fail to refer. (Some parts of the *general* (theoretical) description associated with a term may be dropped altogether without change of reference, but the core causal description can only be added to.) The scientific enquiry is committed to holding the core causal description as true for as long as it is *found* to be so; so the term associated with the core causal description is *held* to refer for as long as the line of enquiry continues. The longer the enquiry, the more evidence there is that the term associated with the core causal description does in *fact* refer. If or when a description of the properties necessary and sufficient to be a member of a NK does become available, it will have been found

---

[164] Psillos (1999), p. 287.

[165] Psillos believes that such criticism can be satisfactorily dealt with (as Psillos (1999), p. 313, n. 4 makes plain), but he doesn't have sufficient space for the details.

that the core causal description was *satisfied* by the referent, and was *about* the referent, all along.[166]

To the objection that, particularly in the early stages of enquiry, the core causal description may be satisfied by a number of different entities not all of which are directly relevant to the enquiry, it can be replied that only the entity or entities which are the *causal source* of the core causal description and which satisfy it are referred to. Further investigation will refine the core causal description by adding quantifiers so that eventually the core causal description will refer to only one entity.[167]

It is certainly possible to criticise the CDTR's above responses. For example, no clear criteria are given for distinguishing what would be part of the core causal description and what would be part of the full theoretical description. This means that if part of what was thought to be the core causal description were falsified, the falsified part could be relegated to the full description and the rest of the core causal description could be retained. It would then be very unclear when to regard a theoretical term as having a referent.

Furthermore, there is a problem with Psillos' claims about reference when he says that, in the process of enquiry, scientists "do refer to the (putative) entity which satisfies the core causal description."[168] The problem is that one can only *refer* to putative entities which actually do exist; that a *putative* entity satisfies a given description is not sufficient for it to exist (unicorns being a case in point). So I think that Psillos would be better to say that, in the process of enquiry, scientists 'talk about'[169] the putative entity which satisfies the core causal description. But then he will not have a theory of reference, only a theory of 'talking about' things. In Part 8, this kind of criticism is enlarged and applied to all the CTRs and CDTRs considered in this chapter.

---

[166] Psillos admits that "the whole idea of the specification of a core description involves an element of *rational reconstruction* of the actual problem situation in which an entity was originally posited." (Psillos (1999), p. 297.)

[167] "[I]t is perfectly possible that a theoretical term begins its life as part of some abstract speculations about the causes of a set of phenomena, and subsequently becomes part of a rather firm theory which associates with it a core causal description." (Psillos (1999), p. 299.

[168] Psillos (1999), p. 295.

The conclusion of Part 7, indeed of Parts 1 to 7, is that the various causal theories of reference here considered do not satisfactorily account for the reference of NK and physical magnitude terms. Since none of the causal or causal-descriptive theories of reference here is judged sufficient to sustain externalist semantic claims, Feyerabend's *thesis I,* stated as:

(a) *thesis I*: "the interpretation of an observation language is determined by the theories which we use to explain what we observe, and it changes as soon as the theories change." [170]

has little to fear from those quarters. (In fact, the causal-descriptivist trend, with its talk of 'reference-determining core causal descriptions' and "overdetermination of reference by theory"[171] might be construed as something of a move in Feyerabend's direction). And since *thesis I* is what motivates the claim that when theory changes, the references of terms may change, the meaning variance thesis is not convincingly challenged by these causal theories of reference; in particular, the claim:

(b) statements of T_1 and T_2 may be logically independent in the common domain

is not refuted by these causal theories of reference because they are not adequate theories of reference.

Should the conclusion - that the CTRs (including CDTRs) in this chapter are inadequate - be regarded as insufficiently supported (particularly by the criticism directed at Kitcher and Psillos here in Part 7), all is not lost. In Part 8 I present an argument for why substantive accounts of reference, such as all those considered here in Chapter 3, constitute an *un*acceptable argument against claim (b). Part 8 will also discuss briefly what relevance the rejection of substantive theories of reference has to the IT.

Part 8: The Flight To Reference and the IT

Here I present Michael Bishop's and Stephen Stich's argument against the strategy of using substantive theories of reference for resolving philosophical issues which are not merely about reference. Bishop and Stich label this strategy 'the flight to reference'. A substantive theory of reference is one "that takes reference to be some sort of complex relationship between referring terms and entities or classes of

[169] The distinction between 'referring to' something and 'talking about' something is discussed by Rorty (1980), pp. 289-91.
[170] Feyerabend (1958), in PP1, p. 31. Italics removed.

entities in the world."[172] So Bishop and Stich have in their sights all the CTRs and CDTRs considered in this chapter, as well as many description theories of reference.

The flight to reference is used by Putnam[173], Devitt,[174] Kitcher[175] and Psillos[176] to argue that the logical independence claim of the meaning variance thesis, claim (b), is false. Their argument assumes something like the following form. First, they use a general account of reference to determine whether or not a scientific term refers, and if so, to what. In the latter case, *continuity* of reference between the terms of successive theories (in their common domain) will be asserted; in the former case there will be *discontinuity* of reference. Any past scientific statement is false which predicates a term which has been deemed referentially discontinuous with the terms of our current theory in the appropriate domain; and all the other past scientific statements (i.e. those whose terms *do* refer) will be either (approximately) true or else false. With the truth values of the statements of *all* scientific theories thus determined (under the interpretation of our current scientific theories) none of the statements of successive scientific theories is logically (or semantically) independent in the common domain: claim (b) is then false.

What is wrong with the flight to reference is that it begins by proposing a theory of reference specifying "an empirical relation [or relations] that must obtain for terms of a certain kind to refer to things in the world"[177]; then it applies this theory to some NK or physical magnitude or theoretical term; but its conclusion "is explicitly about truth or ontology or some matter."[178] In this kind of argument there is a "fatal gap"[179] between premises (about reference) and conclusion (about other things).

The case of 'phlogiston' provides an illustration. Let us say that on a particular occasion, 'o', Priestley uses the term token 'dephlogisticated air' with the intention of

---

[171] Kroon (1985), p. 148.
[172] Bishop, Michael A. & Stich, Stephen P. (1998), p. 34. Deflationary accounts of reference are therefore not targeted.
[173] See Putnam (1975). 'How not to talk about meaning' & 'Explanation and Reference'.
[174] See Devitt (1979).
[175] See Kitcher (1982).
[176] See Psillos (1999), pp. 280-1.
[177] Bishop & Stich (1998), p. 38.
[178] Bishop & Stich (1998), pp. 34-5.
[179] Bishop & Stich (1998), p. 35.

referring to a substance he has produced in a certain experiment. Kitcher's argument begins with the following premises:[180]

(18) On occasion o, Priestly uttered, "Dephlogisticated air supports combustion better than ordinary air."

(19) On occasion o, 'dephlogisticated air' refers to oxygen.

(20) Oxygen supports combustion better than ordinary air.

And ends with the conclusion:

(21) On occasion o, Priestley's utterance, 'Dephlogisticated air supports combustion better than ordinary air' is true.

For (21) to follow from the above premises, a further premise connecting the reference and truth of utterances is required, such as:

(22) An utterance of the form 'Fa' is true iff (Ex) (this token of 'a' refers to x and x satisfies this token of 'F_').

There is historical support for (18), current chemistry supports (20), and Kitcher's theory of reference supports (19). But what support does Kitcher offer (22)? The common response would be to say that (22) must be true, for "No account of the reference relation that failed to make [22] true could possibly be correct."[181] And Bishop and Stich would agree. So the assumption is that Kitcher's account of the reference relation makes (22) come out true.

But this assumption is problematic because other accounts of reference which are *inconsistent* with Kitcher's, *also* want to assume that the reference relations they specify make (22) come out true. Psillos, for example, makes the supported claims:

(23) On occasion o, Priestly uttered, "Dephlogisticated air supports combustion better than ordinary air."

(24) On occasion o, 'dephlogisticated air' fails to refer.

And concludes:

(25) On occasion o, Priestley's utterance, 'Dephlogisticated air supports combustion better than ordinary air' is false.

So Psillos also must assume (22) and that his theory of reference makes (22) come out true. But Kitcher and Psillos can't both be right[182]! Unless one of them demonstrates that his own account of reference supports the truth of (22), neither

---

[180] Statements (17), (18), (19), (20) and (21) are quoted from the example in Bishop and Stich (1998), p. 44.

[181] Bishop & Stich (1998), p. 45.

[182] And, of course, this does *not* show that *both* Kitcher and Psillos are wrong. The (alleged) problem common to both is stated at the end of this paragraph.

Psillos nor Kitcher has "grounds for claiming that the complex substantive relation he describes really is reference."[183] Each of them gives a theory describing the relation between tokens of 'dephlogisticated air' and oxygen; but neither is *entitled* to claim that the relation described is *reference*, where (22) is constitutive of reference, unless it is *shown* that the theory describing the relation makes (22) come out true.

None of those who attempt the flight to reference as a line of attack against MVT claim (b) - neither Putnam, Devitt, Kitcher, nor Psillos - *shows* that the relation he describes in his 'theory of reference' is constituted by (22). Since each fails in this way to *demonstrate* that he is presenting a theory of *reference*, each fails to offer arguments against meaning variance claim (b).

At the end of the day, the theories of reference considered in this chapter seem to me to be inappropriate to the problems posed by Feyerabend's semantic IT. The externalism of these CTRs brooks little understanding of the internalist strain in the IT. As John Preston puts it:

> Attempts to compare incommensurable theories on the basis of the concept of reference ... ignore the fact that reliance on conceptually unmediated word/world relations (such as reference) is not in keeping with Feyerabend's philosophy.[184]

The second problem with these CTRs (in the context of the IT) is that they presuppose a strict realism, whereas what the semantic IT seems to imply is a less strict – or what John Dupre calls 'promiscuous' – realism: "the claim that there are many equally legitimate ways of dividing the world into kinds"[185]. Feyerabend, in his later writings, articulates this position – which he calls 'cosmological' or 'ontological relativism' as follows:

> Scientists [...] are sculptors of reality. That sounds like the strong programme of the sociology of science except that sculptors are restricted by the properties of the material they use [...] What we find when living, experimenting doing research is therefore not a single scenario called 'the world' or 'being' or 'reality' but a variety of responses, each of them constituting a special (and not always well-defined) reality for those who have called it forth. This is relativism because the type of reality encountered depends on the approach taken. However, it differs from the philosophical doctrine by admitting failure.[186]

---

[183] Bishop & Stich (1998), p. 46, n. 8.

[184] Preston (1997), p. 217 n. 14. Oberheim & Hoynengen-Huene (1997) make even stronger remarks to this effect.

[185] Dupre (1993), p. 6.

[186] Feyerabend (1993), pp. 269-70.

So it seems to me that two kinds of semantic theory would offer better analyses or criticism of the semantic IT. The first kind would be one that does not *build* upon an externalist view of reference. This approach is taken by Donald Davidson, whose views will be considered in the next chapter. The other kind of semantic theory would be one which offers a plausible explication of the notions of 'truth' and 'reference' consistent with the ontological pluralism which lies behind the IT.

Part 9: A Kind of Incommensurability

As Martin Carrier observes, "it is the incompatibility of theoretical premises that generates incommensurability in the first place"[187]. Theoretical premises motivate the classifying of entities into natural kinds by being part of theories which help us explain the observed common properties and dispositions of entities. A realist view is that, for the classification to be true, the classification sorts the kinds by their essences. Taxonomic realism is the claim:

> there is one unambiguously correct taxonomic theory. At each taxonomic level there will be clear-cut and universally applicable criteria – essential properties, let us say – that generate an exhaustive partition of individuals into taxa.[188]

Given the difficulties which such a view engenders (earlier parts of this chapter have shown some of the difficulties taxonomic realism created for the CTR) one begins to wonder if natural kinds *have* such essential properties. The meanings of NK terms may not then be constrained by completely nonepistemic essences. However, it will not be my goal in this brief Part 9 to develop or argue for this proposal. Instead, I will limit discussion to incommensurability as a phenomenon which may arise when different theories classify NKs differently. Such a situation describes epistemic rather than semantic incommensurability, but I think that the epistemic model points out the direction for those more brave and more able.

Taxonomic incommensurability is a notion associated with the writings of Thomas Kuhn, but some of Feyerabend's comments from the mid 'sixties make strikingly similar proposals. Feyerabend tells us that incommensurability occurs when "rules according to which objects or events are collected into classes"[189] undergo changes

---

[187] Carrier, in Hoyningen-Huene & Sankey (eds.) (2001), pp. 78-9. Carrier's view here is in line with my exposition of the IT in Chapter 1.
[188] Dupre (1993), p. 27.
[189] Feyerabend (1965b), p. 268.

such that "a new theory entails that all the concepts of the preceding theory have extension zero"[190], or that the new theory:

> introduces rules which cannot be interpreted as attributing specific properties to objects within already existing classes, but which change the system of classes itself.[191]

By way of comparison with Feyerabend, Thomas Kuhn describes what he calls 'the principle of no overlap' as the condition that, in the transition from T_1 and T_2 "no two kind-terms [...] may overlap in their referents unless they are related as species to genus."[192] When the principle of no overlap is not adhered to, when "the kind in the old science cannot be a kind in the new science"[193], then T_1 and T_2 are incommensurable theories.

Taxonomic incommensurability comes close to meeting the four main elements of the MVT. First, *thesis I*: taxonomic incommensurability considers classification not reference, and this gives taxonomic incommensurability a semantic internalist flavour. Second, inconsistency: one theory's system of classification precludes the other's. Thirdly, meaning variance: the intension, and possibly extension, of one or more NK terms will have changed.

The fourth element of the MVT is the logical independence claim. Clearly if T_2's taxonomy is inconsistent with T_1's, the truth of T_1 is not independent of that of T_2. One way round this is to drop the inconsistency claim and therefore the case that the theories are competitors. This is what John Dupre seems to do when he remarks:

> one reason that scientific narratives constructed for different purposes will be incommensurable is that they need to be told in terms of noncoincident kinds.[194]

The incommensurability which Dupre describes is something like the taxonomic incommensurability described above. What gives rise to such different, indeed incommensurable (though not inconsistent and not competing), systems of classification, suggests Dupre, is the science which forms and uses the theory, for "a system of classification is typically an inextricable part of the science to which it

---

[190] Feyerabend (1965b), p. 268.
[191] Feyerabend (1965b), p. 268.
[192] Kuhn, quoted by Carrier in Hoyningen-Huene & Sankey (eds.) (2001), p. 70.
[193] Hacking (1993), p. 295.
[194] Dupre (1993), p. 112.

applies."[195] However, Dupre's view does not fit our paradigm case of incommensurable theories one of which succeeds the other.

To retain the inconsistency element of the MVT, an alternative to logical independence will need to be found. Ian Hacking suggests a kind of pragmatic independence. In this case, one who holds T_2 will "no longer speculate, conjecture, predict, explain, and most importantly, work on the world using the old classifications."[196] The classification of the predecessor theory T_1 can be *understood*, but it no longer suits the purposes of, or makes connexions considered pertinent to, those who have accepted and use the successor. Hacking gives the example of reading a text by Paracelsus. Those competent in Latin or four-hundred-year-old German will be able to translate Paracelsus' words, but his whole style of reasoning and the connexions he makes, are alien to the modern reader; even a Paracelsus expert.

The ideas presented here in Part 9 are tentative and do not even constitute a semantic construal of the IT. The aim has been to sketch a direction to go in to arrive at a satisfactory understanding of the semantic incommensurability.

[195] Dupre (1993), p. 103.
[196] Hacking (1993), p. 295.

Chapter 4

Conceptual Schemes, Translation, and The Meaning Variance Thesis

The world, our world, is depleted, impoverished enough. Away with all duplicates of it, until we again experience more immediately what we have.

Susan Sontag, 'Against Interpretation', p. 6.

To see language in the same way as we see beliefs – not as a 'conceptual framework' but as the causal interaction with the environment described by the field linguist, makes it impossible to think of language as something which may or may not (how could we ever tell?) 'fit the world'. So once we give up *tertia*, we give up (or trivialize) the notions of representation and correspondence, and thereby give up the possibility of formulating epistemological skepticism.

Richard Rorty, 'Pragmatism, Davidson and Truth', in Lepore (ed.) (1986), p. 345.

Introduction

I begin with an overview of Donald Davidson's 'On the Very Idea of a Conceptual Scheme', where he offers arguments against conceptual schemes, arguments which he regards as refuting the IT. In Part 2, I look in more detail at Davidson's views on truth and interpretation which inform the arguments in Part 1. In Parts 3 to 7, I consider objections to Davidson's arguments, but conclude that none of those presented convincingly tackles head on what Davidson proposes. In Part 8, I conclude that Davidson's argument against untranslatability is successful, and that the principle of charity (POC) convincingly opposes and undermines the logical independence claim of the MVT. I also conclude, however, that while Davidson's argument against the possibility that there are untranslatable natural languages holds, his claim to have shown that the very idea of a conceptual scheme is problematic does not succeed. I conclude the chapter with a Davidsonian account of something like semantic incommensurability drawn from Bjørn T. Ramberg.

Part 1: Overview

Donald Davidson has nothing against groups of propositional attitudes, such as beliefs. Nor would he object to the claim that a sentence can describe a state of affairs. What Davidson does not like about conceptual schemes is that they adopt a *staging* role: they present states of affairs in a certain light, from one angle or another; and they take what we are presented with – the world – and *re*-present it. Davidson's trivial point is that a true sentence describes a state of affairs and a false one does not. There is no need to dramatise a state of affairs as something other than what it is. Instead, Davidson recommends that we "re-establish unmediated touch with [...] familiar objects"[1]. Davidson makes clear what he is attacking in his essay 'On the Very Idea of a Conceptual Scheme':

> My target was the idea that on the one hand we have our world picture, consisting of the totality of our beliefs, and on the other hand we have an unconceptualized empirical input which provides the evidence for and content of our empirical beliefs.[2]

Davidson's first step in his attack on the very idea of a conceptual scheme is to lay out the common view of conceptual schemes as being the *tertium quid* of languages. If two people speak the same language, then they share the same conceptual scheme.

[1] Davidson (1984), p. 198.
[2] Davidson (1999), p. 105.

If two people speak different languages then either they share the same conceptual scheme or not. Their sharing the same conceptual scheme entails that they 'divide up' or 'organise' the world and its contents the same; their having different conceptual schemes entails that one of them does these things differently to the other. Davidson's second step is to argue that these common views about dividing up or organising the world are spurious. So he concludes that the idea of conceptual scheme is a bogus one; and this leaves us with language and the world. Giving the details of these three steps takes up practically all of this chapter.

Davidson further maps out the purported relations between conceptual schemes and languages, claiming that, when two different language speakers share the same conceptual scheme, translation between the two languages is possible; when they share different conceptual schemes, it is not. It is the case of untranslatability which Davidson is most interested in:

(1) "two people have different conceptual schemes if they speak languages that fail of intertranslatability".[3]

If Davidson can refute the claim that there *are* different conceptual schemes then, according to his map of the territory, he also refutes mutual untranslatability. And mutual untranslatability is equivalent to incommensurability, for, by Davidson's reckoning:

(2) "'Incommensurable' is, of course, Kuhn and Feyerabend's word for 'not intertranslatable'."[4]

However, Davidson's argument against mutual untranslatability (and incommensurability) is not quite the modus tollens form given above. Statement (1) claims that two different conceptual schemes are necessary for the mutual untranslatability of two languages. Yet Davidson also holds the converse true:

(3) "[t]he failure of intertranslatability is a necessary condition for difference of conceptual schemes."[5]

It seems fair to adduce that Davidson has in mind an equivalence relation between different conceptual schemes and languages which fail of intertranslatability.

[3] Davidson (1984), p. 185.
[4] Davidson (1984), p. 190.
[5] Davidson (1984), p. 190.

Davidson then argues that the roles of organising and arranging attributed to two different conceptual schemes *preclude* their mutual untranslatability and support their mutual intertranslatability. Additionally, Davidson advances what he regards as a refutation of mutually untranslatable natural languages, thereby seeking to rubbish the notions of different conceptual schemes *and* of incommensurable languages/schemes.

The organising (or classifying) role of a conceptual scheme is embodied in "the referential apparatus of language – predicates, quantifiers, variables, and singular terms"[6]. Two different conceptual schemes would classify differently the actual objects in the world such that each conceptual scheme language (L_1 and L_2) would have predicates with no common extensions. Davidson finds such a claim unconvincing because a "language that organizes *such* entities must be a language very like our own."[7] Support for this rebuttal comes in Davidson's use of the principle of charity (POC) to be explained later.

Actually, it is Davidson's comments about the *fitting* role which support most of his argument against total untranslatability. This is because the fitting role of conceptual schemes is concerned with sentences, whereas the organising role is concerned with sentence parts; and Davidson cautions us

> not to suggest that individual words must have meanings at all, in any sense that transcends the fact that they have a systematic effect on the meanings of sentences in which they occur.[8]

Given Davidson's 'top-down' view of the compositionality of meaning, the argument about the referential apparatus of language (and the organising role) defers to that of the truth of sentences (and the fitting role).

When two different conceptual schemes (mostly) fit the facts then (most of) the scheme sentences of L_1 and L_2 are true. Davidson sees no need for the fitting role because he sees no need for this view of truth:

> The trouble is that the notion of fitting the totality of experience, like the notion of fitting the facts, or of being true to the facts, adds nothing intelligible to the simple concept of being true.[9]

---

[6] Davidson (1984), p. 193.
[7] Davidson (1984), p. 192.
[8] Davidson (1984), p. 18.
[9] Davidson (1984), pp., 193-4.

Instead, Davidson proposes:

(4) "The sentence 'My skin is warm' is true if and only if my skin is warm. Here there is no reference to a fact, a world, an experience, or a piece of evidence."[10]

As will become apparent, the establishment and use of this premise (4) will be central in Davidson's arguments against conceptual schemes.

If different schemes (pretty much) fit the world, then (most of) the scheme sentences of L_1 and L_2 are true, in which case "the criterion of a conceptual scheme different from our own now becomes: largely true but not translatable."[11] The reason why this idea of a conceptual scheme is unacceptable is because:

(5) We do not "understand the notion of truth, as applied to language, independent of the notion of translation."[12]

As Part 2 will show, Davidson argues that, from the Tarski-style use of the truth predicate shown in (4), and additional premises, (5) follows. The gist of the argument is that, if you give the necessary and conditions under which a sentence, s, of an unknown language is true, then you give p, the meaning of s, in your own language. The evidence used in going about such a task would be the utterances of an interpretee. If there were *complete* failure of translation between two languages, then such radical interpretation would not be possible at all. Davidson argues that if the evidence gathered by his radical interpreter were such that successful interpretation could not occur, then the interpretee could not be speaking a natural language. To support the claim that "no significant range of sentences could be translated into the other"[13], it would need to be shown that Davidsonian radical interpretation would not work under conditions where an interpretee is speaking a natural language. Showing this, and thereby vindicating the total untranslatability claim, has in my opinion proved unsuccessful, as later parts will try to show.

If *partial* translation failure were possible, then at least one of the interpretee's beliefs would not be expressible in the language of the interpreter. But Davidson maintains that, from his views of the methodology of interpretation, it is not at all clear how there could be such an outcome:

---

[10] Davidson (1984), p. 194.
[11] Davidson (1984), p. 194.
[12] Davidson (1984), p. 194.
[13] Davidson (1984), p. 185.

(6) "Given the underlying methodology of interpretation, we could not be in a position to judge that others had concepts or beliefs radically different [i.e. 'incommensurable' in the Davidsonian sense] from our own."[14]

In the case of radical interpretation, what we have is:

(a) The meaning of s

(b) What the alien believes

From the alien's utterance of s, or assent to s, the radical interpreter knows:

(c) The alien holds that s is true

As Davidson puts it, (c) is "the vector of two forces"[15], (a) and (b): "A speaker holds a sentence to be true because of what the sentence (in his language) means and because of what he believes."[16] The prima facie problem is that the radical interpreter knows neither (a) nor (b), but Davidson maintains that the interpreter has in principle full access to (a) and (b) in the way described in the next paragraph.

Since the interpretee speaks a natural language, he has intentional states, such that "causal links [...] run between states of the environing world and intentional states of the [interpretee]."[17] If the radical interpreter can determine those causal links, he will have epistemic access to the physical conditions under which the alien holds s true. As Davidson puts it:

> I ask myself what sentence of mine I am stimulated to assent to whenever you assent to a particular sentence of yours, and I use my sentence to give the truth conditions of yours.[18]

The truth conditions are expressed in the form of a T-sentence given in (4), thereby matching s with a sentence, p, of the radical interpreter's language. Such truth conditions are based on two main premises. First, that from an interpretee's holding s to be true under certain conditions, the interpreter is justified in holding p true under the same conditions. Second, that the observable features of a particular occasion of alien utterance offer evidence that the alien sentence uttered *is* true.[19] Both of these premises will be returned to, especially the first of which expresses the principle of charity (POC). Further constraints of a formal nature relating the structure of sentences to their truth conditions and values will be such that T-

---

[14] Davidson (1984), p. 197.

[15] Davidson (1984), p. 196.

[16] Davidson (1984), p. 134.

[17] Ramberg (1989), p. 69.

[18] Davidson (1993), p. 39.

[19] "the T-sentence does fix the truth value relative to certain conditions, but it does not say that the object language sentence is true because the conditions hold." Davidson (1984), p. 138.

sentences "giving correct interpretations"[20] will result. So Davidson has found a way to get at (a) and (b) and that way involves the POC.

The general argument against *partial* translation failure proceeds with (6), a statement which begins with Davidson's methodology of interpretation and draws a conclusion about beliefs. A summary of that methodology is: having only evidence for (c), an interpreter must make *general* assumptions (POC) about (b); applying the POC under specific circumstances, the interpreter may look for *specific* knowledge of (b); knowing (c) and (b), the interpreter may adduce (a). *In the case of s being false, the POC is not applicable*, so the interpreter may not combine (c) and (b) to get (a). This looks like it may offer the possibility of partial failure of translation. *Now*, the evidence available to the radical interpreter is the circumstances of other token utterances of s, and confirmed T-sentences for sentences with structural similarities to s. Given evidence particularly of the latter kind (where the POC *is* applied), the radical interpreter could determine (a). Combining (a) and (c), the interpreter can adduce (b).[21]

Satisfactorily formulating the POC has been problematic for Davidson. He has tried "minimize disagreement" and "maximize agreement", but as he admits, "The aim of interpretation is not agreement but understanding."[22] He continues:

> My point has always been that understanding can be secured only by interpreting in a way that makes for the right sort of agreement. The 'right sort', however, is no easier to specify than to say what constitutes a good reason for holding a particular belief.[23]

I will loosely formulate the POC as:

(7) What the interpretee holds as true and asserts, the interpreter also holds as true and asserts under the same conditions.

Bjørn T. Ramberg points out that "It is this very notion of truths-for-languages as somehow the same that drives interpretation."[24] He explains:

> In a true T-sentence, s and p are appropriate to the occasions of empirical observation in exactly the same manner. It is by assuming this sameness of truth, which is the intuitive foundation of Davidson's model of interpretation, that the interpreter is able to understand [the interpretee].[25]

---

[20] Davidson (1984), p. 152.

[21] So evidence for disagreement over specifics is to be had against a background of common understanding and agreement.

[22] Davidson (1984), p. xvii.

[23] Davidson (1984), p. xvii.

[24] Ramberg (1989), p. 76.

[25] Ramberg (1989), p. 76.

For Ramberg, this implies that "[t]he concept of truth that underlies a theory of interpretation is a concept of absolute truth."[26] The POC is the claim that, for each "systematic correlation of sentences held true with sentences held true"[27], what it is for s to be true is the same as what it is for the interpreter's correlated sentence to be true. I have characterised the POC in (7) in terms of common assertability or holding as true, but (7) is used in conjunction with an adequacy condition for use of the truth predicate (4) in order to account for interpretation, as Ramberg's explanation shows. This slippage between holding as true, or assertability and being true will be addressed in Part 4, and again in Part 5. At this point, though, the aim is to clarify that the POC claims that what is true for the alien is true for the interpreter under the same conditions; so the POC a statement about beliefs, for it refers to (the sentences which an alien would utter to express) the beliefs he would have. In particular, the POC (7) is a claim about "general agreement of beliefs"[28].

Assumption (7) is not always applicable, nor does Davidson intend it to be. It expresses a "general policy, to be modified in a host of obvious ways."[29] There may be occasions when the alien lies; he and the interpreter may each have defeasible beliefs or different beliefs. Despite this, the radical interpreter cannot do without presuming the POC (7): it is initially indispensable, yet open to suspension for the nonce in the light of evidence. Why this is so is as follows. If the alien utters a false sentence, either deliberately or erroneously, then it is of no immediate use to the radical interpreter, for "the negative truth-value of a sentence severs the connection between sentence and observable circumstance"[30]. The alien sentences which the interpreter *depends* upon are the true ones about the observable circumstances of utterance. Only after a preponderance of such sentences are recorded in T-sentences will the interpreter be able to know sufficiently members of the extension of the truth predicate (of his own language) 'true-in-alien-language' to recognise false alien sentences *as sentences*. Ramberg puts this point more clearly:

> The only possible incentive the field linguist could have for attributing error or deceit is that a speaker's utterance [...] conflicts with [...] inductively acquired T-sentences. And these [...he] could only have formulated by treating as true the native speakers' previous utterances of [...s] or other expressions in which structural elements of [...s] occurred.[31]

---

[26] Ramberg (1989), p. 76.
[27] Davidson (1984), p. 197.
[28] Davidson (1984), p. 196.
[29] Davidson (1984), p. 152.
[30] Ramberg (1989), p. 72.
[31] Ramberg (1989), p. 70.

It is not that 'exceptions prove the rule of the POC', but rather 'without the POC we never get to discover the exceptions to *or* exemplifications of it'. Without assuming the POC, the interpreter would never have *evidence* that the interpretee is an utterer of truth-bearing sentences.

While the interpretee may have some beliefs which are *different* to those of the radical interpreter (in which case s is false by the interpreter's lights), the "totality of possible sensory evidence, past, present and future"[32] along with the appropriate structural constraints are such that the radical interpreter will be able to interpret the sentence expressing that different belief:

> Attributions of belief are publicly verifiable as interpretations, being based on the same evidence: if we can understand what a person says, we can know what he believes.[33]

Using this approach, Davidson has offered two arguments[34] for the intertranslatability of natural languages. Since there are no non-translatable natural languages, there are no incommensurable natural languages.[35] If there *were* different conceptual schemes, then they would not be translatable[36]; but Davidson has ruled out non-translatability; so there are no different conceptual schemes (nor, therefore, are there incommensurable ones).

Part 2: Meaning: Truth and Interpretation

> An acceptable theory should [...] account for the meanings (or conditions of truth) of every sentence by analysing it as composed , in truth-relevant ways, of elements drawn from a finite stock.[37]

Part 2 begins by presenting Davidson's three adequacy conditions for *any* theory of meaning and then goes on to consider Davidson's own theory of meaning[38]. Since Davidson seeks to extract a theory of meaning from a theory of truth, I spend most of this part presenting the Tarski/Davidson "empirical theory of truth"[39], showing

---

[32] Davidson (1984), p. 193.

[33] Davidson (1984), p. 153.

[34] One against complete translation failure, and one against partial translation failure. Each was briefly presented earlier.

[35] In (2) Davidson equates 'not translatable' with 'incommensurable'.

[36] In (1) Davidson equates different conceptual schemes with nonintertranslatable languages.

[37] Davidson (1984), p. 56.

[38] 'Theory of meaning' in Davidson's "mildly perverse sense". Davidson (1984), p. 24.

[39] Davidson (1984), p. 139.

the reasoning behind (4). The POC (7) is warranted by the claim that this theory of truth is also a sufficient description of any theory of interpretation.

Knowledge of any adequate theory of meaning would enable one to understand and use (all the sentences of) the language (Davidson has in mind a natural language) mentioned by the theory. This is the condition of interpretation. Davidson's first condition on a theory of meaning does not require that the theory of meaning be one that we actually *do* use in interpreting others. Rather, the interpretation condition directs us towards any theory which would enable us to interpret the words of others. Given the first adequacy condition, Davidson's theory of meaning is necessarily a theory of interpretation.

If a theory of meaning for a natural language is to meet the condition of interpretation, then it must enable one to understand an infinite number of sentences and sentences which have not been previously encountered. Such a theory would need to provide a finite number of axioms which can generate an infinite number of sentences thereby accounting for the learnability of a natural language. This is the second adequacy condition on a theory of meaning: compositionality.

A theory for meaning for a natural language must sufficiently describe all the semantic properties of that language. Although natural languages employ extensions and intensions, Davidson requires that a theory of meaning describe the language using only extensional resources. The third adequacy condition, then, on any theory of meaning is that it be extensional. Davidson claims a practical justification for this third condition:

> My objection to meanings in the theory of meaning is not that they are abstract or that their identity conditions are obscure, but that they have no demonstrated use.[40]

If an extensional theory of meaning can sufficiently describe a natural language, then it will indeed have shown that intensions have no demonstrated use in that theory of meaning. As no such theory has yet been presented here, it may be tempting to regard the extensional condition itself as premature and of no yet demonstrated use. However, there are other reasons for requiring that a theory of meaning be couched extensionally. One of these reasons is that "[t]he extensions of

complex expressions are functions of the extensions of their parts"[41], whereas the intensions of complex expressions do not display such compositionality. Accepting the compositionality condition inclines us towards the extensional condition.

An illustration of why an adequate (and so compositional) theory of meaning ought not to employ intensions is found in Davidson's criticism of Frege. Consider the sentences:

(A) Galileo said that the earth moves.

(B) Galileo said that the third planet from the sun moves.

Frege claimed that in intensional contexts such as (A), the referent of 'the earth' is not the earth but the normal sense of the expression 'the earth'. Since the normal sense of 'the earth' is different to that of 'the third planet from the sun', Frege has explained why 'the earth' and 'the third planet from the sun' are not intersubstitutable salve veritate in (A) and (B). The upshot of Frege's view is that "'the earth' has two referents, depending on its context: the earth itself, and the sense of the expression 'the earth' [...] as it features in 'the earth moves'".[42] Indeed, in

(C) Davidson said that Galileo said that the earth moves.

'the earth' refers to the sense of 'the earth' as it features in 'Galileo said that the earth moves'. For Frege, each new intensional context produces a new sense and a new referent; consequently, any referring expression, such as 'the earth',

> has an infinite number of entities it may refer to, depending on context, and there is no rule that gives the reference in more complex contexts on the basis of the reference in simpler ones.[43]

Davidson's claim that "the one thing meanings do not seem to do is oil the wheels of a theory of meaning"[44], and the concomitant extensional condition on an adequate theory of meaning, follows from the compositionality condition, which in turn is supported by the interpretation condition.

One way in which Davidson argues from the three adequacy conditions for his own theory of meaning is as follows. The interpretation condition states that an adequate theory of meaning will pair every well-formed unknown sentence, s, with a sentence

---

[40] Davidson (1984), p. 21.

[41] Evnine (1991), p. 77.

[42] Evnine (1991), p. 91.

[43] Davidson (1984), p. 99.

[44] Davidson (1984), p. 20.

I understand, p. The nature of this pairing will be such that s means that p. The problem for the theory of meaning is that it cannot describe the pairing in terms of 'means that', for the theory must use only extensional resources. An extensional pairing could employ the material biconditional: s if and only if p. This is a metalinguistic statement in the language of the sentence p about the object language (OL) sentence named by s. Since s is a name and not a sentence, we must make the left hand side of the biconditional a metalanguage (ML) sentence in order to make the whole well-formed. This is achieved by attaching the name, s, to a ML predicate. Davidson suggests the dummy predicate 'is T', giving: s is T iff p. Now Alfred Tarski famously showed that to attach a ML *truth* predicate to (the OL sentence) s is to form a sentence which is *materially equivalent* to the ML sentence p, that is:

(T) s is true iff p

where p is the sentence named by s, or is a translation of s. The latter case suits Davidson's purpose of using p to interpret the sentence which s names. To use the truth predicate in place of the dummy predicate could therefore be sufficient to yield an interpretation, p, for each unknown OL sentence named by s. In addition to the interpretation and extensional conditions being addressed, the compositionality condition is met because

> the importance of the theorems does not lie in the theorems themselves, but in their derivation. The power of a Tarskian theory lies in its showing how we can get, from a finite stock of building blocks and logical (recursive) axioms, all and only the true T-sentences for a language.[45]

An adequate theory of meaning will therefore be one which yields sentences of the form (T) as theorems.

Tarski utilised sentences of the form (T) in his proposal, Convention T:

> We wish to use the term "true" in such a way that all equivalences of the form (T) can be asserted, and we shall call a definition of truth "adequate" if all these equivalences follow from it.[46]

As has been said, Tarski has shown that, by assuming that the sentence named by s is the same as, or a translation of, p, it follows that Convention T states when we can rightly say 's is true'. Davidson turns this around and wants to claim that, assuming the sentence named by s *is* true, then Convention T sets conditions under which p is a translation into a known tongue (that is, an interpretation) of s (when the sentence

[45] Ramberg (1989), p. 58.
[46] Tarski (1985).

named by s is not in the language of p). His strategy is therefore to "take truth as basic [...in order to] extract an account of translation or interpretation."[47]

Given the Davidsonian inversion, Convention T alone does not, however, set *adequate* conditions under which p is an interpretation of s when the sentences are those of natural languages. For example,

(S_1) 'La neige est blanche' is true-in-French iff grass is green

is formally correct, but p does not interpret s. I will come back to this; for prior to this problem of adequately constraining interpretation there is the problem of how a radical interpreter would *know* that 'La neige est blanche' is true-in-French, and there is the question of what is the property 'truth' which true sentences have. So before discussing adequate constraints on meaning ascription we must determine adequate constraints on truth ascription, as well as ask what criterion or criteria characterise truth. In so doing, we are addressing the left hand side of the biconditional of a Davidsonian T sentence. Here, we are distinguishing between:

(d) What evidence there is that s is true-in-OL?

(e) What is it for s to be true-in-OL?

Davidson's answer to (d) is: the evidence that s is true-in-OL is that a speaker of OL, holding s to be true, assents to, or utters s. Here, we are assuming that the speaker of OL is expressing a belief generally shared by other speakers of OL, that he is a competent speaker and that is not deluded or lying. Davidson's answer to (e) is partly minatory in character:

> Confusion threatens when this question is reformulated as, what makes a sentence true? The real trouble comes when this in turn is taken to suggest that truth must be explained in terms of a relation between a sentence as a whole and some entity, perhaps a fact or state of affairs. Convention T shows us how to ask the original question without inviting these subsequent formulations.[48]

Convention T answers (e) using the notion of satisfaction. If an open sentence satisfies some condition, then a function maps the variable(s) of the open sentence to an entity or entities. Here, two domains are in correspondence: the free variable(s) and ordered sequence(s). The open sentence 'x is green' is satisfied by {grass, the sky, ice cream, ...} and by {my car, ice cream, the sky, ...} but not by {the sky, grass, ...}. While the open sentence 'x is green' is satisfied by some sequences and not by others, the closed sentence 'Grass is green' corresponds *only* to

[47] Davidson (1984), p. 34.

sequences which satisfy it. Likewise, 'The sky is green' corresponds with, or is satisfied by no sequence. Hence Davidson's comment:

> If the sentence has no free variables – if it is a closed, or genuine, sentence – then it must be satisfied by every function or by none. [...] those closed sentences which are satisfied by all functions are true; those which are satisfied by none are false.[49]

Just as closed sentences are a special case of open sentences, so true (or false) sentences are (each) a special case of satisfaction – satisfied by all sequences or none. For this reason, Ramberg points out: "Satisfaction is what Tarski actually showed us how to define."[50] The recursive or compositional nature of Tarski's approach proposes the use of a finite number of two types of axiom: one kind to show how a complex sentence (open or closed) is satisfied by a sequence in terms of how simpler sentences are satisfied; the other kind of axiom to give the satisfaction conditions for the simplest *open* sentences.[51] Two observations worth noting on this approach to proving T-sentences are as follows. First, while the truth predicate applies only to closed sentences, we require the broader domain of closed and open sentences "to axiomatize the systematic effect of our semantic building blocks on the truth-value of the expressions in which they occur."[52] Second, we cannot say what *makes* sentences true, for each true sentence is satisfied by *all* sequences; as Davidson puts it: "All true sentences end up in the same place"[53]. What we *can* do is "show how they got there [...] by running through the steps of the recursive account of satisfaction appropriate to the sentence."[54] That 'story' is "the canonical proof of a T-sentence"[55].

While "[t]ruth is defined for closed sentences in terms of the notion of satisfaction"[56], we may wonder if this appeal to satisfaction breaks Davidson's own prohibition, mentioned in the previous paragraph, on explaining the property of truth in terms of correspondence. Davidson even admits that "[t]he semantic conception of truth as developed by Tarski deserves to be called a correspondence theory because of the part played by the concept of satisfaction."[57] The traditional

---

[48] Davidson (1984), p. 70.
[49] Davidson (1984), pp. 47 – 48.
[50] Ramberg (1989), p. 40.
[51] See Davidson (1984), p. 131.
[52] Ramberg (1989), p. 43.
[53] Davidson (1984), p. 49.
[54] Davidson (1984), p. 49.
[55] Davidson (1984), p. 138.
[56] Davidson (1984), p. 131.
[57] Davidson (1984), p. 48.

view of a correspondence theory of truth, however, adopts what Davidson describes as 'the strategy of facts'. This approach expresses

> the desire to include in the entity to which a true sentence corresponds not only the objects the sentence is 'about' [...] but also whatever it is the sentence says about them.[58]

Davidson's and Tarski's way is to say that 'Dolores loves Dagmar' is satisfied by {Dolores, Dagmar} in case that Dolores loves Dagmar. {Dolores, Dagmar} are not facts as such, but they are the entities which (may or may not) satisfy 'Dolores loves Dagmar'. The strategy of facts, on the other hand wants "somehow [to] include the loving"[59] as part of the *fact which makes* 'Dolores loves Dagmar' true. Davidson, though, makes a clear demarcation between the semantic and the epistemological issues: *determining* the truth *value* of 'Dolores loves Dagmar' is the concern of epistemology; to say what *is* the *property* of truth-in-English in the sentence 'Dolores loves Dagmar' is a semantic problem, constrained by a Davidsonian theory of theories of truth-in-natural-languages. The Tarski/ Davidson account of truth-in-a-language

> is less ambitious about what it packs into the entities to which sentences correspond: in such a theory, these entities are no more than arbitrary pairings of the objects over which the variables of the language range with those variables.[60]

The entities involved in the correspondence relation of satisfaction are clearly not facts in the common sense of the word. This has led Evnine to venture:

> The fact that what is supposed to 'correspond' to the sentences does not include the relations that apply to the objects makes this sense of correspondence too tenuous to justify the idea that we have an explanation of truth in terms of correspondence.[61]

Evnine feels particularly justified in saying so because of a later comment of Davidson's that "Correspondence theories have always been conceived as providing an explanation or analysis of truth, and this a Tarski-style theory of truth certainly does not do."[62] But perhaps Davidson's later words are not a recantation of correspondence claims; instead, they may simply confirm that the Davidsonian view of correspondence and truth is different to the norm. This stance will be adopted as I draw this presentation of Davidson's answer to (e) to a close with a summing up. Conveniently, this will carry us to where we need to go, for what we will find is that Davidsonian views on truth will lead us from talk of the left hand side to talk of the right hand side of T-sentences.

---

[58] Davidson (1984), p. 49.
[59] Davidson (1984), p. 48.
[60] Davidson (1984), p. 49.
[61] Evnine (1991), p. 137.
[62] Davidson, quoted in Evnine (1991), p. 137.

Davidson wants to give us "correspondence without confrontation"[63]. The main point to this slogan is that "there is no way to somehow confront sentences with something non-linguistic in order to see whether they are true."[64] So correspondence is of no use in determining truth *values* – for *that* a different theory, one of verification, is required (and is given in Davidson's holism and empiricism, later). The strength of the Tarski/Davidson approach to correspondence is two-fold. First, the elements of the theory –i.e., the truth bearer (the sentence) and real entities (sequences of objects) and their relation (satisfaction) - are all clearly defined:

> Propositions, statements, facts, states of affairs, and other assorted relations figure not at all. [...] Most traditional discussion of correspondence theories has centred upon the adequacy of the definition of the relation or the adequacy of the requisite identity and individuation conditions of the *relata*. We avoid most of these worries.[65]

Secondly, where correspondence is of particular use to Davidson is in the determination of truth *conditions*. We have seen that *if* 'Dolores loves Dagmar' is satisfied by {Dolores, Dagmar}, it is so on condition that Dolores loves Dagmar. Truth conditions are important to Davidson's task of interpretation because "to give truth conditions is a way of giving the meaning of a sentence."[66] Now I come to the right hand side of sentences of the form (T).

If s is true if and only if p, then p is a condition on the truth of s. The conditions under which the English sentence 'I gave him the book' is true will depend upon who uttered the sentence, who got the book, and which book that was. Since the specific answers to these questions are not merely rule-determined, but are instead a matter of circumstances, the proof of a Davidsonian T-sentence will be partly empirical and not just syntactical (as Tarski's theorems for formal languages were). So Davidson tells us to "relate language with the occasions of truth in a way that invites the construction of a theory."[67] Such 'occasions of truth' are taken to be the circumstances under which s is uttered:

> We have agreed that the evidential base for the theory will consist in facts about the circumstances under which speakers hold sentences of their language to be true.[68]

---

[63] Davidson, quoted in Ramberg (1989), p. 47.
[64] Ramberg (1989), p. 44.
[65] Platts (1979), p. 35.
[66] Davidson (1984), p. 24.
[67] Davidson (1984), p. 44.
[68] Davidson (1984), p. 152.

These 'facts' given by p will be the truth conditions of s: strictly speaking they are *evidence* for the truth of s, for no finite amount of empirical evidence can add up to the semantic statement that s is true. This point was mentioned earlier, and Davidson reminds us of it when explaining truth conditions:

> The T-sentence does fix the truth value relative to certain conditions, but it does not say the object language sentence is true *because* the conditions hold.[69]

Davidson can say this because he is talking about one or some T-sentences. Were we to have *all* the T-sentences for a natural language, that is, "[i]f we knew that a T-sentence satisfied Tarski's Convention T, we would know it was true"[70]. Since a natural language would be capable of producing an infinite number of T-sentences, such knowledge is not forthcoming:

> A true statement for Davidson is simply one we would assert when *all* the evidence is in. No statement is ever indefeasible, but that is [...] for the epistemic reason that we never possess all the evidence there might be – and not because of a discrepancy between all possible knowledge and the way things really are.[71]

Truth conditions are epistemic, but in Davidson's use of Convention T they tend towards a limit which is semantic: it is this calculus which enables Davidson to propose an empirical *coherence* theory of truth and holistic approach to meaning.

Davidson's coherentist approach to truth is displayed in his taking consistency in the use of words as evidence that a sentence is held true. If I believe that snow is white and that snowmen are made of snow then this is evidence that I would hold that snowmen are white. He tells us:

> I called my view a coherence theory because I held (I still do) that there is a presumption that a belief that coheres with the rest of our beliefs is true. But obviously this doesn't make every such belief true.[72]

It is this expectation of consistency between utterances which solves the problem presented earlier in:

(S_1) 'La neige est blanche' is true-in-French iff grass is green.

If a radical interpreter were to set about forming a large number of sentences of the form (T), a pattern would begin to emerge. It would be observed that a preponderance of utterances containing 'neige' would occur around snow and those containing 'blanche' around things which are white. The circumstances in which these many sentences are held true, and the consistent assent and dissent to

---

[69] Davidson (1984), p. 138.
[70] Davidson (1984), p. 138.
[71] Ramberg (1989), p. 46.

different utterances and sentences containing 'neige', would rule out the truth condition given in (S_1). We see that

> The work of the theory is in relating the known truth conditions of each sentence to those aspects ('words') of the sentence that recur in other sentences, and can be assigned identical roles in other sentences.[73]

When we require that "the totality of T-sentences should [...] optimally fit the evidence about sentences held true by native speakers"[74], then we find that the right hand side of the biconditional in (T) will be an interpretation of the sentence mentioned on the left hand side. We have made the transition from a theory of truth to a 'theory of meaning'.

Davidson's use of Convention T for the interpretation of natural languages, wherein the interpreter must, on condition that p, attach the ML truth predicate 'true-in-OL' to s in order to interpret s, proceeds on the assumption of the POC (7). This principle of interpretation has many facets. I have already mentioned some of these in relation to (d), where it is assumed that a *competent* speaker of OL is *acting in good faith.* Here, the very notion of a competent speaker who acts in good faith is one who speaks and acts rationally. The POC therefore makes the cognitive claim that "if a creature has propositional attitudes then that creature is approximately rational"[75]. More will be said in later parts about the charitable "conceptual link between truth and rationality on the one hand and intentional description on the other"[76].

In Part 1 it was said that the POC is a claim about the use of the truth predicate of any interpreter's language. When Tarskian claims about the truth predicate are applied to natural languages in the context of radical interpretation, it is found that to be able to make claims that there are necessary and sufficient conditions under which the ML truth predicate applies to names of OL sentences, it is necessary to proceed on the basis of the POC. In this way, the names of OL sentences satisfy the ML truth predicate, and the necessary and sufficient conditions under which they do so are the ML interpretations of the OL sentences. This way of arriving at an interpretation makes it clear that the POC is a general semantic claim. For individual sentences of the OL, it will be found that this general semantic claim does not hold;

---

[72] Davidson, in Stoeker (ed.) (1993), p. 37.
[73] Davidson (1984), p. 25.
[74] Davidson (1984), 'Radical Interpretation', p. 139.
[75] Davidson, quoted in Evnine (1991), p. 112.
[76] Stich (1991), p. 44.

yet the general semantic claim is indispensable for, without it, the exceptions (where p is not an interpretation of s) would never be discovered. For this reason, the POC is more than a heuristic: it is not merely an instrument which yields interpretations, but part of the very notion of 'truth-in-a-natural language', and so of interpreting a natural language. For this reason Ramberg describes the POC as "a *condition of the possibility* of interpretation"[77].

Davidson's argument against complete translation failure is summed up as:

> nothing [...] could count as evidence that some form of activity could not be interpreted in our language that was not at the same time evidence that that form of activity was not speech behaviour.[78]

The language is slightly contorted, but what Davidson seems to be saying is: that which counts as evidence that some behaviour, say the utterance of s, *is* speech behaviour will also count as evidence for the interpretation of s; and enough of such evidence would, under Davidson's description of radical interpretation, lead to the translation of s. The only condition under which the methodology of radical interpretation would not yield translation into a familiar tongue is when the behavioural evidence is not an utterance from a language. So Davidson describes his argument against complete translation failure as "transcendental"[79].

Davidson's opposition to partial translation failure concerned those utterances where the POC did not hold. On these occasions, the interpreter and interpretee would have *different* beliefs under the same conditions. But were it such that the *total* behaviour of the interpretee was not evidence that both interpreter and interpretee had mostly the same beliefs under the same conditions, then radical interpretation would not be possible; in which case (see argument against complete failure), the interpretee is not speaking a language[80]. As long as the interpreter and interpretee have different beliefs under the same conditions on a minority of occasions, translation will go ahead for *every* utterance.

---

[77] Ramberg (1989), p. 74.

[78] Davidson (1984), p. 185.

[79] Davidson (1984), p. 72.

[80] "If we cannot find a way to interpret the utterances and other behaviour of a creature as revealing a set of beliefs largely consistent and true by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything at all." Davidson (1984), p. 137.

The implications of all this for the semantic IT are that statements of T_1 and T_2 will always be semantically comparable; and ruling out the possibility of failure of translation in turn rules out (according to Davidson) incommensurable theories or conceptual schemes. The claim that *all* the meanings of terms have changed, such that all the truth values of statements of T_1 are independent of the truth of statements of T_2, will be false, for it entails, for Davidson, the claim of *complete* untranslatability. Likewise, the weaker claim that, due to meaning change, *some* statements of T_1 are logically independent of those of T_2 will also be false because it entails the false partial untranslatability claim.

However, it is Davidson's justification of the POC which is more directly relevant to my presentation of Feyerabend's IT. So Ramberg remarks:

> The principle of charity serves in one form or another as the foundation for Davidson's much cited arguments against incommensurability and the possibility of our being fundamentally mistaken about how things are.[81]

The argument against complete untranslatability justifies the claims of the POC, namely, that when two people speaking different languages each make an utterance about their common physical circumstances, then, as long as they tend to (be disposed to) express agreement, the utterance of one will be evidence which the other can use to interpret what his interlocutor is saying. And the argument against partial untranslatability showed that when on occasion they do *not* express agreement, the POC, must still be applicable most of the time, if the interpreter and interpretee are language speakers. This *charitable agreement under common causal interaction with the environment* limits the degree to which the claims of two theories will differ in a common domain; it also ensures that in a *common domain*, the truth claims of one speaker are not semantically *independent* of what another speaker would be disposed to assert. So the POC rules out the MVT's logical independence claim.

To save the strong or weak form of the logical independence claim of the MVT, a successful argument against the POC would do the job. This is not so easy to do because of the POC's transcendental justification:

(f) If Davidson's theory of interpretation were to work for all natural languages, it must employ the POC

(g) Davidson's theory of interpretation would work for all natural languages

(h) Therefore we are justified in employing the POC

Given the careful nature of Davidson's argument, few would challenge (f). Undermining the conclusion (h) will involve explaining why Davidson's theory of interpretation would not work, or else showing that (h) does not follow from (f) and (g). Both of these approaches have weaknesses, but other, more indirect arguments have also been used to undermine (h). So, for example, Michael Devitt accepts the validity of the above transcendental arguments, but rejects them on the grounds that they are instrumentalist in nature (as the verificationist character of the arguments may suggest). Another, very different, indirect attack on the POC is mounted by Stephen Stich who accepts (h), but claims that the POC is not as philosophically potent, or interesting, as Davidson thinks. These indirect rebuttals of the POC will be considered in Parts 5 and 6.

Part 3: Common Objections

Here, I consider briefly a couple of common objections to Davidson's theory of interpretation and say why they are not sufficient (at least in the form presented) to derail Davidson's position. In subsequent Parts, I go on to look at other arguments (many of which seek to undermine the POC) against Davidson's theory of interpretation.

A common criticism of Davidson's meta-theory of interpretation is that the extensional adequacy condition is unacceptable. Such criticism generally amounts to the claim that natural languages use intensional constructions which cannot, for the purpose of interpretation, be sufficiently described purely extensionally.[82] Davidson, however, has had some success in dealing with specific types of intensional construction, such as indirectly reported speech. Davidson's defence is that critics of the extensional adequacy condition have not offered a detailed argument to support their objection; and an objection of a merely general nature does not offer a convincing challenge in the face of Davidson's specific successes. In addition, Davidson can appeal that his project is a research programme, and claim that the difficult intensional parts of language will succumb later. There are other aspects of natural language, such as irony and paradox, which would also pose problems for

[81] Ramberg (1989), p. 70.

[82] See for example: Blackburn (1984), p. 288; Taylor (1998), pp. 148-9; Grayling (1997), p. 251.

Davidson since his use of Convention T requires formal, first-order logic[83]. Davidson acknowledges that these aspects are of concern:

> I have always been aware that it is a big question whether, or to what extent, such theories can be made adequate to natural languages; what is clear that [sic] they are adequate to powerful parts of natural languages.[84]

Concerns, however, are not arguments. Davidson maintains that very general doubts about the suitability of Convention T to natural languages are not enough to show that his theory of interpretation is unsuccessful:

> To show this, or even that it is unlikely to work, they [critics] would either have to produce a priori reasons why you can't get there from here, or come to grips with the arguments which aim to show in some detail how it could be done.[85]

A further, final example of an ineffectual attack is the claim that Davidson's theory of interpretation is irrelevant, for it does not describe how we actually go about the business of interpretation. To support the irrelevancy claim it is sometimes added that Davidson offers no explanation for how a first language is acquired. However, the irrelevancy claim is itself irrelevant! As was pointed out earlier, Davidson is only saying what *would* work as a theory of interpretation, not what *is done* by interpreters:

> I am outlining what I claim could succeed, not what does. [...] I have never claimed to know how children learn their first language. (In fact, it is a mystery to me how we can correctly describe the contents of a partly formed mind [...] I have never claimed to give an account of how field linguists arrive at their theories.[86]

The next four protests address Davidson's position by attempting to show either that his arguments are not logically valid, or that one or more of his premises is false *a priori*.

Part 4: The Presumption of Truth

Simon Evnine maintains that Davidson cannot presume, without supplementing his argument for the method of radical interpretation, that all, or most, of the sentences which the interpretee, or indeed, of the interpreter, are disposed to utter are *true*. What Evnine has in mind is that Davidson's argument as presented so far is not

---

[83] Notwithstanding Davidson's admission that he was "hasty" to rule out use of modal logics, possible world semantics and substitutional quantification as able to meet the demands of Convention T, "[t]he well-known virtues of first-order quantification theory still provide plenty of motivation, however, to see how much we can do with it." see Davidson (1984), pp. xv-xvi.

[84] Davidson, in Stoeker (ed.) (1993), p. 83.

[85] Davidson, in Stoeker (ed.) (1993), p. 84.

[86] Davidson, in Stoeker (ed.) (1993), pp. 80-1.

sufficient to fend off the sceptic. The pessimistic induction that one globally false theory succeeds another opens up a space where one group of people hold a globally false theory and another group hold another globally false theory. Under these circumstances there is general agreement among the members of *each* group, but the method of radical interpretation would fail to secure intergroup translation. Then the POC would not be justified and it is possible that each group holds logically independent general theories in a common domain.

I have taken the POC (7) to express the claim that an interpreter and alien have mostly the *same* beliefs, but it may be recalled from Part 2 that Davidson makes the further claim that those beliefs are mostly *true*. Davidson must therefore warrant his claim that "What makes interpretation possible [...] is the fact that we can dismiss *a priori* the chance of massive error."[87] As far as Evnine is concerned:

> The method of radical interpretation takes us as far as ensuring that [...] the interpreter must take the interpretee to be largely a believer of truths. But it does not appear to guarantee the truth of the beliefs of either.[88]

For such a guarantee, claims Evnine, Davidson requires[89] his omniscient interpreter argument, which goes as follows:

> there is nothing absurd in the idea of an omniscient interpreter; he attributes beliefs to others, and interprets their speech on the basis of his own beliefs, just as the rest of us do. Since he does this as the rest of us do, he perforce finds as much agreement as is needed to make sense of his attributions and interpretations; and in this case, of course, what is agreed is by hypothesis true.[90]

In this case, when the interpretee is in massive error, then very little he says will be true or (therefore) agreed. Then (nearly) all of the interpretee's utterances about the common physical situation of utterance will *be* false. The evidence thus tainted, the omniscient interpreter will pair interpretee utterances with the wrong truth conditions, and p will not give the meaning of s. Davidson concludes that "massive error about the world is simply unintelligible"[91] (of natural language speakers) because even the most knowledgeable radical interpreter could not interpret one who is in massive error. Since Davidson's constraints on radical interpretation are adequate for the interpretation of all natural languages, global error can only be attributed to one who does not speak a natural language. From all this, two questions arise. First, what are we to make of Davidson's omniscient interpreter

---

[87] Davidson (1984), pp. 168-9.
[88] Evnine (1991), p. 141.
[89] Evnine (1991), p. 142.
[90] Davidson (1984), p. 201.

argument? And second, is Evnine right to maintain that Davidson needs the omniscient interpreter argument to demonstrate that the interpretee *is* largely a believer of truths?

To answer the second question first, Davidson's answer is that the omniscient interpreter argument is an adjunct to the main argument upholding the charitable claims that the beliefs of all natural language speakers are mostly common and mostly correct. This is what Davidson says *before* introducing the omniscient interpreter:

> It may seem that the argument so far shows only that good interpretation breeds concurrence, while leaving quite open the question whether what is agreed upon is true. And certainly agreement, no matter how widespread, does not guarantee truth. This observation misses the point of the argument, however.[92]

To see why Davidson thinks this, I will turn to the first question.

Davidson tells us that "all interpretation, whether radical or not, must be constrained in certain ways, and therefore [...] all natural languages must have certain properties."[93] One of those properties is "correspondence between observed utterances and specifiable features of the environment"[94]. Part of that correspondence is a causal relation between a speaker and his environment; another is the intentional relation. But Davidson's mention of "correspondence" does not indicate that he is offering a correspondence view of truth (in the traditional sense). He is merely saying that the environment – or world – *is involved* in the translation, and so the very notion, of a natural language. The possibility of global error, while rejected by Davidson, is not *directly* refuted. Davidson just points out that, in cases where either the interpretee or the interpreter has mostly false beliefs, then interpretation is not possible; and if interpretation is not possible, then one – or both - of the parties is not a speaker of a language. So "Davidson does not provide metaphysical assurance of our connection with reality, he simply makes the point that if we give up the world, we must also give up language."[95] The omniscient interpreter argument is striking because it makes this very point, but makes it to the

---

[91] Davidson (1984), p. 201.

[92] Davidson (1984), p. 200.

[93] Davidson, in Stoeker (ed.) (1993), p. 78.

[94] Ramberg (1989), p. 47.

[95] Ramberg (1989), p. 47. Such a light metaphysical touch should not come as much of a surprise when it is recalled that the Davidson/Tarski semantic conception of truth is a view with a minimum of metaphysical ballast.

nth degree: epistemological scepticism does not make semantic sense. I will consider Davidson's position a little further.

Davidson does not deny that "a sentence may be false in spite of the indications of all *available* evidence"[96]; what he rejects is "that a sentence might be false in spite of the indications of all *possible* evidence"[97]. Davidson does not attempt to give criteria for individual sentences by which we can judge whether an individual sentence is true, or its terms refer. Instead, he shows why it is wrong to claim that the majority of all *possible* sentences of a natural language could be false[98]. It is in this way that Davidson rejects scepticism, and I believe that, contra Evnine, he does not need the omniscient interpreter argument to do so.

By way of assessing Davidson's argument against the possibility of natural language users having mostly false beliefs I consider briefly the case of the brain in the vat. Colin McGinn points out that it is consistent with the POC that:

> my brain and the brain in the simulation machine or in the vat could be physically indiscernible and yet we would, on the Davidsonian view, experience and believe totally different things [...] I believe that there is a brown rabbit running by and I have a visual experience as of a brown, rabbit-like creature running by; they believe (say) that an electrode is sending n volts into their occipital lobe and they have an experience with just this content. But there is no difference in what is going on in our brains.[99]

In the case where every belief of the vat brain could be false, Davidson's view may be that radical interpretation would be impossible because the radical interpreter would never be able to differentiate truth conditions for vat brain utterances (assuming the vat brain had an articulation device); or, if the radical interpreter were the scientist with good access to the physical conditions of the vat brain, then he would be able to give truth conditions of vat brain utterances in terms of degrees of electrical stimulation to specific locations. Either case would rule out any notion of content *other* than that which could be publicly verifiable. In the first case it would not be possible to know that the vat brain was in global error and able to

---

[96] Ramberg (1989), p. 47. My italics.

[97] Ramberg (1989), p. 47. My italics.

[98] Richard Rorty's comments (Rorty (1980). p. 311) seem to suggest, rather misleadingly I think, that Davidson doesn't really have much of an argument: "If you ask [...] Davidson why he thinks that we ever talk about what really exists or say anything true about it, [he is] likely to ask you what makes you have doubts on the subject. If you reply that the burden is on [him], and that [he] is forbidden to argue from the fact that we would never know it if the sceptic were right to the impossibility of his being right, [...Davidson] might [...] reply that [he] will not argue in that way. [He] need not invoke verificationist arguments; he need simply ask why [he] should worry about the sceptical alternative until [he] is given some concrete ground of doubt." But surely Davidson *does* have an argument – a semantic argument – against the sceptic.

speak a natural language. In the second case, where translation can go ahead, the vat brain would not be in global error, for it would be talking about electrical impulses and such like.

According to Davidson, understanding an utterer in global error would not be possible. But what if the interpreter were in the *same* global error as the interpretee? The 'soft' Davidsonian defence here is to say that the rest of us would never know because our radical interpretation of them would fail. But I don't think that this is Davidson's main defence. Rather, I think that Davidson's point is that, even for the two in shared global error, it is not clear how two people in shared global error could interpret each other: how could two such language users *systematically* map all their own or each other's sentences (i.e. *false* sentences) to common causes? Consequently, the extreme sceptical charge that *all* natural language speakers are in one great shared global error is not rebuffed *directly* by Davidson – but it *is* rebuffed, and the rebuttal is Davidson's general point that radical interpretation must deal mostly with true sentences for they are the ones which can be systematically mapped to the environment. For this reason I disagree with Simon Evnine that Davidson needs the omniscient interpreter argument.

Part 5: Devitt's Discontent

Michael Devitt rejects the interpretation adequacy condition which Davidson places on any theory of meaning. Davidson's approach to semantics, instead of asking the question 'What is meaning?', has pursued another question which he thought would be less intractable, namely "What would it suffice an interpreter to know in order to understand the speaker of an alien language, and how could he come to know it?"[100] Davidson has redirected the task of semantics from that of giving "an explanatory correspondence notion of truth [...] explained in terms of genuine, objective reference relations"[101] to that of describing "how to construct theories of interpretation"[102] where meaning is rendered but not explained. Devitt's line of attack is to attempt to show that, because of problems caused by this 'interpretative

[99] McGinn (1986), p. 361. The problem, for McGinn, is that, as a result of Davidson's approach, "there is a rather extreme failure of the supervenience of the mental on the cerebral". McGinn (1986), p. 361. Davidson does not regard such supervenience as a problem.

[100] Davidson, in Stoeker (ed.) (1993), p. 83.

[101] Devitt (1991), p. 180.

[102] Devitt (1991), p. 187.

perspective', Davidson is forced to adopt the POC; yet the POC is, for a number of reasons, objectionable.

Devitt's 'take' on Davidson is presented in the next few paragraphs. Then come Devitt's objections to Davidson's enterprise. The thrust of these objections is that Davidsonian semantics in general, and the POC in particular, are tainted with instrumentalism. If the instrumentalist charge can be upheld, then the POC need be considered little more than a heuristic device, in which case the arguments supported by the POC (including those against the IT) are seriously undermined; for rather than Davidson having shown that language users are mostly in agreement and mostly speakers of the truth, he will then just have shown that these are good assumptions to make when placed in the position of a radical interpreter.

Devitt charges Davidson with 'semanticalism', the view that

> Our theory of the world has need of explanatory semantic notions which are basic and inexplicable in non-semantic terms - for example, in physical terms.[103]

Davidson admits that "the truth predicate is not defined, but must be considered a primitive expression"[104]:

> Not that the concept of truth that is used in T-sentences can be explicitly defined in non-semantic concepts, or reduced to more behaviouristic concepts. Reduction and definition are [...] too much to expect.[105]

As regards other semantic concepts like satisfaction and reference, "we know all there is to know about them when we know how they operate to characterize truth."[106]

For Devitt, the problem with such semanticalism is that its account of meaning lacks a clear description of how language hooks on to the physical world. This issue of reference or correspondence truth will turn out to be *the* issue between Devitt and Davidson. According to Devitt, Davidson gets around the anti-physicalist import of semanticalism by two epistemic means. The first of these is to proceed on the basis that "meaning is determined entirely by observable behaviour, even readily

[103] Devitt (1991), p. 182.
[104] Davidson (1984), p. 216.
[105] Davidson (1984), p. 223.
[106] Davidson (1984), p. 223.

observable behaviour."[107] However, this approach merely creates another problem, namely:

> readily observable behaviour is far too thin a basis to do all this work: to determine interpretations, to determine attitudes, *and*, let us not forget, *to explain truth*.[108]

Thus Davidson is forced to resort to a further constraint on the evidence – the POC. With charity to hand, Davidson can finally proceed with the interpretative perspective, a perspective which "is an attempt to have irreducible semantic facts while retaining physicalistic respectability."[109] This is what Devitt sees when he looks at Davidson. Next come his critical arguments.

First I consider Devitt's attack on charity. To employ the POC in order to interpret is, says Devitt, to "seek understanding by imposing an interpretation."[110] Devitt regards the POC as asserting what "is constitutive of a person's having beliefs and expressing meaningful utterances"[111]: the interpretee's utterances express beliefs which would mostly agree with the interpreter's and which are themselves largely consistent. Such an imposition of meaning is unacceptable to Devitt because the imposition of an interpretation does not guarantee the *correct* interpretation, and "What is the point of attaching a meaning to a person's words if they don't *really* have that meaning?"[112]

As regards this first cluster of objections, Devitt's charge that the POC is constitutive of, or imposes, meaning, seems to me to lack teeth for two reasons. First, Devitt does not spell out in what sense 'constitutive of meaning' or 'imposing meaning' is true of the POC. What I think is clear is that the POC does not apply in every case, so it does not impose meaning in every case: the POC constrains the *overall* interpretation, but it is suspended locally. If the POC imposes meaning, then it does so not 'across the board' but indirectly in a significant number of instances. So I feel that the accusation that the POC imposes meanings needs to be more accurate or refined for the discussion to come further.

---

[107] Devitt (1991), p. 190, quoting Davidson.
[108] Devitt (1991), p. 190.
[109] Devitt (1991), p. 191.
[110] Devitt (1991), p. 192.
[111] Devitt (1991), p. 192.
[112] Devitt (1991), p. 198.

The second reason why this first attack on the POC seems to lack teeth is that Devitt does not make clear what is *wrong* with the constraint which the POC does constitute (as I have described it above). Just as the POC is not a hard and fast rule, it is surely also not, as Ian Hacking thinks[113], a rule of thumb. The POC is how an interpreter *must* begin. Why it is so that the POC is a necessary opening gambit was explained in Part 1; yet Devitt does not appear to address this explanation or the nature of this 'imposition'. To re-phrase (with a slightly different emphasis) my second reason why Devitt's imposition of meaning charge lacks teeth: Devitt does not show that the assumption that the POC *is* correct in a majority of cases would *ever* result in the *wrong* meaning being ultimately 'imposed'.

Devitt's second cluster of objections hang around the notion that Davidson is instrumentalist in his approach to semantics. Devitt substantiates this objection in three ways. The first is that the POC is merely an instrument for yielding interpretations: "It seems as if the Davidsonian adopts [it] simply because without [it] no interpretation would be possible."[114] If a theory of meaning need not provide interpretations (as Devitt maintains), then the POC would lack any justification.

A similar objection is also raised by Jerry Fodor and Ernest Lepore[115], to which Davidson responded:

> I do not believe I have presented any argument that makes the need for charity depend on radical interpretation; the argument goes the other way round.[116]

The POC is not justified by the task of radical interpretation; it is employed even by speakers of the same language[117]. So it is not the case that the POC is an instrument only in the exotic cases; rather "all understanding of the speech of another involves radical interpretation"[118] in the sense that all linguistic understanding involves the use of the POC. If we dispense with charity, we dispense with people talking with each other; that the former is a *requirement* of the latter is surely all there is to the instrumental nature of the POC. While the POC is *instrumental* to linguistic communication, I don't see that this makes it instrumentalistic (i.e. merely a useful

---

[113] Hacking (1975), p. 147.
[114] Devitt (1991), p. 198.
[115] In Stoeker (1993)
[116] Davidson, in Stoeker (1993), p. 78.
[117] As Davidson shows in his example of the ketch and the yawl in Davidson (1984), p. 196.

fiction for explaining linguistic behaviour). It is true that the POC is part of a strategy to render meanings rather than to say what meaning is, and that the only argument for accepting the POC is that it is instrumental in this way. But Devitt has not undermined the POC's *a priori* status and shown that charity is *merely* a pragmatic or instrumental*ist* notion.

The second charge in the instrumentalist cluster of three concerns reference: Davidson's "[t]alk of reference in that theory is a mere instrument for yielding the T-sentences"[119].To Devitt's dismay, Davidson does not explain the truth of a sentence in terms of the reference of its parts; nor does he explain the meaning of a term as its referent. I will take this second charge together with the third. The third charge in the instrumentalist bunch is related to the issue of reference. Devitt maintains that, for Davidsonian, 'truth' is merely warranted belief[120]. The truth of a sentence, for Davidson, is a matter of "facts about the structural properties of utterances, behavioural facts, and environmental facts"[121]; but Devitt insists "these are not the physical facts but the physical *evidence*"[122]: so Davidson's entire interpretative project amounts to positing (and assimilating) lots of empirical hypotheses in the form of T-sentences; but "[t]o suppose that the only facts a theory must make contact with are the evidential ones is instrumentalistic."[123] A more proper (i.e., noninstrumental) realism about semantic facts (such as the contents ascribed to utterances) "will suppose, by contrast, that there is a factual realm underlying the evidential one with which he is trying to make contact."[124]

By way of a reply to Devitt's criticisms of Davidson's views on reference and truth, I note that Davidson himself distinguishes two complaints related to his theory's notions of truth and reference:

1. his "theory of truth does not throw light on the semantic features of the basic vocabulary of predicates and names"[125]

and

---

[118] Davidson (1984), p. 125. Here, Davidson appears to shift from presenting a theory of interpretation that *would* work to one which actually *is* employed.

[119] Devitt (1991), p. 191.

[120] Quine also senses some confusion of truth and warranted belief on Davidson's part: Quine (1981), p. 39.

[121] Devitt (1991), p. 185.

[122] Devitt (1991), p. 185.

[123] Devitt (1991), p. 185.

[124] Devitt (1991), p. 185.

[125] Davidson (1984), p. 217.

2. his "theory of truth gives no insight into the concept of truth."[126]

As regards complaint number 2, Davidson admits that he uses truth as a primitive notion. However, that is not to say that Davidson's theory tells us nothing about truth. For example, his theory "reveals how the truth of every sentence of a particular L depends on its structure and constituents."[127] It also gives an extensional characterisation of the truth predicate for any natural language. That his theory of interpretation does not present a general conceptual analysis of truth would be of no hindrance as regards the purpose of interpretation to which Davidson puts the theory. Hence his response to complaint 2: "The point may be granted without impugning the interest of the theory."[128] However, Davidson's semantic theory of truth *does* claim "to give a complete account of the truth of *sentences*"[129], and to do so a notion of reference would be required, for "[t]ruth [...] clearly depends on the semantic features of the elements; and where the elements are names or predicates, what features can be relevant but reference?"[130] This brings me to complaint 1.

Davidson does not deny the reference relation in language: "[i]f the name 'Kilimanjaro' refers to Kilimanjaro, then no doubt there is some relation between English (or Swahili) speakers, the word, and the mountain."[131] A truth conditional account of meaning, such as Davidson offers, assumes the notion of reference (and, as Part 2 has shown, of satisfaction):

> Explaining the truth conditions of a sentence like 'Socrates flies' must amount to saying it is true if and only if the object referred to by 'Socrates' is one of the objects referred to by the predicate 'flies'.[132]

Yet, while the bare idea of correspondence between word and object is assumed, it is not put to any further use, for it is Davidson's belief that the reference relation is not semantically interesting; that is, reference fails to explain semantic concepts (of correspondence, truth, and such like) in terms other than semantic ones. As regards reference,

> it is inconceivable that one should be able to explain this relation without first explaining the role of the word in sentences; and if this is so, there is no chance of explaining reference *directly* in non-linguistic terms.[133]

---

[126] Davidson (1984), p. 217.
[127] Davidson (1984), p. 218.
[128] Davidson (1984), p. 218.
[129] Davidson (1984), p. 218. My italics.
[130] Davidson (1984), p. 216.
[131] Davidson (1984), p. 220.
[132] Davidson (1984), p. 216.
[133] Davidson (1984), p. 220. My italics.

It is this claim, particularly its final main clause, which is in many ways the crux of the matter between Devitt and Davidson.

Devitt believes that some causal-historical theory of reference will describe the reference relation in non-semantic terms. According to the causal-historical theory of reference, I could determine which person 'Socrates' refers to and what objects satisfy 'flies'; and then I could determine whether 'Socrates flies' is true. For obvious reasons, Davidson calls this approach to determining the truth value of a sentence 'the building-block theory'. The flaw (from where Davidson is standing) in this account of determining the truth value of a sentence is explained by Ramberg:

> While we might be able to formulate a causal theory of reference without using the concept of truth [...] testing such a theory presupposes knowledge of the truth-value of sentences, knowledge which we have come by independently of the theory to be tested.[134]

Chapter 3 has shown how Devitt's own CTR works by stipulating references in grounding sentences, and from the reference groundings, the truth of sentences (at the object level) is determined. Truth at the object level is externally determined, according to the CTR. But what of the CTR itself - how can the truth of the *theory*, truth at the meta-level, be justified? It cannot be done by building up from the reference of terms to the truth of the theory: for that, a meta-theory of reference to determine the reference of the theory's own terms would be required. So Davidson's point is that, though Devitt might try "to give a non-linguistic characterization of reference, [...] of this there seems no chance"[135], for we just come back to where we started, trying to determine the truth value of sentences (but this time at the meta-level – the sentences of a causal theory of reference). So Davidson believes that the building block approach is doomed to fail to account adequately for the truth of sentences.

Davidson would agree with Devitt that truth conditions can be *described* in terms of the reference relation, and that the truth value of a sentence can be *described* as composed of "the referential properties of its parts."[136]; but Davidson does not accept that the truth value or conditions of a sentence are *determined* by the referential properties of its parts. For his part, Devitt accepts that "[t]he Davidsonians are of

---

[134] Ramberg (1989), p. 27.
[135] Davidson (1984), p. 221.
[136] Devitt (1991), p. 174.

course right to emphasize that the evidence for a semantic explanation is at the level of sentences"[137]; yet he comes to a different conclusion:

> But this does not support the view that truth, but not reference, is a place of 'direct contact between linguistic theory and events, actions or objects described in nonlinguistic terms'"[138].

But surely Devitt is only warranted in drawing this conclusion if he can refute Davidson's *a priori* point, outlined in the previous paragraph, that to *verify or justify* any semantic theory (of truth or reference, for example) we will need to use the notion of *truth*, and truth is predicated of *sentences*.

On this a priori foundation, Davidson builds an empirical theory of truth and interpretation. While "the only way to find out whether a particular expression refers to a particular object is to see how that term affects the truth-value of the sentences in which it occurs"[139], knowing truth-values of sentences is not sufficient to determine reference or interpretations. We must then hypothesise in relation to what physical evidence a sentence could be considered true, and we do this by giving truth conditions. Having provisionally assigned truth values and truth conditions, we have a first pass at forming T-sentences. So Davidson's 'holistic method' (in contrast to the 'building-block method') "assigns no empirical content directly to relations between names or predicates and objects."[140] Instead, the holistic method assigns empirical content at the level of sentences; and this surely implies that true sentences *are* the points of contact between the physical and the linguistic.

When Davidson himself tells us that he assumes reference in order to implement his theory of interpretation[141], this perhaps sounds to Devitt as if Davidson is saying, 'It makes no difference whether a term refers (to a particular physical object) or not, as long as I can *assume* a correspondence relation which gives me the truth conditions I need to form the T-sentences which will enable me to predict utterances accurately.' This, then, is the charge of epistemological instrumentalism[142]; but it does not seem to me to be an accurate description of Davidson's position. True, Davidson assumes correspondence as a means of accounting for truth conditions;

---

[137] Devitt (1991), p. 186.
[138] Devitt (1991), p. 186.
[139] Ramberg (1989), p. 26.
[140] Davidson (1984), p. 223.
[141] Davidson (1984), p. 222.
[142] "The suggestion that theories are true or false but that that fact plays no role in our understanding [...of the theory...] will be called epistemological instrumentalism." Newton-Smith (1981), p. 30.

but the constraints of coherence truth, applied in conjunction with empirical evidence of circumstances of utterance, ultimately yield an equivalence between a pair of sentences most of which *are* true, and all of which are true under the conditions given by the other.

Davidson's view of reference and satisfaction does not imply that language has no correspondence with physical objects; all it implies is that any such correspondence is "beyond the reach of direct verification."[143] Yet it may be recalled from Part 2 that Davidson has "semantical notions of satisfaction and reference [which appeal] to an ontology of sequences and the objects ordered by the sequences"[144]. This Davidsonian ontology appears to lack any physicalistic import. Ramberg clears up this issue of what kind of objects Davidsonian correspondence corresponds to:

> Davidson calls himself a realist because the only way to construct a semantic theory of truth, to give the truth conditions of sentences, is to postulate a relation between language and the world. But this relation does not serve justificatory purposes of any kind. He is a coherentist because the only way to test claims to truth, to determine the truth-value of sentences, is to see how they cohere with other truths.[145]

Devitt's view that Davidson has a "special sort of non-physical understanding"[146] of objects and facts is perhaps based on the assumption that, were Davidson to have assumed that the referents (of, for example, scientific terms) are physical, then he would advocate using them "to give a rich content to each sentence directly on the basis of non-semantic evidence"[147]. But Devitt would be wrong to make this assumption. Davidson does not deny that scientific objects are physical, only that their being physical helps much in answering our semantic questions; for "we can be no more, or less, sure about meanings than about facts in the world."[148] It is this epistemological position that gives Davidsonian semantics an antirealist hue, but Davidson still maintains his realist credentials:

> Realism about correct interpretation does not, for me, entail that what someone means by his words is independent of what is understood by others, nor does it imply that what expressions in a natural language mean is independent of how speakers understand one another.[149]

---

[143] Davidson (1984), p. 133.
[144] Davidson (1984), p. 133.
[145] Ramberg (1989), p. 47.
[146] Devitt (1991), p. 191-
[147] Davidson (1984), p. 225.
[148] Ramberg (1989), p. 47.
[149] Davidson, in Stoeker (ed.) (1993), p. 83.

Davidson finds that "[d]oing without reference [as an evidential or explanatory notion] is not at all to embrace a policy of doing without semantics or ontology."[150] The objects to which he is overtly committed in his description of the reference relation are minimistically conceived; but a fuller consideration of the nature of T-sentences has shown that true natural language sentences are correspondence true. Like the POC, reference is, for Davidson, an instrument to advance an interpretation; but that does not make it instrumentalist. This winds up the reply to Devitt's charge that Davidson has an instrumentalist conception of reference.

The final and related charge of instrumentalism is that Davidson's theory of interpretation never gets beyond talk of warranted belief to talk of what is true: Davidson fails to move from consideration of evidence *for* the facts to description *of* the facts. I think that this accusation has been successfully dealt with already in the discussion of Simon Evnine's claim about the omniscient interpreter argument in Part 4.

Devitt's bold (or reckless?) comment that, for Davidson, "the only 'reality' determined [...] consists of T-sentences"[151] seems not to acknowledge the theory-independent nature of the facts, or the epistemic implications of being speakers and interpreters of natural languages. If speakers are not mostly talking to each other about the facts, how are they talking at all and what are they talking about?

To conclude briefly, I do not think that Devitt has managed to undermine the POC (7) or the Davidsonian theory of interpretation (4) and its a priori claims (5) and (6). With his realist credentials intact, I think it fair to conclude that Davidson's argument against nontranslatable languages and incommensurable (in Davidson's sense) conceptual schemes and for the POC remains intact.

Part 6: Charity Gets Stiched Up

Stephen Stich does not directly challenge Davidson's arguments for the POC; instead, he argues that the POC is not significant or philosophically interesting. In

[150] Davidson (1984), p. 223.

particular, Stich seeks to demonstrate that "the demarcation between states that are intentionally describable and states that are not is going to be vague, context sensitive, and observer relative; it will not be stable, or objective, or sharp."[152] Stich's tactic is to enervate the POC and show that it is then not up to the job it is supposed to do in the argument against untranslatability.

The charitable constraint on beliefs is that most beliefs are common to all language speakers. A further Davidsonian premise, discussed in Part 4, is that most beliefs are *true*. Since beliefs which are in fact true are mutually consistent, most beliefs are consistent. This final premise will be called the rationality constraint and if it is found to be unacceptable, one of the other premises, both of which are essential to Davidson's argument against untranslatability, will be undermined.

To the extent that an interpretee's utterances are not subject to the rationality constraint, attempts at interpretation are undermined. For example, an interpreter may attribute to an interpretee the belief 'If p then q'. Then the interpretee asserts that p. From this belief, p, the interpretee then infers (without any change in physical circumstances) that not-q[153]. The intentional characterisations of the interpretee's beliefs are now in question. "If, moreover, our subject as we interpret him regularly infers in this silly way, our discontent with our scheme of interpretation will grow even more acute."[154] So if a belief is to have any content ascribed, it must interact with other beliefs "in certain systematic ways"[155], that is, "a way which more or less mirrors the laws of logic."[156] If a person believes that If p then q and that p, then, if we are to understand him, he must generally infer that q. This is a general example of the rationality constraint on beliefs.

The rationality constraint has been regarded as a strong constraint by some but not by others. The strong form, the perfect rationality constraint, insists on consistency of all held beliefs, and acceptance of all logical consequences of held beliefs, as a condition of beliefs being intentionally describable. The weak form, the minimum

---

[151] Devitt (1991), p. 191.
[152] Stich (1990), p. 52.
[153] Merely simultaneous acceptance of p, if p then q, and not-q is required for irrationality, but I follow Stich's argument.
[154] Stich (1990), p. 35.
[155] Stich (1990), p. 49.
[156] Stich (1990), p. 37.

rationality constraint is considered in two varieties. The *free* minimum rationality constraint claims that no fixed set of valid inferences or consistent beliefs is needed to understand a belief, but that, for any cluster of inferences, there must be "some reasonably substantial subset of the inferences that would be required of a perfectly rational cognitive agent."[157] I will return to this constraint shortly. The *tapering* minimum rationality constraint is the view that:

> intentional characterizability, like baldness, is a matter of degree. As the distance between perfect rationality and the rationality displayed by the system at hand increases, the intentional characterizability of the system decreases.[158]

However, it seems fair to conclude that Davidson does not advocate the perfect rationality constraint. Davidson's theory allows for some difference in beliefs, and such difference could be due to inconsistent beliefs on the part of interpreter or interpretee. So Stich suggests that "[p]erhaps the most prominent advocate of the minimum rationality view is Donald Davidson"[159].

Stich's argument against the minimum rationality constraint on beliefs proceeds as follows:

> If what we are doing in offering an intentional characterization of a person's cognitive state is identifying it by way of its similarity to a hypothetical state of our own, then we should expect that as subjects get less and less similar to us in salient respects, we will increasingly lose our grip on how their cognitive states might be intentionally characterized.[160]

Stich illustrates this claim by asking us to imagine a row of people. The second person in the row has identical beliefs but one to the first person, the third person identical beliefs but one to the second person, and but two to the first person, and so on; so "each adjacent pair are *very* psychologically similar to one another"[161]. Vital to Stich's argument, though, is the further claim that "there are no interesting or significant discontinuities: there is no natural or theoretically well-motivated way to divide these people into two classes."[162] This claim, if true, calls into question the status of the POC in belief content ascription; for in order to describe intentionally the beliefs of people in the row, "we will be forced to divide them up into two radically different groups. The ones relatively close to me have intentionally

---

[157] Stich (1990), p. 40.

[158] Stich (1990), p. 41.

[159] Stich (1990), p. 41. Though Stich is not entirely sure about this, for he tells us that "There are a number of passages in which Davidson sounds instead like an advocate of the perfect rationality view." (ibid.). Is Davidson inconsistent or is Stich not clear about the POC?

[160] Stich (1990), p. 49.

[161] Stich (1990), p. 52.

[162] Stich (1990), p. 52.

characterizable states; the ones very far away do not."[163] This division "is without any psychological significance"[164], claims Stich. That is, the cognitive states of the people in the row, along with their cognitive processes, are *not* constrained by charity: the POC offers no justifiable way of distinguishing 'real' beliefs from belief-like states (those cognitive states *not* sanctioned by the POC), or 'real' inference from inference-like processes (those cognitive processes not sanctioned by the POC). The distinctions made by the POC are "observer-relative [and] situation-relative"; they are dependent upon the first person in the line (taken as the interpretative norm) and the location of the interpretee in the line. There is, says Stich, "no natural or theoretically significant boundary"[165] found in the line. The rationality constraint is therefore not of any demonstrable significance.

It seems to me that it is by no means certain that Stich's argument addresses the POC as Davidson conceives it. For Davidson, interpretation may only occur if there is a background of shared belief. The division in the line of people, if there were such a division, is therefore clear: the row of people would extend as long as each member would be able to evidence beliefs (in the form of utterances) a majority of which are in agreement with the interpreter's beliefs. Davidson's point is that an alien who would not be able to demonstrate substantial agreement is an alien without beliefs altogether and without a natural language. Stich seems to think it possible to *extend* the line in terms of difference of beliefs, for he envisages a line with all-but full agreement to all-but no agreement. Yet, as far as I can see, such a line does not figure in Davidson's view of interpretation. That is, Stich's line presupposes a tapering minimum rationality constraint; but Davidsonian charity envisages a free minimum rationality constraint. Stich has presented an argument against the former but not the latter.

In an example which parallels Stich's illustration of tapering minimal rationality, Davidson asks us to consider a row of different language speakers in which it is suggested that "the relation of translatability is not transitive"[166]:

> The idea is that some language, say Saturnian, may be translatable into English, and some further language, like Plutonian, may be translatable into Saturnian, while Plutonian is not translatable into English. Enough translatable differences

[163] Stich (1990), p. 53.
[164] Stich (1990), p. 53.
[165] Stich (1990), p. 52.
[166] Davidson (1984), p. 186.

> may add up to an untranslatable one [...] Corresponding to this distant language would be a system of concepts altogether alien to us [i.e., ones we could not understand].[167]

The point Davidson makes is that if the differences in meanings are *translatable* between two natural languages, they will *never* add up to even partial untranslatability between *any* natural languages. Part 1 of this chapter pointed out that even partial translation failure is not possible under radical interpretation, and made the following distinctions:

(a) The meaning of s

(b) What the alien believes

(c) The alien holds that s is true.

The POC does not ultimately constrain any *individual* belief; so an interpreter may not combine (c) and (b) to get (a). However, the evidence still available to the radical interpreter is the circumstances of *other* token utterances of s, and of sentences with structural similarities to s. Given evidence particularly of the latter kind (where the POC *is* applicable in a majority of cases), the radical interpreter could determine (a). Combining (a) and (c), the interpreter can adduce (b). While the alien may have some different beliefs to those of the radical interpreter in the same circumstances, (in which case s is false by the interpreter's lights), the "totality of possible sensory evidence, past, present and future"[168] along with the appropriate structural constraints are such that the radical interpreter will be able to interpret the sentence expressing that different belief:

> Attributions of belief are publicly verifiable as interpretations, being based on the same evidence: if we can understand what a person says, we can know what he believes.[169]

In Stich's example, the variable is not the language spoken but the beliefs. So the interpreter would already understand (a) and he would know (c); so he could always determine (b). Stich wants to address the claim that enough *differences* in belief would at some point add up to having no intentionally characterizable beliefs. Davidson, if he agreed with such a claim, would perhaps also object that it clumsily comes at the problem from the wrong end; for when there are too many *differences of* belief, there are no beliefs at all to be reckoned with. Avoiding the clumsiness, Davidson could say that without mostly the *same* beliefs, an interpreter has no linguistic evidence that an interpretee has beliefs at all. And if an interpreter can ascribe content to *one* alien belief, he can in principle do so to *all*. Stich's row seems

---

[167] Davidson (1984), P. 186.

[168] Davidson (1984), p. 193.

[169] Davidson (1984), p. 153.

to be a tacit denial of this point, for it suggests that there can be believers with *some* contentful beliefs, but not enough of them for content to be ascribed.

I conclude that Stich's argument against the tapering minimum rationality constraint is no argument against the free minimum rationality constraint, and so is no argument against the Davidsonian POC or the method of radical interpretation.

Part 7: Conclusion

The foregoing discussions suggest to me that Davidson shows that natural languages are intertranslatable. This conclusion implies that there is no principled bar on communication between people who speak different languages. While this is no mean achievement, it is Davidson's success in justifying the POC which has put paid to the MVT's claim that in the common domain a theory and its successor may, due to meaning variance, have few common consequences because the terms of one are logically independent of the other. The POC has established that language speakers, when placed under common causal interactions with their environment, will, most of the time, be disposed to assent to sentences which have the same meanings. In the cases where they are not so disposed, the overall constraint of charity means that an interpreter will always be able to state, in the ML, the conditions under which an OL sentence is true. By the same token, a natural language speaker who holds a theory T_2 will in principle always be able to state the conditions under which utterances of a speaker of a theory T_1 are true. Consequently, the truth claims of the latter are never in principle semantically *independent* of those of the former.

My main disagreement with Davidson's paper 'On the Very Idea of a Conceptual Scheme' is not with his conception of language, but with his view of conceptual schemes. I see no good reason to claim, as Davidson does[170], that all different conceptual schemes are incommensurable, and that incommensurable conceptual schemes are somehow equivalent to languages which fail of intertranslatability. While such notions as conceptual schemes or world views may well imply forms of representationalism which Davidson rejects, I do not believe that Davidson's semantic arguments show that these notions are fundamentally incoherent. Rather,

[170] See Part 1, claims (2) and (3).

I am inclined to agree with Larry Laudan that "establishing relations of mutual translatability is a precondition for determining that they *are* different conceptual schemes."[171]

Many of Feyerabend's comments suggest that the notions or taxonomy of one theory do not fit those of another theory, but he usually veers away from claiming explicitly that their languages cannot be *translated.* Certainly the later Feyerabend goes out of his way to *deny* that schemes or "cultures are more or less closed entities"[172] which are not open to understanding by outsiders. In the transition from classical physics to quantum theory:

> every stage of the transformation was discussed. There were clear problems; they worried both the radicals and the conservatives. Many people suggested solutions. These solutions, too, were understood by the contending parties [...] The final clash between the new philosophy and its classical predecessor found its most dramatic expression between the debate between Bohr and Einstein. Did Bohr and Einstein talk past each other? No.[173]

Of course there may have been misunderstandings between the two great men, but they were not doomed to misunderstanding each other forever! In the last Part, I trace a Davidsonian description of communication pathology in such encounters and regard this as a kind of incommensurability.

Part 8: Another Kind of Incommensurability

Bjørn T. Ramberg provides "an analysis of the semantics of incommensurability"[174]. Ramberg's analysis regards the roots of incommensurability as being embedded not in the nontranslatability of a language but rather in its mistranslation. Ramberg holds that such a "semantic obstruction"[175] is in principle only a temporary, or temporal, communication problem which can be overcome once the role of radical interpreter is adopted.

Donald Davidson has pointed out that we resort to use of the principle of charity and do "off the cuff interpretation"[176] even when understanding speakers of the same

---

[171] Laudan (1996), p. 13. My italics.
[172] Feyerabend (1995), p. 151.
[173] Feyerabend (1999), p. 267.
[174] Ramberg (1989), p. 115.
[175] Ramberg (1989), p. 119.
[176] Davidson (1984), p. 196.

natural language. However, interpretation is not always necessary: speakers of the same language can frequently rely on one another to use words in similar ways. It is when the regularities disappear that interlocutors need to begin interpreting. So interpretation is what needs to be resorted to when reliance on convention does not yield a right interpretation.

Ramberg argues that a semantic problem occurs when one speaker relies on the wrong conventions to understand his interlocutor. In such a case, one is trying to "interpret others [by] applying linguistic conventions to which they are not a party."[177] Correct translation of the misunderstood interlocutor has not been achieved because "interpretation, [that is, alteration of the theory of truth] rather than reliance on convention, is required to a greater degree than is usual"[178], but interpretation has not been (sufficiently) employed. The resulting misunderstanding, claims Ramberg, is the stuff of incommensurability[179]:

> Incommensurability in discourse can only begin to occur once we *think* we have begun to agree on linguistic conventions, but in actuality remain confused as to which language we are using.[180]

Ramberg is justified (thanks to Davidson) in discounting there being nontranslatable languages, and so I think that he is wise to separate the issue of nontranslatability from the useful one of mistranslation. The question is: 'How close is Ramberg's proposal to the semantic IT?

From Chapter 1 it will be recalled that two successive theories are semantically incommensurable if they are inconsistent, if the terms of the former have different meanings in the latter, and if they are logically or semantically mutually independent. The passive interpreter (as opposed to the active radical interpreter who is carefully studying utterance dispositions), as represented by the Received View, assumes that the conventions governing 'impetus' are the same as those governing the use of 'momentum' and translates 'impetus' as 'momentum' so that theory reduction is supported. Were the passive interpreter to become an active radical interpreter, however, he will have some evidence suggesting that, while the impetus theory, contains explicable error, impetus is momentum; but he will *also*

---

[177] Ramberg (1989), p. 132.

[178] Ramberg (1989), p. 131.

[179] Ramberg's explanation implies that the interpretee's terms have meanings different to those of the interpreter, and this clearly covers the meaning change element of the MVT.

[180] Ramberg (1989), p. 132.

have some evidence that impetus theory is false and that impetus is not momentum. (The Davidsonian approach does not tell us whether specific entities exist or not, or what the translation of any specific sentence is.) The case where impetus is not momentum is a textbook example of meaning change between two successive theories. When this case applies, but the interpreter assumes, without bothering to consult the evidence, that impetus is momentum, then Rambergian incommensurability is the result.

It is clear, then, that Ramberg's semantic proposal for the incommensurability thesis really only addresses the meaning change element. For the rest of this short section, I suggest slight additions to Ramberg's proposal in order to incorporate the conditions of inconsistency and something like logical independence. The hope is that the result will be something yet closer to the semantic proposals of the IT.

The only kind of disagreement which Ramberg's suggestion necessarily includes is disagreement over an interpretation. One person's utterance, s, is translated as p when p is *not* a correct translation of s. But from this disagreement it does not follow that s and p are inconsistent. My first little addition to Ramberg's description is that s and p must be mutually inconsistent. This simple requirement is that of Feyerabend's semantic IT. An interesting consequence of this requirement as regards incommensurable conflicts is that the inconsistency lies between s and its *mis*translation, p. So correct translation is a necessary and sufficient condition for resolution of this kind of conflict. Of course, this is not to say that the conflict between the impetus theory and Newtonian mechanics magically disappears when we are good interpreters! Some of the beliefs of one who holds the impetus theory are, when *correctly* translated, inconsistent with one who holds Newtonian mechanics. At this point, then, it is probably a good idea to distinguish on the one hand between conflict due only to mistranslation, and on the other hand genuine conflict[181].

The logical independence requirement will have to be dropped for reasons given in Chapter 1. What Ramberg's approach cleverly hints at is that incommensurable

[181] The distinction may explain Feyerabend's remark that theories "may be incommensurable in some interpretations, not incommensurable in others." Feyerabend (1978), p. 68, n. 118.

conflicts have an air of intractability: the mistranslation can persist for a long time. Ramberg offers a mechanism whereby this may occur: one thinks that one's own linguistic conventions yield the right translation. I think that the idea of intractability should be enlarged on because it seems to capture something like the logical independence claim of the semantic IT.

The POC and other premises of Davidson's radical translation put general constraints on the beliefs of speakers of natural languages. Without the agreement and rationality constraints, Davidson argues, interpretation would be inconceivable:

> If we cannot find a way to interpret the utterances and other behavior of a creature as revealing sets of beliefs largely consistent and true by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything at all.[182]

The rationality or consistency constraint on beliefs is motivated by the thought that too much irrational inference on the part of an interpretee would undermine attempts at interpretation. The rationality constraint can also work the other way: *mis*interpretation can undermine understanding the reasoning of the interpretee, so that the interpretee's inference patterns seem wildly irrational. In a situation where an interpreter has misinterpreted an interpretee in the way described so far in Part 9, and where the misascribed belief is viewed as part of a reasoning process (such as theoretical explanation), then the interpretee's reasoning could well appear suspect to the interpreter. This gives the conflict another dimension.

The new dimension, seen in terms of an undermined rationality constraint, could look something like this. Let's say that an interpretee has three consistent beliefs, that:

(i) If p then q

(ii) p

(iii) q.

Then let's say that some utterance types expressing the third belief are misinterpreted (in the Ramberg way) as asserting that not-q. The interpreter, perceiving that the interpretee is being unreasonable, must decide if the apparent irrationality is explained by poor cognitive performance on the part of the interpretee, or poor interpretation on his own part. Since the interpreter is under the impression that they are 'speaking the same language' he will more likely feel

[182] Davidson (1984), p. 137.

inclined to try and convince the interpretee that *he*, the interpretee, is being irrational. This scenario seems to create conditions suitable for a certain kind of protracted conflict which I think can be usefully described as incommensurable. For, ironically, the more an interpretee tries to justify himself and argue using utterances of the type misinterpreted as expressing that not-q, the more irrational he will appear to the interpreter. For example, "How can you say that q and *also* say that not-q?" is not an uncommon question (or implied accusation that the other person is obtuse) in this kind of conflict. It is usually met with the reply: "I didn't say that". This response is often regarded as further evidence of the interpretee's poor cognitive performance.

The *apparent* failure of the rationality requirement (due to misinterpretation) is my proposed substitute for the logical independence of the MVT. What I am proposing instead is *perceived* inconsistency *within* a theory from the point of view of Ramberg's duff interpreter. This suggestion at least has the merits of being consistent with the inconsistency claim of the MVT and of attempting to include a substitute for logical independence. It also captures what I think is the flavour of the problematic logical independence claim, which is that 'the other guy's theory does not make sense to me'[183].

Were the interpreter to stop relying on linguistic conventions and begin to radically interpret, the above conflict would be transformed. Continuing in the Davidsonian vein, the interpretee would be found to be largely rational and largely in agreement with the interpreter; and the interpreter, in understanding aright, could find out where their theories actually differed (if at all). Then there would be a *different* conflict - and a different *kind* of conflict.

Conflict Resolution theorists speak of the need, in the case of intractable conflict, to transform the conflict, where "transforming an intractable conflict into a tractable one involves changing the understanding of that conflict."[184] On way of going about such a transformation is by controlled communication, an approach to conflict founded on the premise that:

---

[183] See Feyerabend (1978), p. 70.

[184] Kriesberg, Terrell, and Thorson (eds.) (1989), p. 5.

> The process of resolution of conflict is essentially the process of testing whether information is received as was transmitted, and whether what was transmitted was sent deliberately and contained accurate information.[185]

The MVT describes incommensurable misinterpretations, misinterpretations which cause intractable conflicts where (at least) one side does not understand what the other is saying. The kind of incommensurability thesis here proposed is conceived of as "a diachronic relation, not a synchronic one"[186] between the languages (in Davidson's sense) of the holders of different theories.; for "[i]n incommensurable discourse, participants who take themselves to be speaking the same language, actually are not."[187]

[185] Burton (1969), p. 55.
[186] Ramberg (1989), p. 131.
[187] Ramberg (1989), p. 130.

General Conclusion

It is sometimes thought that, in his 1962 critique of the Received View of scientific theories, Paul Feyerabend was attacking a man of straw. The incommensurability thesis (IT) which ensued as part of that attack is then regarded as misguided from the very start. I have opposed such a view and have argued that Paul Feyerabend's 'Explanation, Reduction and Empiricism' was a well-justified critique of the Received View of scientific theories. The derivability, consistency, and meaning invariance conditions *do* logically follow from Nagelian views about theory reduction and development.

Admittedly, Carl Hempel did not share Nagel's view that a (well-confirmed) successor theory would always, in principle, reduce its predecessor; so Hempel did not hold that a successor theory would always, in principle, be able to explain its predecessor. The derivability condition is not pinnable on Hempel. However, a consequence of the double-language model of scientific theories is that two well-confirmed theories in a common domain would be mutually consistent and their (observation) terms would be meaning-invariant.[1] When such theories are mutually inconsistent, Hempel insists that such a state of affairs is an anomaly which will, in principle, be eventually resolved by observationally based crucial experiments falsifying one of them. In this case, though, the falsified theory is not, technically speaking, the same theory which was empirically adequate in the common domain, for new correspondence rules must apply.[2] Hempel also maintains that, when a hypothesis or theory is consistent with a currently well-confirmed theory, it is to that hypothesis's (or theory's) advantage. The Hempelian picture is therefore that successive theories are likely consistent and (observationally) meaning invariant; but establishing that a predecessor is a logical consequence of its successor requires empirically grounded correspondence rules rather than *a priori fiat.* In short, the derivability, consistency, and meaning invariance conditions do not sit as well on the Hempelian Received View as they sat on the Nagelian; but it would be an exaggeration to claim that the three conditions have little to do with the Hempelian Received View.

[1] In the foregoing pages I try to spell out – and distinguish - exactly what meaning invariance means for observation and theoretical terms.

This misfit between Hempel and the three conditions does not matter very much, for Feyerabend, in opposing the three conditions, addresses examples in which Hempel says the three conditions *do* apply. Feyerabend demonstrates how Hempel's (and Nagel's) previously accepted examples of theory reduction and explanation *fail* to meet the requirements which reduction and explanation imply (namely, derivability, consistency, and meaning invariance). Feyerabend's argument from example, along with his other two kinds of argument against the Received View, presuppose that inconsistencies at the theoretical level work through to the observational level; this contrasts with the logical empiricist view that observational terms inform the theoretical ones. Feyerabend's arguments against the Received View therefore rely on the view that there are internal and holistic constraints on the meanings of scientific terms.

The semantic IT claims that there are some general successive theories such that:
(1) the derivability, consistency and meaning invariance conditions do not hold, even within a common domain.
The IT also claims that:
(2) the three above conditions fail to hold because of meaning holism
and that
(3) such meaning holism allows that (parts of) successor theories may be logically independent of their predecessors, even in the common domain.
Feyerabend's attacks on the Received View focus on showing how some successive, empirically adequate theories are inconsistent within a common domain. Since theoretical inconsistency is the thrust of his arguments, Feyerabend is wrong to claim that such inconsistent theories are logically independent.

The theory of meaning which Feyerabend proposes in support of semantic internalism and holism, the Pragmatic Theory of Observation (PTO), is found wanting.[3] But this is not sufficient cause to dismiss the IT.

---

[2] Since new observations, experiments, or measuring instruments will have been used.

[3] Also found wanting is Feyerabend's claim that the PTO proposes the indeterminacy of translation thesis; but a number of other Quinean elements are found in the PTO.

Since the arguments supporting (1) (and (3)) have presupposed (2), a largely externalist view of meaning, if convincing, would fatally undermine the IT. Causal theories of reference have been used with this purpose in mind. However, I judge that the CTRs of Putnam, Devitt, Kitcher, and Psillos fail to convince for a number of reasons, many of which revolve around the need for some "structural component"[4] in addition to ostension.[5]

These, and other criticisms of using theories of reference to combat the IT, lead me to wonder whether a theory of reference based on the view that there is only one true way to separate individuals into kinds is the wrong way to think of reference. Feyerabend seems to wonder this too. An alternative would be a theory of reference which embraces the view that there are various – even mutually inconsistent – ways of correctly dividing the world into kinds. Under this approach to reference, two competitor theories could be true in a common domain and yet each theory could refer to many objects simply not referred to at all - or objects whose existence is even forbidden - by the other theory. This pluralist understanding of reference might then further explicate the semantic IT by showing how something *like* (3) can be proposed in conjunction with (1) and (2). However, I make no attempt to give any further details of such a theory of reference.

An other way to attack (3) is to argue about truth rather than reference. If two theories share a common domain but are logically independent of each other, then the truth of one will imply nothing about the truth of the other theory. Donald Davidson argues that, were this the case, translation from one theory to the other would fail; and that permanent and principled failure of translation between two natural languages is unacceptable; so theories cannot exhibit in a common domain the logical independence asserted in (3). I think that Davidson's argument is convincing, but I argue that nontranslatability is not, or ought not to be considered, a part of the semantic IT.

---

[4] Devitt, M. & Sterelny, K. (1987), quoted in Stanford, P. Kyle & Kitcher, P. (2000), p. 100.

[5] "Something about the mental state of the grounder must determine which putative nature of the sample is the one relevant to the grounding, and should it have no such nature the grounding will fail. It is very difficult to say exactly what determines the relevant nature." ibid., p. 101.

I outline a Davidsonian account[6] of the IT in which (3) is replaced by temporary (but possibly long-term) failure to translate correctly. In this scenario, chronic mistranslation on the interpreter's part leads the interpreter to regard (some of) the interpretee's inferences as irrational; and this, in turn, perpetuates the interpreter's view that the interpretee is not making sense. I take it that this captures something of the gist of (3) (or what (3) ought to be claiming in stead of logical independence).

It seems to me that researching the semantic claims of the IT still has plenty of mileage. Element (1) of the IT is now generally accepted and plenty of work is being done on internalism and holism (element (2)). As for finding adequate semantic substitutes for (3), fewer suggestions have been forthcoming. The tendency here is to offer epistemological alternatives to (3). But if the semantic IT is a valid research programme, as I think it is, then its value and interest surely lies in finding out what it is about *language* that makes so fraught with difficulty our understanding another's general theory or world-view.

[6] Developing ideas of Bjørn T. Ramberg.

Bibliography

1. Works By Paul Feyerabend

(1958). 'An Attempt at a Realistic Interpretation of Experience', *Proceedings of the Aristotelian Society*, 58. Reprinted in PP1, pp. 17 – 36.

(1960a). 'Review of N.R.Hanson, *Patterns of Discovery*,' Philosophical Review, 69, pp. 247 – 52.'

(1960b). 'On the Interpretation of Scientific Theories, *Proceedings of the 12th International Congress in Philosophy*, vol. 5, reprinted in PP1, pp. 37 – 43.

(1960c). 'The Problem of the Existence of Theoretical Entities'. Translated by Daniel Sirtes and Eric Oberheim, in PP3.

(1962). 'Explanation, Reduction and Empiricism', in H. Feigl & G. Maxwell (eds), *Scientific Explanation, Space, and Time: Minnesota Studies in the Philosophy of Science*, vol. 3. (Minneapolis: University of Minnesota Press), pp. 28 – 97. [Also in PP1]

(1963a). 'Materialism and the Mind – Body Problem,' The Review of Metaphysics, 17. Reprinted in PP1, pp. 161 – 75.

(1963b). 'How to be a Good Empiricist: A Plea for Tolerance in Matters Epistemological', in B. Baumrin (ed.), *Philosophy of Science: The Delaware Seminar*, vol. 2. (New York: Interscience Press), pp. 3 – 39.

(1964a). 'Review of N.R. Hanson, *The Concept of the Positron*', Philosophical Review, 73, pp. 264-6.

(1964b). 'The Structure of Science', BJPS, 16, pp. 237ff. Reprinted in PP2, pp. 52 – 64.

(1965a). 'Problems of Empiricism' in R.G. Colodny (ed.), *Beyond the Edge of Certainty*, University of Pittsburgh Studies in the Philosophy of Science (Englewood Cliffs, NJ: Prentice Hall), pp. 145 – 260.

(1965b). 'On the Meaning of Scientific Terms', *Journal of Philosophy*, 62, pp. 266-74. [Also in PP1]

(1965c). 'Reply to Criticism: Comments on Smart, Sellers, and Putnam', in R.S. Cohen & M.W. Wartofsky (eds.), *Boston Studies in the Philosophy of Science*, vol. 2: In Honor of Philipp Frank (New York: Humanities Press), pp. 223 – 61. [Also in PP1]

(1965d). 'Review of K.R. Popper, *Conjectures and Refutations*', *Isis*, 56

(1968a). 'Science, Freedom, and the Good Life', *Philosophical Forum*, 1, pp. 127 – 35.

(1968b). 'Outline for a Pluralistic Theory of Knowledge and Action', in S. Anderson (ed.), *Planning for Diversity and Choice* (Cambridge, Mass.: MIT Press), 275 – 84.

(1969a). 'Science Without Experience', *Journal of Philosophy*, 66, Reprinted in PP1, pp. 132 – 5.

(1970). 'Consolations for the Specialist', *Criticism and the Growth of Knowledge*, Imre Lakatos & Alan Musgrave (eds.) (Cambridge: CUP), repreinted in PP2, pp. 131 – 167.

(1975). Against Method. (London: Verso).

(1978). *Science in a Free Society*, (London: New Left Books).

[PP1] (1981). *Realism, Rationalism and Scientific Method*, Philosophical Papers volume 1. (Cambridge: CUP).

[PP2] (1981). *Problems of Empiricism*, Philosophical Papers volume 2. (Cambridge: CUP).

(1987) Farewell to Reason. (London: Verso).

(1993). *Against Method*, 3rd edition. (London: Verso).

(1995). *Killing Time*. London: University of Chicago.

(1999a). *The Conquest of Abundance*. (Chicago: Chicago University Press).

(1999b). *For and Against Method*. Matteo Motterlini (ed.). (Chicago: Chicago University Press).

[PP3] (1999c). *Knowledge, Science and Relativism*, Philosophical Papers 3. John Preston (ed.). (Cambridge: CUP).

2. Other Works

Abbot, Barbara (1999). 'Water = H_2O', *Mind*, 108, pp. 145 – 8.

Achenstein, P. (1964). 'On the Meaning of Scientific Terms', *Journal of Philosophy*, 61, pp. 475 – 510.

Aurelius, Marcus (1964). *Meditations*. Translated by Maxwell Staniforth. (London: Penguin).

Bishop, Michael A. (1991). 'Why the Semantic Incommensurability Thesis is Self-Defeating', Philosophical Studies, 63, pp. 343-56.

Bishop, Michael A. & Stich, Stephen P. (1998). 'The Flight to Reference, or How Not to Make Progress in the Philosophy of Science', *Philosophy of Science* , 65, pp. 33 – 49.

Blackburn, Simon (1984). *Spreading the Word.* (Oxford: Clarendon Press).

Boyd, R., Gasper, P., & Trout. J.D. (eds.) (1991). *The Philosophy of Science.* (Cambridge, Massachusetts: MIT Press).

Brown, Harold I. (1983). 'Incommensurability', *Inquiry,* 26, pp. 3-30.

Brown, Harold I. (1984). Book Review of Feyerabend [PP1], *International studies in Philosophy,* 16, pp. 90-3.

Burton, John (1969). *Conflict and Communication: The Use of Controlled Communication in International Relations.* (London: Macmillan).

Butts, R.E. (1966). 'Feyerabend and the Pragmatic Theory of Observation', *Philosophy of Science,* 33.

Chang, Ruth (1998). *Incommensurability, Incomparability and Practical Reason.* (USA: Harvard University Press).

Coffa, J.A. (1967). 'Feyerabend on Explanation and Reduction', *The Journal of Philosophy,* 64, pp. 500 – 8.

Couvalis, George (1989). *Feyerabend's Critique Of Foundationalism.* (England: Avebury).

Davidson, Donald (1984). *Inquiries into Truth and Interpretation.* (England: Clarendon Press), pp. 183-198.

Davidson, Donald (1986a). 'A Coherence Theory of Truth and Knowledge', in Lepore (ed.) (1986).

Davidson, Donald (1986b). 'A Nice Derangement of Epitaphs', in Lepore (ed.) (1986).

Davidson, Donald (1993). Replies in Stoeker, Ralf (ed.) (1993).

Davidson, Donald (1999). Replies in Hahn (ed.) (1999).

Devitt, M. (1979). 'Against Incommensurability', *Australian Journal of Philosophy,* 57, pp. 29 – 50.

Devitt, M. (1981). *Designation.* (USA: Columbia University Press.)

Devitt, M. (1991). *Realism and Truth,* 2nd edition. (New Jersey: Princeton University Press).

Devitt, M. & Kim Sterelny (1987). *Language and Reality.* (Cambridge, Massachusetts: MIT Press)

Donnellan, Keith (1977). 'Reference and Definite Descriptions' in Schwartz (ed.) (1977).

Dupre, John (1993). *The Disorder of Things.* (Cambridge, Massachusetts: Harvard University Press)

Dupre, John (2001). *Human Nature and the Limits of Science.* (Oxford: Clarendon Press).

Earman, John, & Fine, Arthur (1977). 'Against Indeterminacy', *Journal of Philosophy*, 74, pp. 535 - 8.

Evnine, Simon (1991). *Donald Davidson.* (California: Stanford University Press).

Field, Hartry (1973). 'Theory Change and the Indeterminacy of Reference', *Journal of Philosophy*, 70, pp. 462 - 481.

Fine, Arthur (1967). 'Consistency, Derivability and Scientific Change', *The Journal of Philosophy*, 64, pp.231 - 40

Fine, Arthur (1975). 'How to Compare Theories: Reference and Change', *Nous* 9, pp. 17 - 32.

Fine, Arthur (1991). 'The Natural Ontological Attitude', in Boyd, Gasper and Trout (1991), pp. 261 – 277.

Føllesdal, Dagfinn. (1973). 'Indeterminacy of Translation and Under-Determination of the Theory of Nature', *Dialectica*, 27, pp. 289-301.

Giedymin, J. (1971). 'The Paradox of Meaning Variance', *British Journal for the Philosophy of Science*, 21, pp. 30 – 48.

Gillies, Donald (1993). *Philosophy of Science in the Twentieth Century.* (Oxford: Blackwell).

Goosens, William K. (1977). 'Underlying Trait Terms', in Schartz (1977). pp. 133-154.

Grayling (1997). *An Introduction to Philosophical Logic.* 3rd edition. (Oxford: Blackwell).

Hacking, Ian (1975), *Why Does Language Matter To Philosophy?* (Cambridge: CUP).

Hacking, Ian (1982). 'Language, Truth and Reason' in Hollis, M. & S. Lukes (eds.) (1982).

Hacking, Ian (1983). *Representing and Intervening.* (Cambridge: CUP).

Hacking (1993). 'Working in a New World: The Taxonomic Solution', in Paul Horwich (ed.). *World Changes.* (London: MIT Press).

Hahn, Lewis Edwin (ed.) (1999). *The Philosophy of Donald Davidson.* (Chicago and La Salle, Illinois: Open Court).

Hempel, Carl G. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science.* (New York: Macmillan).

Hempel, Carl G. (1966). *Philosophy of Natural Science.* (New Jersey: Prentice-Hall).

Hesse, M. (1968). 'Fine's Criteria of Meaning Change', *The Journal of Philosophy*, 65, pp. 46 – 52.

Hesse, M. (1983). Comment on Kuhn's 'Commensurability, Comparability, Communicability', in P.D. Asquith and T. Nickles (eds.), *PSA* 1982, vol. 2. (Michigan: Philosophy of Science Association). pp. 704 – 11.

Hollis, M. & S. Lukes (eds.) (1982). *Rationality and Relativism.* (Oxford: Blackwell).

Hookway, Christopher (1988). *Quine.* (Cambridge: Polity Press).

Horwich, Paul. (1998). *Meaning.* (Oxford: OUP).

Hoyningen-Huene, P. (1993). *Reconstructing scientific Revolutions: Thomas S. Kuhn's Philosophy of Science.* (London: University of Chicago Press).

Hoyningen-Huene, P., (2000) 'Paul Feyerabend and Thomas Kuhn.' *The Worst Enemies of Science? Essays in Memory of Paul Feyerabend.* J. Preston, G. Munevar, D. Lamb. (eds.). (Oxford: OUP). pp. 102 – 114.

Hoyningen-Huene, P., E. Oberheim, & H. Andersen (1996). 'On Incommensurability', *Studies in History and Philosophy of Science*, 27, pp. 131 – 141.

Hoyningen-Huene, P., Howard Sankey. (eds.) (2001). *Incommensurability and Related Matters.* Boston Studies in the Philosophy of Science, vol. 216. (Dordrecht: Kluwer).

Hull, R.T. (1972). 'Feyerabend's Attack on Observation Sentences', *Synthese*, 23, pp. 374 – 99.

Kitcher, Philip (1978). 'Theories, Theorists and Theoretical Change', *The Philosophical Review*, 87, pp. 519 – 547.

Kitcher, P. (1983). 'Implications of Incommensurability', in P.D. Asquith & T. Nickles (eds.), *PSA* 1982, vol. 2, Philosophy of Science Association, East Lansing, Michigan, pp. 689 – 703.

Kraut, Robert (1986). 'The Third Dogma', in Lepore (ed.) (1986).

Kroon, Frederick W. (1985). 'Theoretical Terms and the Causal View of Reference', *Australian Journal of Philosophy*, 63, pp. 143 – 166.

Kroon, Frederick W. (1987). 'Causal Descriptivism', *Australian Journal of Philosophy*, 65, pp. 1 – 17.

Kuhn, Thomas (1996). *The Structure of Scientific Revolutions*, 3rd edition. (London: University of Chicago Press).

Lambert, K. & Brittan, Gordon G. (1992). *An Introduction to the Philosophy of Science.* 4th edition. (California: Ridgeview Publishing Company).

Laudan, Larry (1996). *Beyond Positivism and Relativism.* (Oxford: Westview Press).

Leeds, Stephen (1997). 'Incommensurability and Vagueness', *Nous*, 31, pp. 385-407.

Leplin, J. (1969). 'Meaning Change and the Comparability of Theories', *British Journal for the Philosophy of Science*, 20, p. 69 – 75.

Leplin, Jarret (1979). 'Reference and Scientific Realism', *Studies in History and Philosophy of Science*, 10, pp. 265 – 284

Lepore, Ernest (ed.) (1986). *Perspectives on the Philosophy of Donald Davidson.* (Oxford: Basil Blackwell).

Martin, Michael (1971). 'Referential Variance and Scientific Objectivity', *BJPS*, 22, pp. 17-26.

McGinn, Colin (1977). 'Charity, Interpretation and Belief', *Journal of Philosophy*, 74, pp. 521-35.

McGinn, Colin (1986). 'Radical Interpretation and Epistemology' in Lepore (ed.) (1986).

McGowan, Mary Kate (1999). 'The Metaphysics of Squaring Scientific Realism With Referential Indeterminacy', *Erkenntnis*, 50, pp. 87 – 94.

Mellor, D.H. (1996). 'Natural Kinds', in Pessin, Andrew, & Goldberg, Sanford, (eds.) *The Twin Earth Chronicles.* (London: M.E. Sharpe).

Miller, Alexander (1998). *Philosophy of Language.* (London: UCL Press).

Miller, A.I. (1991). 'Have Incommensurability and Causal Theory of Reference Anything to Do with Actual Science?', - incommensurability, no; causal theory, yes.' *International Studies in the Philosophy of Science*, 5, pp. 97 – 108.

Musgrave, Alan (1989). 'NOA's Ark – Fine For Realism', *The Philosophical Quarterly*, 39, pp. 383-98.

Nersessian, N & Andersen, H. (1997). 'Conceptual Change and Incommensurability', *Danish Yearbook of Philosophy*, 32, pp. 111-51.

Næss, A. (1964). 'Pluralistic Theorizing in Physics and Philosophy', *Danish Yearbook of Philosophy*, 1.

Newton-Smith, W. (1996). *The Rationality of Science.* (London: Routledge).

Newton-Smith, W. (1982). 'Relativism and the Possibility of Interpretation' in Hollis, M. & S. Lukes (eds.) (1982).

Nickles, T. (1973). 'Two Concepts of Inter-Theoretic Reduction', *Journal of Philosophy*, 70, pp. 181 – 201.

Oberdan, T. (1990). 'Positivism and the Pragmatic Theory of Observation', in Fine, A., M. Forbes & L. Wessels (eds.) (1990), pp. 25 – 37.

Oberheim, E. & Hoynengen-Huene, P. (1997). 'Incommensurability, Realism and Meta-Incommensurability', *Theoria* (Spain) 12, pp. 447 – 465.

Oberheim, E. & Hoynengen-Huene, P. (2000). 'Feyerabend's Early Philosophy', review of Preston (1997). *Studies in the History and Philosophy of Science*, 31, pp. 363 – 75.

Papineau, David (1979). *Theory and Meaning*. (Oxford: Clarendon Press).

Papineau, David (1996). 'Theory-Dependent Terms', *Philosophy of Science*, 63, pp. 1 – 20.

Perovich, Anthony H. Jr. (1991). 'Incommensurability, Its Varieties and Its Ontological Consequences', in Munevar, G, *Beyond Reason, Boston Studies in the Philosophy of Science*, vol. 132, (Kluwer Academic Publishers: Dordrecht, The Netherlands), pp. 313-328.

Platts, Mark (1979). *Ways of Meaning*. (London: Routledge and Keegan Paul).

Popper, Karl R. (1959). *The Logic Of Scientific Discovery*. (London: Routledge).

Preston, John (1989). 'Folk Psychology as Theory or Practice? The Case for Eliminative Materialism. *Inquiry*, 32, pp. 277-303.

Preston, John (1992). Book Review of Couvalis (1989), *International Studies in the Philosophy of Science*, 6, pp. 155-8.

Preston, John (1997a). *Feyerabend: Philosophy, Science and Society*. (Cambridge: CUP).

Preston, John (1997b). Feyerabend's Polanyian Turns, *Appraisal*, 1, Supplementary Issue, pp. 30-6.

Preston, John (1997c), 'Feyerabend's Retreat from Realism', *Philosophy of Science*, 64 (Supplement, Part 2, Symposia Papers), Lindley Darden (ed.) (USA: PSA), pp. 421-31.

Psillos, Stathis (1997). 'Kitcher on Reference', *International Studies in The Philosophy of Science*, 11, pp. 259 - 72.

Psillos, Stathis (1999). *Scientific Realism*. (London: Routledge.)

Putnam, H. (1975). *Mind, Language and Reality: Philosophical Papers* vol.2. (Cambridge: CUP).

Putnam, H. (1977a). 'Is Semantics Possible?', in Schwartz (ed.) (1977).

Putnam, H. (1977b). 'Meaning and Reference', in Schwartz (ed.) (1977). pp. 119-132.

Putnam, H. (1981). *Reason, Truth and History*. (Cambridge: CUP)

Putnam, H. (1990). *Realism with a Human Face*. (Cambridge, Massachusetts: Harvard University Press).

Putnam, H. (1996). 'The Meaning of Meaning', in Pessin, Andrew, & Goldberg, Sanford, (eds.) *The Twin Earth Chronicles.*(London: M.E. Sharpe).

Quine, W.O. (1981). *Theories and Things.* (Massachusetts: Harvard University Press).

Quine, W.O. 'Two Dogmas of Empiricism', in *From A Logical Point of View, 2nd edition.* (Massachusetts: Harvard University Press).

Ramberg, Bjørn T. (1989). *Donald Davidson's Philosophy of Language.* (Oxford: Basil Blackwell).

Rorty, Richard (1980). *Philosophy and the Mirror of Nature.* (Oxford: Blackwell).

Rorty, Richard (1986). 'Pragmatism, Davidson and Truth', in Lepore (ed.) (1986).

Russell, B. (1917). *Mysticism and Logic.* (New Jersey: Barnes and Noble Books).

Sankey, H. (1991). 'Feyerabend and the Description Theory of Reference', *Journal of Philosophical Research*, 16, pp. 223 – 232.

Sankey, Howard (1994). *The Incommensurability Thesis.*(England: Avebury).

Sankey, Howard (1997). 'Kuhn's Ontological Relativism', *Issues and Images in the Philosophy of Science.* D. Ginev and R.S. Cohen (eds.) (Dordrecht: Kluwer). pp. 305 – 320.

Sankey, Howard (1998) 'Taxonomic Incommensurability', *International Studies in the Philosophy of Science*, 12, pp. 7 – 16.

Sankey, Howard, 'The Language of Science: Meaning Variance and Theory Comparison'. http://www.hps.unmelb.edu.au/student/biographies/howardpaper7.htm

Schaffner, K. (1967). 'Approaches to Reduction', *Philosophy of Science*, 34, pp. 137 – 147.

Schwartz, Stephen P. (ed.) (1977). *Naming, Necessity and Natural Kinds.* (Cornell University Press).

Shapere, D.S. (1966). 'Meaning and Scientific Change', in R.G. Colodny (ed.), *Mind and Cosmos.* (University of Pittsburg Press: Pittsburg USA). pp. 41 – 85.

Shapere, D. (1982). 'Reason, Reference and the Quest for Knowledge', *Philosophy of Science*, 49, pp. 1-23.

Shapere, Dudley (1989). 'Evolution and Continuity in Scientific Change', *Philosophy of Science*, 56, pp. 419-37.

Sklar, L. (1967). 'Types of Inter-Theoretic Reduction', *British Journal for the Philosophy of Science*, 18, pp. 109 – 24.

Sontag, Susan (1966). *Against Interpretation and Other Essays.* (New York: Farrar, Strauss & Giroux).

Stanford, P. Kyle & Kitcher, Philip (2000). 'Refining the Causal Theory of Reference for Natural Kind Terms', *Philosophical Studies 97*, pp. 99.129.

Sterelny, Kim. (1983). 'Natural Kind Terms', in in Pessin, Andrew, & Goldberg, Sanford, (eds.) *The Twin Earth Chronicles.* (London: M.E. Sharpe).

Stich, Stephen (1990). *The Fragmentation of Reason.* (London: MIT Press).

Stoeker, Ralf (ed.) (1993). *Reflecting Davidson.* (Berlin: Walter de Gruyter).

Suppe, Frederick (ed.). (1977). *The Structure of Scientific Theories*, 2nd edition. (Urbana and Chicago: University of Illinois Press).

Suppe, Frederick (1991). 'The Observational Origins of Feyerabend's Anarchistic Epistemology', in Munevar, G, *Beyond Reason, Boston Studies in the Philosophy of Science*, vol. 132, (Kluwer Academic Publishers: Dordrecht, The Netherlands), pp. 297 – 311.

Tarski, Alfred (1944). 'The Semantic Conception of Truth and the Foundations of Semantics'. http://www.ditext.com/tarski/tarski.html

Taylor, Kenneth (1998). *Truth and Meaning.* (Oxford: Blackwell).

Townsend, B. (1971). 'Feyerabend's Pragmatic Theory of Observation and the Comparability of Alternative Theories', *Boston Studies in the Philosophy of Science,* 8, pp. 202 – 11.

Werth, R. (1980). 'On the Theory-Dependence of Observations', *Studies in the History and Philosophy of Science*, 11, pp. 137 – 43.

Zahar, E. (1982). 'Feyerabend on Observation and Empirical Content', *BJPS*, 33, pp. 397 – 433.

Zemach, Eddy M. (1996). 'Putnam's Theory on the Reference of Substance Terms', in Pessin, Andrew, & Goldberg, Sanford, (eds.) *The Twin Earth Chronicles.* London: M.E. Sharpe.