Empiricism and Philosophy

Nathan Sinclair

Degree: Doctor of Philosophy

The University of Sydney

2008

This thesis is submitted in partial fulfillment for the degree of doctor of philosophy and is entirely my own work except where indicated in the references.

To Kate,
with and without whom I could not have written this

Contents

III Philosophy   203

Introduction

> We ought not then be lead astray by the contentious argument you quoted. It would make us lazy, and is music in the ears of weaklings. The other doctrine produces energetic seekers after knowledge; and being convinced of its truth, I am ready, with your help, to inquire into the nature of virtue.

> (Socrates in *the Meno*)

In nineteen ninety-eight I experienced a physical, an emotional and an intellectual crisis. I was beaten and stabbed[1] four times on Petersham station, I separated from the mother of my first two (and then only) children, and I lost faith in the analytic-synthetic distinction. Though it dominated for a short period of time, the physical pain soon subsided; and as it gradually became clear that my children were going to largely recover from the harm inflicted upon them in my separation with their mother, so did the emotional. The longest enduring crisis, and the one which over time caused me the greatest unhappiness, was my loss of faith in the analytic-synthetic distinction.

The ankle-tap which lead to my rejection of the analytic-synthetic distinction was provided by Michael McDermott in a course titled "Ayer and Quine". That course introduced me to Quine's polemic against the existence of a distinction between analytic and synthetic truths. Till that time I had held a constellation of beliefs that may be best tersely described as a form of Australian realism; and, influenced by David Armstrong, I took Wittgenstein's Tractatus Logico-Philosophicus to be one of the central texts of that school. (At the time I was unaware of the work's latent phenomenalism[2]) Even at the time the Tractatus was not taken as holy writ; Armstrong held that the atoms into which possible states of affairs could be divided were to be discovered by empirical investigation rather than by analysis of the meanings which were these atom's semantic counterparts. For my own part I held a less rigid view of atoms, believing that

---

[1] with a key protruding through the knuckles

[2] I now have the impression that the only thing stopping Wittgenstein from an open admission that the atomic facts were sensory was, to borrow his term, an inability to make sense of the "grammar" of the phrase 'is an atomic fact'.

like many mathematical spaces, if the space of possible worlds had at least one basis or spanning set of atomic elements then it had at least countably many such sets. Like Armstrong however, I thought that such spanning sets, and hence the range of space that was to be spanned, were to be discovered in the first instance by scientific empirical investigation.

My loss of faith in the analytic-synthetic distinction was all the more painful because I had not realised that it had been based upon faith at all. Eventually, after much struggle to find a defence of the distinction, and despite finding many flaws in all the oft-recited arguments against its existence, I came at last to realise that my own belief in the analytic-synthetic distinction had been based upon the cardinal Wittgensteinian sin of presuming that my intuitions came with labels attached. This loss did not occur overnight, nor without much searching about for alternative justifications, and my master's thesis, written in the dying days of my faith, was a disconnected hodge-podge of various attempts to revive the distinction.

My final rejection of the distinction came at a high price. Prior to rejecting it I had taken the production of (interesting) valid arguments to be an essential part of philosophical practice, and thought validity to be analyticity of the conditional. I was thus left without any account of why anyone should listen to anything I had to say, nor why I should listen to the pronouncements of others (particularly other philosophers). Scientific theories could still be recognised as having some merit, for they gave issue to predictions that turned out to be palpably correct, (though how this merit might be recognised ahead of time remained a mystery) but it seemed that philosophers had nothing of value to add to scientist's brute practices of modifying their theories in the light of accumulating evidence. Getting the right answer, making predictions that came true was the only standard by which theories were to be judged, and all the rest was suspected of being hot air.

Though such a state is not conducive to wide reading, I scanned the literature avidly for any hope of an alternative. Just because I did not have any account of the distinction between good and bad arguments, nor my ability to discern the difference, that did not imply that there was no such distinction to be made. (Concern that my only reason for believing this was an unfounded intuition that the contrary conditional was invalid was put aside for practical purposes.) At one time we were unable to account for the ability of chicken-sexers to discriminate between male and female chicks but chicken-sexers had the ability nonetheless and their exercise of it was a useful and productive enterprise, though at least in their case there was an account of why it was useful.

My investigation into the literature in hopes of an alternative account of the epistemic basis by which the valid arguments of a novel foreign tongue might

be identified revealed, to my surprise, that though the rejection of the analytic-synthetic distinction was almost universal and widely heralded, I was almost alone in being troubled by it. In hindsight I believe that this almost universal and curiously undogmatic slumber has two main historical causes: The first is the wide-spread rejection of the very idea of a sharp distinction between science and philosophy (which in turn is largely due to Quine's arguments that "One effect of abandoning them [i.e. the analytic-synthetic distinction and reductionism] is, as we shall see, a blurring of the supposed boundary between speculative metaphysics and natural science"), and the second was Lewis' account of philosophical methodology as the systematisation and simplification of our pre-existing intuitions and the consequent extraordinary popularity of beginning philosophical papers with appeals to intuition.

One of the oddities of the historical role of the belief that there is no sharp distinction between science and metaphysics is the extraordinary discrepancy between the arguments which have been taken to justify it, and the conclusions that philosophers have drawn from it. Though the most influential arguments for this belief are Quine's, and are entirely epistemic and based upon empiricist principles, philosophers have used this belief to justify the rankest unempirical metaphysical speculation. Quine's point was that metaphysical speculations were to be judged by the same standard as scientific speculations — i.e. whether they contributed to the development of the simplest and strongest overall empirical theory fitting with our experiences. Good metaphysics, like good science, simplifies and clarifies, and bad metaphysics, like bad science, either breaks the fit with experience or pointlessly muddies the waters. The conclusion that many philosophers citing Quine as their authority for this belief appear to have drawn from it however, is that the most unempirical speculation deserves the same respect and credibility as science.

Though the central problem struck me as one of justifying my philosophical practice, it can at least be raised without any assumptions as to the nature of philosophy nor indeed that philosophical practice is tied together by any essential feature at all. Suppose there is an English speaking slave boy locked in a barren and featureless room with a device emitting random strings of English words. Once the boy has come to terms with his plight, the vast majority of the strings emitted by the device have little effect on his beliefs; some however, cause striking revisions. In a suitably attentive and intelligent slave boy with just the right degree of previous education the string

> If the line AB is a radius of the circle centered at A and the line BC
> is a radius of the circle centered at C and AC is a radius of both
> the circle centered at A and the circle centered at C, then AB is

> the same length as AC since they are both radii of the same circle,
> and BC is the same length as AC since they are both radii of the
> same circle, and AB is the same length as BC since they are both
> the same length as AC and therefore all three sides of the triangle
> ABC are of the same length.

is apt to cause the belief that the sides of triangles inscribed between the
centres and a point of intersection of two circles sharing a radius are all the same
length. Perhaps the slave boy has not entertained any opinion on the matter
prior to being thus taught that it is true, and exposure to this string only causes
him to add to his stock of beliefs, but exposure to other strings is apt to cause
outright revision of his previously held opinions on such matters as the length
of the sides of a square double the area of a unit square, or whether justice is
the interests of the stronger.

The boy's surroundings do not, at least in the ordinary sense of the term,
provide evidence for some of the beliefs he adopts in the revisions occasioned
by such strings nor against the beliefs he abandons. And many of the beliefs he
revises do not concern the mechanism by which the sequences of words to which
he is exposed are produced nor (in some sense) his immediate surroundings.
Despite this we, his English speaking peers, hold some of the beliefs he adopts
as a result of exposure to these strings to be justified and others to be unjus-
tified. If he believes that there are flying cars as a result of hearing the string
"There are flying cars" or that there are horseless carriages as a result of hearing
"There are horseless carriages" his opinion is unjustified, but exposure to some
strings produces, by our lights at least, not only true belief but knowledge. The
question that immediately arises is whether there is any rhyme or reason to our
assessments.

Socrates' suggestion was that the justified beliefs that might be produced by
such means were those recalled from a previous existence, and that slave boys
were justified in believing them because they had known them in a previous
incarnation when they were not in human form. Even apart from doubts over
re-incarnation, one is left wondering how to tell false memories from true, and
what justifies these beliefs in the first place.

Another response (consonant with the belief that the line between science
and philosophy is blurred) is to attack the question as ill-conceived, and any
answer to it as an insignificant artefact of the way the question is set up. The
vitiating assumption of the example, according to the most popular and cogent
version of this criticism, is that concerning the boy's previous education. Vary
this prior evidence (or collateral information) and the range of beliefs that the
boy can be induced to adopt by exposure to strings of words in his cell can be

made to vary as widely as one wishes (and hence is of no further significance). The beliefs he adopts as a result of such confined exposures are still based upon empirical evidence, albeit largely the empirical evidence he was exposed to before his incarceration. There is thus no sharp line to be drawn between those beliefs that are due purely to the slave boy's linguistic competence together with his exposure to strings of words, and those that are also due in part to his other experiences both prior to his incarceration and during it (for even the dullest oubliette continues to provide evidence of something). The two justifying factors of previous experience and linguistic competence are inextricably intermingled in each of the beliefs he revises and hence the very idea of distinguishing a class of beliefs whose adoption can be justified (for competent speakers) by exposure to an argument or sequence of words alone is mis-conceived.

Still, against this criticism it can be held that some of the revisions the boy may be induced to adopt do not alter his empirical expectations. New beliefs are adopted and old rejected but the empirical content of the whole system is left unchanged. Release the slave boy, let him roam about and perform experiments as he will, no observation, no experience, could confirm the theory he held before making such a revision that did not also confirm the theory he was disposed to hold afterwards, nor infirm one and not both, yet when isolated in his prison cell (or even worse locked in with Socrates) the merest line of dialogue not only causes him to make such revisions, but justifies them as well.

At this point it might be objected that a change in belief without change in empirical content is no change at all, and that all that is altered in such revisions is the form in which the slave boy expresses his unchanging beliefs (i.e. the range of sentences to which the boy is disposed to assent). Even so, the question remains of just what was wrong with the boy's prevous expression of his beliefs. Are the norms breached by defective expressions of theories systematic, capable of being derived from a finite or recursive set of elementary principles; and are these principles categorical and primitive or can they be derived from other norms which already command our allegiance? To what extent is the choice between different expressions of the same empirical content arbitrary and conventional, and driven only by the benefits of convention? These are the questions, which first struck me as questions about philosophical practice, to which this thesis is devoted.

My objections to the currently fashionable philosophical practice of setting out to capture intuitions may best be quickly illustrated, though not substantiated, by a good example. By my lights the most egregious appeals to intuition are those in which the only justification offered for deferring to these intuitions is the citation of someone else doing so. The following is a prime example taken from the fifth slide of the last seminar I attended before writing this, which I

went to by chance and without foreknowledge of its title or contents.

> The Indeterminacy Intuition: Since the assertion is not now de-
> termined to be true/false then the assertion lacks a truth value.
>
> Determinacy Intuition One: At the end of the next day after the
> sea-battle has/has not happened, the assertion was true/false.
>
> Constraint: A good solution must respect both intuitions (cf.
> MacFarlane 2003 p. 322)

The respect given to these intuitions deserves emphasis. In this case, as
in many others, such declarations served to establish the rules of the ensuing
philosophical game, and the assurance that the game had a point was deferred
to cited authority. On turning to page 322 of *The Philosophical Quarterly* (*!*)[3]
our hopes of a justification for these essential constraints are further deferred
but at least we are assured of an argument for them: "Some approaches save
the indeterminacy intuition, others the determinacy intuition. In SSII and SS3 I
discuss these one-sided approaches and argue that they are all unsatisfactory. A
satisfactory account of future contingents must give both intuitions their due."[4]

Turning to section II (titled *The Indeterminacy Intuition*), the closest there
is to an argument for the eponymous intuition is given in the first paragraph
which I quote in full:

> Suppose that at some moment m_0 there is an objectively possi-
> ble future history h_1 in which there is a sea battle the next day, and
> another h_2 in which there is no sea battle the next day. These al-
> ternative histories are both objective possibilities, not just epistemic
> ones. It is not just that we do not know whether or not there will be
> a sea battle, or even that we could not know, but that both courses
> of events are real possibilities. Whether the world is objectively in-
> deterministic in this sense is, of course, a substantive scientific (and
> perhaps metaphysical) question. I do not here presuppose an affir-
> mative answer to this question. All I am presupposing is that talk
> about the future would not be incoherent in an objectively inde-
> terministic world. Determinism may be true, but it is not for the
> semanticist to say so.

To strip away the verbiage, the proposed justification of the indeterminacy
intuition is that it might be true together with a none too covert appeal to

---

[3]Indeed not only was this paper published in *The Philosophical Quarterly* but it was the
winner of its essay prize for 2002!

[4](MacFarlane, 2003)

the analytic-synthetic distinction. (The relevant tractarian dicta are 2.0211^5 and 2.0212^6). I have shared the author's yearning for the analytic-synthetic distinction, but it's ostensible rejection was what set off this business of appealing to intuitions in the first place. To rely upon this distinction to justify the postulation of objective possible future histories is simply perverse.

MacFarlane's justification of the second intuition in section III (titled *The Determinacy Intuition*) is somewhat better:

> If we think about how to assign a truth value to Jake's utterance at m_0, the indeterminacy intuition seems overwhelming. But now what about someone who is assessing Jakes utterance from some point in the future? Sally is hanging on to the mast, deafened by the roar of the cannon. She turns to Jake and says 'Your assertion yesterday turned out to be true'. Sally's reasoning seems unimpeachable:
>
> > Jake yesterday asserted that there would be a sea battle today
> > There is a sea battle today
> > So Jake's assertion was true.
>
> When we take this retrospective view, we are driven to assign a determinate truth value to Jake's utterance: this is the determinacy intuition.

The proposed justification for the determinacy intuition is that we all strongly believe it to be true, and this it must be admitted, is a pretty good justification as far as it goes. But if it is so widely and powerfully believed then why not simply admit it as an assumption? The only pressing questions left at this point are why anyone bothered to dress up these assumptions as intuitions at all and whether there can be any intellectually respectable motivation for *citing* such so-called justifications.

Of course a single example cannot establish the bankruptcy of a whole methodology, anything can be done badly, but this is not an isolated example. Appeals to intuitions are rife in current philosophical literature and are commonly justified either by covert appeal to the analytic-synthetic distinction or by the citation of previous authorities. Instances include (Philips, 1986), (Wasserman, 2004), (Stroud, 2006), (Reibetanz, 1998) and (Kment, 2006). More is at stake than mere style in these appeals for, at least to judge by the modern literature, the whole edifice of possible worlds and modal logic is raised upon such foundations. (See (Kripke, 1972), (Thomason and Stalnaker, 1968),

---

[5]If they world had no substance, then whether a proposition had sense would depend on whether another proposition was true.

[6]In that case we could not sketch any picture of the world (true or false).

(Lewis, 2001), (Lewis, 1998), (Barcan Marcus, 1968), (Elder, 1998) for examples or citations of the quantification over modal operators being justified upon such intuitive grounds). This philosophy is too credulous for me, for all the talk of simplicity there is little to no agreement on just how much simplicity is sufficient to warrant rejecting an intuition, nor on which theories possess it, and the capturing of intuitions just doesn't seem like much sport. I fell in love with a different philosophy, a critical philosophy capable of overturning our intuitions, not merely capturing them.

Though I found most modern philosophical publications of little assistance, two unpublished manuscripts by Michael McDermott were essential to the development of this thesis. *Quine on Science and Metaphysics* reaffirmed my confidence in the belief that Quine's holism implies the distinction between science and philosophy is perfectly sharp; and *From a psychological point of view* provided a glimpse of what, in chapter 2 of the current work, I have called weak (or non-reductive) empiricism. Where McDermott appeared to disagree with me was in the value of Quine's arguments against the analytic-synthetic distinction. Though I agreed with him that Quine's most overt arguments against the distinction were flawed I thought those flaws were not critical and that there was at least a hint of a solid argument against the analytic-synthetic distinction in the combination of *Two Dogmas of Empiricism* and *Word and Object* and this, together with what I had learnt from my own failed attempts to redeem the distinction, convinced me that it must be abandoned.

The first draft of this thesis was primarily an attempt to assemble a blancmange of vegan substitutes for the analytic-synthetic distinction. Coherence, simplicity, computational complexity, first-third person symmetry, reflective equilibrium, and common-sense were all pressed into service. I was prepared to settle for a motley assortment of reasoning norms rather than an over-arching principle of rationality. The results however were not appetising; even though philosophical disputes could be distinguished I could provide no account of the grounds upon which any but trivial philosophical disputes might be decisively resolved. The specification of the empirical evidence upon which our theories are based also proved more difficult than expected and I was stumped by the lack of inter-subjectively identifiable stimuli.

The second draft pursued the more modest goal of merely clarifying the arguments against the analytic-synthetic distinction and some of the problems attendant upon its rejection. In the course of writing it I realised that the requirement of inter-subjectively identical stimuli was a relic of reductive empiricism and could be abandoned. But the major improvement was in the statement of the (more of less Quinian) argument against the analytic-synthetic distinction. I was still left without an account of the resolution of philosophical disputes

however, until I read the following quotation from Marx's correspondence in David Stove's article *A Farewell to Arts*:

> "It's possible that I shall make an ass of myself. But in that case one can always get out of it with a little dialectic. I have, of course, so worded my proposition so as to be right either way."

Marx may have penned these words to Engels in jest, and to attempt to mount a serious criticism of the theories Marx advanced in earnest and on full reflection based on this excerpt alone would be foolish, but here at last I had an example of intellectual malfeasance that I could make sense of. Furthermore it seemed, and still seems to me, that any theory (or theory expression) that was free from such defects would be one in which the generation of our empirical expectations was made completely explicit, that the lack of any grounds for a decisive resolution to a dispute between proponents of empirically equivalent and fully explicit theories would still leave scope for the resolution of significant philosophical disputes, and that the removal (or at least identification) of such defects would be a useful and productive enterprise.

This third and (near) final version (so let us all hope) consists of three parts. In the first I attempt to specify the fundamental principle of holistic (or weak) empiricism. In the second, I present the main argument against the analytic-synthetic distinction and attempt to give an account of logical truths' immunity to revision which does not depend upon their analyticity, and in the third I argue that (more or less Quinean) holistic empiricism implies there is a perfectly sharp distinction between philosophy and science and propose the grounds upon which significant philosophical disputes may be resolved.

From the point of view of those in the midst of the current tradition, I am aware that it looks as though I am harking back to a past time, before Kripke and Lewis inspired the current enthusiasm for modal logic and possible worlds. "Trapped in twentieth century philosophy" is the phrase that leaps to mind. My supervisor has often enjoined me to regard my thesis as a tale of how philosophy might have gone if it had taken a different course in about 1970. But from my point of view it is current philosophy that is harking back, to concepts and meanings, to Aristotelian essences, to metaphysical (as opposed to logical) atomism, to mystical capacities to intuit modal truths, to anything before (more or less British) empiricism because empiricism looks like it implies that there is no point to philosophy. We appear to have hit a dead end and are casting back along the trail; the policy is intellectually sensible, but it is strange to see *philosophers* following it covertly.

Outline

> What else but constitutive connections among concepts is there
> for a philosophical analysis to be the philosophical analysis of? And,
> if there are no philosophical analyses, what are analytic philosophers
> for?
>
> <div align="right">(Fodor, Concepts)</div>

The question of what philosophers do is another fundamental question that admits of an obviously correct but unenlightening one word answer — "Reason". This is not to say that other people do not reason too and in their professional capacities — just as carpenters use mathematics, and administrative staff write computer programs. All sorts of people reason in their day-to-day lives, but reasoning is philosophers' professional skill and it is because of their reasoning that philosophers' claims concerning the deepest consequences of our everyday beliefs and the most fundamental principles that justify them, warrant being taken seriously.

One traditional objection to logical positivism is that it is self refuting — that the central tenets of logical positivism imply that those tenets themselves are meaningless (or defective in some manner that precludes their being true), and hence must be false or nonsense. Putnam gave succinct expression to a version of this objection when he argued "if it is true that only statements that can be criterially verified can be rationally acceptable, that statement itself cannot be criterially verified and hence cannot be rationally acceptable". The equally well known defence against this charge if that the principles of logical positivism (or at least those of the particular version of logical positivism being defended) are analytic — i.e. true by virtue of meaning alone. The merit of this defence is disputed (by Putnam and others). But even granting it, a more serious objection remains.

The positivists did not recommend adopting their principles on empirical grounds — They did not charge their critics with holding theories that were empirically defective nor recommend their own as offering a better fit with experience. Their opponent's theories, including their intuitions concerning the

truth of logical positivism, were at least simple enough for those opponents to have learnt them, and also possessed whatever virtues are conferred by inertia and prior occupancy. Their own principles seem to leave little room for their *arguments* for those principles. To quote Putnam again, "If there is such a thing as rationality at all — and we commit ourselves to believing in *some* notion of rationality by engaging in the activities of *speaking* and *arguing* — then it is self-refuting to *argue* for the position that it is identical with or properly contained in what the institutionalized norms of the culture determine to be instances of it."[7] (i.e. holding true the statements which are supposedly analytic or true by virtue of meaning conferring linguistic norms).

This second charge, that positivism leaves no scope for reasoning nor for the justification of such philosophical doctrines as positivism itself, has also been thought to apply to holistic (more or less Quinean) empiricism[8]. The most widely touted defence of holistic empiricism against this charge has been to insist that there is no sharp line between science and philosophy, and that even abstruse philosophical doctrines such as holistic empiricism are justified by their connection to experience, albeit only indirectly and as part of a larger whole. The main purpose of this thesis is to show that both this charge and the defence against it are mistaken; that while holistic empiricism does imply that there is no analytic-synthetic distinction it yet leaves scope for the exercise of reason and does not imply that the distinction between philosophy and science is in any way indistinct.

The purpose of part one is to develop a clear statement of holistic empiricism.

Chapter 1 is devoted to clarifying reductive empiricism and showing why it is false. The most significant and interesting claim advanced in this chapter is that the common view that our belief in external objects is justified because the postulation of external objects offers the best explanation(s) of our observations and the success of our predictions radically under-estimates the strength of the true justification of our belief in external objects, which is that it is it is impossible to even state (let alone explain the correctness of) the empirical content of sophisticated scientific theories without postulating the existence of external (or at least unobserved) objects at some level. Viewed aright, the argument is an empirically based justification of the belief in external unobserved objects.

Chapter 2 is devoted to the production of a clear specification of the limits of human knowledge according to weak or holistic empiricism. The approach taken is based upon translation, and the main project is to clearly present austere idioms into which, according to holistic empiricism, all the statements we can be justified in believing true can be translated. The fundamental non-logical

---

[7](Putnam, 1981, p. p111)
[8]See (Putnam, 1983d) and (Sher, 1999) among others

components of these austere idioms are a finite number of predicates each true of just those subjects-at-times exposed to a single type of stimulus, and terms expressing discernible temporal relations between stimuli and set membership. Key difficulties are justifying the shift from experience to stimuli and identifying the stimuli themselves.

One aspect of the shift from experiences to stimuli is that it cuts out introspection as a source of knowledge. The first objective of Chapter 2 is to show that introspection is dispensable — that no projectible scientific laws that can be known with the aid of introspection cannot be known without it — and the second is to show that there are good reasons (other than a prior commitment to empiricism) which warrant the exclusion of introspective observations from the evidential base upon which our scientific theories are built.

The next task in Chapter 2 is to identify and provide individuation conditions for stimuli. Although Quine is extraordinarily cagey (even by his standards) about the relationship between stimuli and theories in *Word and Object*, the use of stimuli in that work requires them to satisfy three key constraints: identical stimuli must be indistinguishable to (each of) the subjects that can be subjected to them; insofar as they can be distinguished, any two stimuli must be compatible with exactly the same theories, i.e. there must be no theory for which just one of two identical stimuli (or two stimuli sequences composed of identical stimuli arranged in the same order) comprises a "recalcitrant datum"; and different subjects must be capable of being exposed to the same stimuli. This last constraint is essential for Quine's account of the identification and translation of observation sentences.

Section 2.2 thus begins as a search for objects satisfying these three constraints. Eventually however, after considering surfaces above, below and at the boundaries of subjects bodies as loci of stimulation, it is argued that there are no such stimuli to be had. The lack of inter-subjectively identical stimuli, more or less, has lead philosophers such as Donald Davidson to reject empiricism altogether[9]; the position advocated in section 2.2 however, is that the requirement of inter-subjective identification of stimuli is a relic of reductive empiricism that may be discarded without damage to the claims of weak empiricism or the accompanying standard of translation. Weak empiricism (and the accompanying account of translation) remain plausible without inter-subjectively identical stimuli provided exposure to stimuli is discernible to the subject's peers, *and* it can be determined which dispositions to assent to sentences in response to such exposures are resilient enough to alterations in subject's theories and expectations.

As the search for inter-subjectively identical stimuli is eventually abandoned,

---

[9]see (Davidson, 1984b)

the consideration of various loci for such stimuli may appear pointless, and indeed it is superfluous to the main purpose of this thesis (i.e. to show that holistic empiricism is not incompatible with the exercise of unempirical reason). However, in the course of considering various loci for inter-subjectively identical stimuli other serious objections to holistic empiricism are raised and refuted. Wittgenstein's criterial operationalism (an unhappy halfway house between reductive and non-reductive empiricism) is criticised in section 2.2.1. Putnam's account of the a posteriori discovery of natural kind's essences (and hence necessary truth discoverable a posteriori) and the role of social allegiance and role in determining the extensions of our terms is criticised in section 2.2.1. Though these critiques are proposed in the context of the project of individuating stimuli and the kinds explicitly considered tend to be psychological states, the arguments against these unempirical doctrines are readily generalisable to other kinds. Though these responses to the most influential objections to empiricism are not required to show that holistic empiricism is not self-refuting (in the way that Putnam held logical positivism to be self-refuting) in the current philosophical climate they are required in order to raise interest in the claim.

Having fixed the elements known by direct acquaintance the remainder of chapter two is devoted to specifying (but not justifying) the logical machinery by which references to objects known by description may be introduced. Against the background of this logical machinery, the class of sentences that subjects may be justified in holding true can be viewed as those that can composed from a finite number of two-place predicates each attributing exposure to a stimuli to a subject at a time (i.e. of the form S(x,t) and true of objects just at times when they are being exposed to the corresponding stimuli), predicates expressing the (local) order in which times occur, and an indexical term which at any given time refers to that time, and a two place predicate expressing set membership, by means of the logical operations of conjunction, negation, predication and quantification. Though there is some discussion of the inclusion of each of these fundamental elements, it falls far short of a conclusive justification of their inclusion. (Again, the main purpose of Part One is merely to produce a clear statement of weak empiricism's doctrine.) Further terms may be introduced into such austere idioms by means of Carnapian meaning postulates (though Lewis' technique for giving outright definitions of theoretical terms is closely considered and envied).

Not only is the choice of the primitive non-logical terms of the idioms proposed in Part One justified only inconclusively, the choice of logical terms and the logical operations by which sentences may be formed from all these (and theoretical terms introduced) are not justified at all. The main purpose of Part Two is not only to justify the choice of logical terms and to show why logical

truths are immune to revision (without change of meaning) but, equally importantly, to show that this justification does not extend to other non-logical truths.

The first objective of Chapter 3 is to promote conceptualist (more or less Wittgensteinian) accounts of necessary truths and our knowledge of them on empiricist grounds. Short shrift is given to externalist accounts of modality based upon object's intrinsic essential properties. The old empiricist's arguments against such views are well known (as are Quine's arguments against quantifying over modal operators) and are as good as ever (and Putnam's account of essential features discoverable a posteriori was rejected in chapter two). Attention thus quickly turns to internalist accounts of modality according to which modality is imposed upon (our thoughts about) the world by our minds. The consideration of internalist accounts starts on a roughly historical note and the first objects considered as candidates for foundations of modality are sense-data. The constraints sense-data are required to fulfil in order to serve as foundations are delineated and sense-data are found wanting (in §3.1).

The next objects considered as candidates for modal foundations are concepts. It is suggested that concepts are purpose posited to be free of the defects besetting sense-data and various arguments for and against conceptualistic accounts of necessity and our knowledge that truths are necessary are considered. Reflecting on these arguments now, the case made against conceptualism is perhaps stronger than I had thought, and the (Wittgensteinian) defence less airtight. In the context of this thesis however excessive generosity towards conceptualism is acceptable at this point, for a much better argument against conceptual/linguistic accounts of necessity is to be presented in Chapter four.

In preparation for this argument Chapter three ends by considering psychologism, for the argument to be advanced in Chapter four is an epistemic argument and the proponents of the strongest forms of anti-psychologism are only too happy to grant that there is no public empirical evidence[10] of concepts (or indeed of intentionality). The very strongest form of psychologism, according to which empirical psychological investigations can justify knowledge of necessary truths, is rejected. However the more moderate claim that (instances of) concepts can be identified with the posits of empirical psychological theories is defended and it is insisted that there must be some public empirical evidence of concepts for conceptual analysis to be a collaborative exercise and serve its supposed function of settling arguments (and in particular of quashing various forms of skepticism[11]).

---

[10]The term "evidence" is not to be paraphrased as "stimuli" in the manner advocated in Chapter two, but understood in the ordinary sense in which (particular) loud beeps and diffraction patterns are evidence.

[11]see (Fodor, 1998, p. 70)

Chapter 4 is devoted to presenting a clear argument against the epistemic accessibility of the analytic-synthetic distinction. The argument is closely based upon, though not an exegesis of, the arguments presented by Quine in *Two Dogma's* and *Word and Object*. The chapter begins by considering some of the very bad arguments commonly attributed to those papers and putting them aside. In addition to some trivial interpretations of Quine's arguments, the claim that holistic languages do not contain distinct classes of statements that must be held true come what (experience) may by competent speakers is also rejected. Indeed it is argued that any holistic language in which the empirical contents of our sophisticated scientific theories can be expressed must contain some statements that cannot be revised. Despite this, the good argument continues, *the* unrevisable or analytic truths of a subject's language cannot be determined because the empirical evidence cannot uniquely determine which language a subject is speaking. Indeed not only is there for any (minimally sophisticated) language speaker a range of more and less restrictive languages, containing more and fewer analytic truths, which she may be taken to be speaking (compatibly with her dispositions to hold sentences true); but any competent speaker of a relatively restrictive language is, ipso facto, also a speaker of all the (strictly) less restrictive languages. The choice of which of these languages to take a subject as speaking is arbitrary, and by the same token so too is the choice of which of the sentences he is stably disposed to hold true to take as analytic. The least restrictive languages capable of doing justice to the empirical contents of our sophisticated scientific theories have only the minimal range of analytic truths capable of supporting the inferential connections required to articulate the connections between observations and prediction; this minimal range constitutes the logic of the language and it is to be chosen on the grounds of the articulation it provides between (sentences reporting) observations and predictions. The claim that this minimal logic of human languages capable of expressing sophisticated empirical theories is quantificational calculus is then defended upon empirical grounds.

Before passing on to the choice of logical machinery which closes Chapter four, several objections to the main argument against the epistemic accessibility of the analytic-synthetic distinction are considered. Of these the most significant are based upon Kripke's objection to dispositionalistic semantics, and some time is spent going over the problems raised by subjects' fallibility and those Kripke raised in connection with "quus".

Having spent Part One in delineating the evidence upon which our theories are based, and Part Two considering and then rejecting concepts and conceptual entailment, in Part Three we turn directly to the question of what is left of philosophy and pure reason.

Chapter 5 is devoted to arguing that far from implying that the distinction between science and philosophy is indistinct holism implies that the distinction is perfectly sharp. In particular, though individual statements cannot be divided into scientific and philosophical claims, holism implies that disputes are never over individual statements in isolation but only ever over the choice between whole theories, and a perfectly sharp line can be drawn between scientific and philosophical disputes.

Chapter 6 is devoted to attempting to specify the grounds upon which philosophical disputes may be decisively resolved. While some disputes may be resolved as being merely verbal or due to mutual incomprehension, such resolutions are symmetrical and give no reason for preferring one theory over another. Simplicity is considered as a grounds for decisive resolutions, but, while it is admitted as ground for resolving scientific disputes it is found wanting as a basis for resolving philosophical disputes. Attention then turns to ways in which theories may be defective independently of their empirical contents.

Part I

Evidence

> The product of scientific practice is the maze of theories that make up a science.
>
> (A.F. Chalmers, *What is this thing called science*)

The reason for beginning our investigation of philosophy by considering the nature of scientific theories is the evidence. If all evidence is amenable to scientific discovery, and the logical articulation of our scientific theories is adequate to the recognition and declaration of every humanly discernible pattern in our observations of the evidence, then nothing humanly known is beyond the reach of science, and if philosophy makes any indispensable contribution to the development of human knowledge it must make a contribution to the development of scientific theories.

It might be that philosophy makes a dispensable contribution to human knowledge; that philosophers do make discoveries, discern novel truths, and so on, but that all these discoveries would be made (or surpassed) by scientific investigations in due course. While philosophy would still be a worthwhile field of study in that case, its value would arise from a sort of happenstance — that we discover certain truths more quickly or easily by application of philosophical rather than scientific practices (supposing there is a sharp line to be drawn between the two) would be a sort of peculiar fact concerning our psychology and perhaps the pragmatic difficulties of mounting field expeditions. One way to delineate the scope for philosophical contributions to the development of knowledge (if any) and determine whether they are due to the happenstances of psychology or pragmatics is to explain how philosophy may contribute to the development of scientific theories.

Chapter 1

Strong Empiricism

> They have seriously considered the possibility that sense-data as
> well as material things might appear to have properties they did not
> really have. ... by adopting this convention they come to treat sense-
> data as if they were themselves material things or characteristics
> of material things; and in that case the terminology of sense-data
> becomes superfluous
>
> (A.J. Ayer, *The Foundations of Empirical Knowledge*)

Strong empiricism is marked by commitment to the empiricist doctrine of
content — viz. that all statements are *about* perceptions or sense-data. The
substance of this claim is unclear so long as we are unclear about the relation of
being about or having as content[1]. Clarity is at least improved by refining the
claim into the following: Each significant indicative sentence[2] can be translated
into a language in which all terms refer to objects of immediate perception.

The improvement is only a matter of degree, while clarity is improved some-
what by dispensing with talk of being about in favour of the comparatively clear
notions of reference and translation, better yet would be to avoid the word-world
relationship of reference altogether and deal with the purely internal properties
of a sense-data language and avoid the issue of reference altogether. For one
thing the claim is, as it stands, slightly too strong — reference is a success term.
This form of the empiricist theory of content requires there to be objects of im-
mediate perception, sense-data, for the terms of the supposed phenomenalistic
language to refer. Undoubtedly many empiricists did advocate an ontology of

---

[1] Are sentences about the objects they refer to, or the states of affairs which those objects
and the universals they instantiate comprise? With that question begins another and different
thesis.

[2] We rely upon grammarians of English to identify the indicative sentences, overcoming this
linguistic chauvinism where it suits by extending the term to cover the translations of such
sentences.

sense-data, but we may profitably distinguish this ontological claim from the semantic component of the empiricist doctrine of content, and simply note the restricted range of putative references, or postulates, of the sense-datum language without worrying over whether they succeed in referring (or what such success consists in). With this modification in place the doctrine becomes: Each significant indicative sentence can be translated into a language in which the only objects postulated are objects of immediate perception.

This again shifts the weight, which is now borne by the notions of translation, postulation and object of immediate perception; each deserving further explication.

1.0.1   Translation

Naively one presumes that good translation preserves meaning, and that translation is a reflexive, transitive and symmetric relation between sentences. But this may be to presume too much. Identity of meaning, and hence the individuation and reification of meanings, may be a notion we are required to live without. Theories of translation are but one type of empirical theory, and we are just begun considering how such theories are related to the evidence and what that evidence is; this notion of meaning might rely upon extra-empirical resources. A principle of translation that must survive even the toughest empiricist scruples, is that translation preserves evidence. Whatever evidence or pieces of evidence are, and whatever the correct explication(s) of the relationship of being evidence for, which they bear to statements (or beliefs or theories or assertions), one object cannot be a translation of another if there is evidence for or against just one of them.

1.0.2   Postulation

Postulation is what is left in focus when one is concerned purely with what is required of a theory for representation of objects to occur and not how world and theory be related in order to establish reference. Thus isolated, the theoretical structure of postulation can be clearly specified. A term is a putative reference, a name or description, when it is subject to the machinery of predication, quantification, and identity; machinery embodied in phrases such as "a", "the", "the same", "another", "everything which is", "all" and "such that". This is only a partial list of the supporting machinery of reference with which a term is required to interact in order for it to count as a reference at all. A complete description of this machinery (and its workings) amounts to the specification of quantificational calculus, in which such machinery is crystallised.

Two reasons for adopting this quantificational standard of postulation are that it is completely explicit (in application to theories already in quantificational form) and the quantificational notation on which it relies is adequate for the expression of our theories. A third is that the role of the existential quantifier is close enough to ordinary use of the English terms "exist" or "is" to warrant treating the formal version as an explication of the ordinary notion of existence. There may be other ways of reforming ordinary usage, other notations than quantificational calculus, and alternate explications of "exists" with accompanying standards of postulation. Each such alternative provides a corresponding explication of the empiricist doctrine of content and another version of strong empiricism. The question of which of these alternative standards is correct is as misplaced as asking which translation of arithmetic into set theory is correct (or indeed which set theory). Any alternative doing as well on the score of explicitness and adequacy as the quantificational standard, and doing equal justice to ordinary usage, would be equally good.

The explicitness of the quantificational standard is irreproachable. Its adequacy to the expression of our scientific theories is, however, a subject of criticism. In particular there is no accepted translation of the attributions of propositional attitudes (e.g. "John hopes that it will rain", "Jane believes that the Earth is round", but also less obvious instances such as "Willard is hunting lions") into quantificational notation. If, however, any significant extension beyond the limits of quantificational calculus were beyond the reach of any (in principle) humanly scrutable proof procedure, then its adequacy to the expression of human scientific theories would be guaranteed, for it is by such procedures that scientists seek to derive predictions from their theories.

We shall return to the question of the adequacy of quantificational calculus for scientific theorising in section 4.6. For the purpose of arguing that strong empiricism is false, we do not need to show that every scientific theory can be expressed in quantificational form, but only that the empiricist doctrine of content is false of some statement in some scientific theory that is expressed in quantificational form.

Examples of empirical theories in quantificational notation are not generally found in text-books or journal articles. Despite this, it is no longer uncommon to find scientific theories, or at least empirically significant slabs of them, translated into absolutely formal notation — not for the purpose of clarification but in order to automate the generation of predictions. New technologies offer more tempting rewards for formalisation of theories than the traditional returns of clarity and precision. Computational models partly free scientists from the often tedious process of deriving predictions for particular occasions from their store of universal truths, and allow finely detailed theories to be applied on scales that

beggar the imagination and dwarf the intellect. The enhancement in the range
of observable consequences that can be generated in practice not only increases
the scope of cases to which existing theories may be applied, but also provides
new ways of testing and distinguishing novel theories.

Such models have been developed and fruitfully applied in fields as diverse
as meteorology, quantum physics, biology, genetics, economics, physiology and
chemistry, and the construction of such models continues to be a boom industry.
The inputs and outputs of such programs are, or can be rendered as, sentences
— sentences used to make assertions about the matters being modelled (not
that the *program* asserts anything by producing such outputs, but they are
the sentences which its *users* might use to report the predictions their model
vindicates) — and for each such program there is a theory in quantificational
notation capturing as theorems, of the class of material conditionals having an
input as antecedent and output as consequent just those whose consequents the
program produces as outputs in response to their antecedent as input.

1.0.3 Sense-data

The distinguishing feature of sense-data is supposed to be the quality of our
epistemic access to them. Coloured patches of visual fields and sensations of
heat or cold have often been taken as paradigmatic examples.

The goals of early British empiricists included delimiting the claims which ex-
perience could give us good reason to believe, pruning our science and common-
sense of any pretence to knowledge outside this sphere, and clarifying the rela-
tionship of what was left to the available evidence [3] . At some stages (particu-
larly where the project was influenced by Descartes' quest for certainty) the only
reasons accepted as good were those supposed to transmit absolute certainty,
guaranteeing the truth of the beliefs they supported. This conception of good
reason, at its narrowest restricted to deductive implication, foundered for the
lack of certain foundations for the generalisations required of a fruitful scientific
theory (and common-sense). Any prediction goes beyond a strict statement of
the evidence for that prediction, and relies upon generalising from the acknowl-
edged data to bridge the gap; without the capacity to make such predictions
not much of science is left.

While the pursuit of certainty did not reach its intended goal, it did reveal
the reliance of science upon non-deductive inference, and in this revelatory spirit
empiricists might still hope to discern all the gaps through which errors might
enter into our theories. By starting from certain experience and then tracing

---

[3]I have in mind Locke and Hume particularly, but these goals were widely shared by their
peers.

the inferences from this foundation to our scientific hypotheses, the unverified assumptions made in scientific practice and the roles they played would be clearly displayed. If all talk were translated into talk of sense-data then the assumptions would show up in sentences which postulated sense-data that one was not experiencing[4].

In pursuing each of these tasks, British empiricists (and their fellow travellers) assumed that the senses provide certain foundations for empirical theorising. Neither projects is promoted by a reduction to sense-data if sense-data are known only as partially and dimly as the objects familiar to us from ordinary science and common sense. If sense-data are postulated to explain our experiences, like everyday chairs and tables, then strong empiricism is pointless; theories quantifying over sense data would be mere alternative sciences, without any special claim to epistemic transparency.

The stringency of this requirement upon sense-data is extraordinary, and leads strong empiricism afoul of the following well-known objection[5]. Let us assume that sense-data include colours (or coloured patches of the visual field, or coloured whatevers, take your pick) and of these there are three dubbed "A", "B", and "C", of which A and B are of indistinguishable hue, B and C are of indistinguishable hue, but A and C distinguishable. Instances of such triples can be provided by visiting a cinema and viewing a sufficiently slow fade-to-black[6].

Such intransitively indistinguishable triples present a dilemma for strong empiricism: Are the apparently indistinguishable sense-data identical or not? Strong empiricists cannot claim that A and B merely appear to be identical but are in fact not identical without vitiating the supposed epistemic clarity of a sense-data language. On the other horn, the cost of maintaining that these objects are identical is that identity becomes intransitive, A is identical to B and B to C but not A to C. Such intransitivity of identity is incompatible with the principle of substitutivity of identicals, and to surrender that principle is to give up on identity itself, at least in the quantificational form upon which our account of postulation is based.

The dilemma does not turn upon any special feature of colours, analogous cases can be constructed for pitch, temperature and sensory extension (the property by which sense-data present objects as being extended), but let us not be distracted by watching grass grow. Nor can it be dismissed as an artefact of the quantificational standard of identity; the difficulty rests lies not in a limita-

---

[4]Versions of empiricism differ over whether the postulation of objects of immediate perception that one *had* experienced were speculative.

[5]Originally raised, as far as I can tell, as an objection to Bertrand Russell's ontology of sensibilia.

[6]Well, they would be provided if perception involved sense-data, and that's the target of the reductio ad absurdam.

tion on our capacity to describe such sense-data, but in our limited capacity to distinguish them.

The problem is that simple inspection of a colour (or length or position or weight etc.) is not enough to determine the colours from which it is distinguishable. A and B *look* the same, and yet A is distinguishable from C and B isn't; not that the distinction between A and B cannot be registered in language, but that it doesn't register upon direct inspection. That A and B can be distinguished in the presence of C doesn't help, for we might take more B for C (they look the same) and mistakenly conclude that everything being seen is C when in fact nothing is.

In presenting this dilemma we have assumed that sense-data are coloured, or at least present other objects as being coloured, extended, and so on and that such presentation is dependent on the properties and relationships of sense-data and we might be wrong about all this. Sense-data may be radically different from anything we have considered. But it is not open to a strong empiricist to argue this, at least not if she wishes to proclaim sense-data's absolute epistemological priority. If we need to infer the existence and nature of sense-data from other, uncertain, considerations (e.g. because our epistemology posits absolutely secure foundations from which we insecurely generalise) then sense-data are incapable of providing the required epistemically transparent foundation.

A strong empiricist may hope to weather this dilemma by lowering her sights. She might still hope to posit objects of which our judgements are less error-prone than those concerning objects posited in our more familiar science, and then reconstruct our familiar science in terms of her novel posits. Such reconstructive projects are not unique to empiricism, much the same motives led philosophers to attempt to reduce mathematics to logic. The distinctive feature of empiricist attempts is that the objects to which they seek a reduction are supposed to be directly observable.

While we took empiricists' hopes for certain foundations seriously, we could rely upon incorrigibility as a mark of the observational. At that point the mere fact we *thought* objects of immediate perception were coloured, together with our limited abilities to distinguish colours, was sufficient to show that there were no sense-data in the sense required to deliver the desired epistemic transparency. Without such a clear criterion of observationality the category of empiricist reductions (and hence of strong empiricism) is less clear cut.

The problem is not that our classification of theories is fallible, but that each form of strong empiricism comes with its own account of what it is we directly observe, and hence its own standard of what counts as strong empiricism. The situation comes to resembles a religious dispute — many creeds each claiming to be the one true church, each with their own self-serving criteria of authenticity.

Such disagreement vitiates the usefulness of a category of strong empiricism within a taxonomy of philosophical theories. We are left with a vague cluster of positions involving one or another reform of the term "observation"[7]. Even rationalism may be counted as a type of empiricism if pure reason is allowed as a form of perception (perhaps with concepts or Platonic forms as its objects).

1.1 The refutation of strong empiricism

The pursuit of a clear and precise account of strong empiricism is a lost cause. In the absence of a common account of the objects of direct observation, or even plausible criteria for assessing claims of direct observation, we are doomed to live with a degree of vagueness and ambiguity over the range of theories that count as forms of strong empiricism.

One might easily think this puts strong empiricism beyond criticism (except on the very count of vagueness) — that any substantial criticism can only be of this or that form of strong empiricism (this or that way of distinguishing observation from other forms of belief fixation). Even if the form of strong empiricism we judge most plausible is wanting, that may simply reflect our inability to discern a better alternative. Surprisingly this is not so, for though it remains unclear what exactly to count as strong empiricism, it is clear that no such reductive program can succeed.

There are three strategies for refuting strong empiricism. The best known and most often attempted is to argue that there are truths known by reason alone. Such truths have often been supposed to include necessities and generalities known with certainty. Empiricists have in general agreed that no experience can justify knowledge of necessity, nor absolute confidence in a generalisation extending over all experience[8]. More to our point is that such truths cannot be about experience if experiences are contingent and known only in particular instances. Empiricists have made various attempts to explain our knowledge of such truths; Mill denied necessity, while Wittgenstein, Ayer and other analytic philosophers attempted to explain necessity in terms of analyticity.

A successful refutation of strong empiricism based upon this strategy would show not merely that some beliefs are about something other than experience, but that some beliefs are justified by something other than experience. Since the etiologies of particular psychological states are themselves composed of particular contingent objects and their causal interactions, that something other cannot

---

[7]On the other hand, strong empiricism may remain a useful category in the history of philosophy if there are useful generalisations concerning the origins, development or effects of the movement.

[8]"No matter how often [a general proposition] is verified in practice, there still remains the possibility that it will be confuted on some future occasion" (Ayer, 1936, p. 64)

be an entirely natural phenomena. Unnatural realms are not to be posited upon
a mere whim, they must pay their way in explaining and predicting what natural
science cannot; the capacity to know necessary truths is one case in point.

The evaluation of the most compelling of these arguments — consideration
of whether or not the capacity to reason, and in particular the capacity to
resolve arguments that cannot be resolved by empirical test (i.e. to engage in
philosophy), is beyond the capacity of naturalistic empiricism to explain — is
the over-riding purpose of the remaining sections of the current work. For the
moment however, let us concentrate on refutations of strong empiricism which
are not also refutations of empiricism *tout court*.

Arguments following either of the remaining two strategies do not contend
that there is a source of knowledge other than experience. The first of these
two is to attempt to show that individuals hold beliefs whose truth is indepen-
dent of their experiences. McDermott's argument from life-insurance exemplifies
this approach[9]. The argument begins by noting that people usually purchase
life-insurance because they believe their death will trigger a payment to the
beneficiaries of the insurance policy. People find it sensible to make such pur-
chases despite knowing that they will not be around to witness the pay-off or its
consequences because their beliefs are about other people, and not merely sense
impressions of them. Such arguments are less than convincing because their
proponents concede that there is empirical evidence for the beliefs in question
and hence it is open to a strong empiricist to attempt to identify the content of
such beliefs with the evidence for them. In the case of life-insurance, a strong
empiricist may maintain that what is purchased is the assurances of the insur-
ance agent and the reduced anxiety of the beneficiaries, all of which may be
observed by the purchaser. Nor is the objection insuperable that a purchaser is
deceived when payment is not made despite the assurances she receives on her
death-bed. A strong empiricist must claim that the the deceased has received
the full value for his insurance purchase and maintain that it is other prospec-
tive purchasers who suffer from such deception, from having their confidence in
insurance agents' promises reduced[10]. That we feel it is the deceived purchaser
who has been wronged must be dismissed as a product of our deep inculcation
in a social system which allocates blame and injury in such a way as to give in-
centives to each player to maintain the stability of the system. Within our legal
system it is easier for the representatives of the deceased and her beneficiaries

---

[9]McDermott cites C.D. Broad as the earliest source of this argument that he knows of.

[10]That the making of the payment to the beneficiaries may be a matter of their experiences
does not help. If sense-data include such hidden objects as the experiences of other people,
Ayer's criticism quoted at the beginning of this section comes to bear. Both this and the
subsequent refutation are intended to show only that there must be unobserved objects which
make our statements true, not reveal the nature of such objects, and in particular not to show
that they could not be other people's experiences.

to mount a case than to stage a class action on behalf of prospective purchasers, hence we allocate injury to the deceased, rather than to those insurance purchasers who survive him.

While the strong empiricist's response seems unintuitive, that is not enough to show that it is false. What is required, if the argument from life-insurance (or others following the same strategy) is to convince any but the already converted, is a demonstration that the statement in question ("The beneficiaries of the policy will receive a payment after my death") is not equivalent to any statement about experience.

The third strategy for refuting strong empiricism is weaker yet. Like the second it does not rely upon the existence of any unempirical source of knowledge, but where arguments following the second strategy turned upon statements whose truth was independent of experience, arguments following the third strategy concede that our claims do not go beyond the empirical. Instead the third strategy is to argue that the postulation of unobserved or hidden objects is required to provide the requisite empirical content — that no theory which does not postulate hidden objects can have the empirical content of sophisticated empirical theories such as those of our sciences. The problem for this third approach is that there is a formal proof of the contrary.

1.1.1 Craig's theorem

Craig's theorem states that if a class of sentences can be captured as the theorems of any one theory at all, then they may be captured by a theory which does not contain any expressions beyond those used in those theorems themselves and truth functional connectives. One may capture just those theorems of any given theory which do not contain expressions that are somehow objectionable, without relying upon those objectionable expressions in the proofs of those theorems.

In the case of strong empiricism, the candidate objectionable terms are those which refer to objects other than experiences or impute properties which are not the observable properties of experiences[11], the theorems to be captured are those which concern observables and Craig's theorem demonstrates that we can have a theory which postulates no unobservable objects or properties and captures as theorems the same claims about experience as a more liberal empirical theory.

Craig's theorem appears to provide a knock-down response to arguments following the third strategy since such arguments hinge upon showing that unobservable objects are indispensable supports of our scientific theories' empirical

---

[11] As we have already seen, this is so strict as to leave empiricists nothing to talk about, but the current point is that there is no logical reason why such strictures cannot be met.

contents; but this appearance does not survive a closer inspection of the proof by which Craig's theorem is established and the role proofs and theorems play in theories.

Logical proofs have been traditionally viewed as establishing a necessary connection between their premises and their conclusions. Nowadays, while that view of logic remains orthodox, it is also widely disputed. Though the nature of the connection between premises and conclusion is debatable, it is clear that logical arguments decisively establish the truth of their conclusions among those who accept their premises and the rules of the inference in question. This is not a truism – the rules of a system might be vague or unclear; people who agree to the rules of a game and the location of the pieces may yet be unable to resolve disputes over the outcome of particular games[12]. Logical systems avoid such unsettled results by having obvious, and decisive standards for determining whether any particular step in a proof is in accordance with that system's rules of inference.

The orthodox explication of "obvious and decisive standards" relies upon the notion of recursiveness — a class of rule compliant inferential steps is clear and obvious iff it is recursive. In terms of familiar objects, a class is recursive just in case there is a computer program which can always determine whether any given object a member of the class or not[13].

This explication of "obvious steps" in terms of recursive procedures is of key significance for arguments following the third strategy for refuting strong empiricism. If the class of acceptable steps in the derivation of predictions from observations is required to be recursive, then arguments following the third strategy may succeed by showing that there is no recursive class of inferential steps by which our scientific predictions can be derived (from a recursive class of axioms) which does not rely upon formulae postulating hidden objects in the intermediate steps of those derivations.

If the classes of predictions to be derived were themselves recursive, arguments following the third strategy would be doomed from the outset; but the classes of predictions derivable from any except our most banal scientific theo-

---

[12] In the game of Go the best known oddity in the Japanese rules covers the special situation called "bent four in the corner", in some situations this rule gives different results from those of Chinese rules, and in others again it is not clear whether the "bent four in the corner rule" is applicable at all. These are just some of the many strange situations in the bestiary of Go oddities over which advocates of the same naive rule systems can find themselves at odds.

[13] More technically: The recursive functions on the natural numbers include any k-ary function which returns the value of its nth argument, any constant function, and the successor function. A recursive function is either one of these initial functions or one that may be cobbled together from them by means of any combination of composition (where e.g. $F(\vec{x}) = G(H_1(\vec{x}), H_2(\vec{x}), \ldots, H_n(\vec{x}))$) induction (where e.g. $F(0, \vec{x}) = G(\vec{x})$ and $F(y+1, \vec{x}) = H(F(y, \vec{x}), y, \vec{x})$) or minimization (which returns the least member of a class). A recursive class is one for which there is a recursive function mapping just its members to non-zero values.

ries are recursively enumerable rather than recursive. For each recursive class there is a foolproof method of determining whether any given object object is a member or not; for recursively enumerable classes however there is no more effective method of determining membership than reviewing a list of members hoping to come across an entry corresponding to the object in question[14]. Such checks are comparable to those of an electoral officer checking for a prospective voter's name against an infinitely long electoral role which isn't in alphabetical order — that the desired name hasn't yet been found is no guarantee that it doesn't occur further down the list.

Proofs may be checked but theorems can only be enumerated, and this provides an opening for a refutation of strong empiricism following the third strategy. If the derivations of a recursively numerable class of predictions from (recursively enumerable) axioms was shown to require the use of intermediate formulae concerning matters that are not directly observable [15] (i.e. positing unobserved objects or objects in unobserved states) then strong empiricism would be refuted. Craig's theorem turns upon a technique for constructing intermediaries free from such commitments.

Craig's theorem relies upon two devices, one is the familiar technique of Gödel numbering, associating, by means of a recursive function, each finite sequence of formulae of the source theory (containing objectionable terms) with a number; the other is a trivial technique of associating numbers with formulae by means of repetitive conjunction. The number associated with any formula is simply however many repetitive conjuncts it contains. "A & A & A", for example, is associated with the number 3. Since both the Gödel function and the class of proofs of the source theory are recursive, so too is the class of Gödel numbers of proofs, so too the class of repetitiously conjoined formulae in which the number of repetitions is the Gödel number of the proof (in the source theory) of the tediously repeated conjunct. Craig's ingenious suggestion is to simply take each wordy member of this class of sentences which thus encode the Gödel number of their own proof (in the objectionable source theory) as axioms. The class of axioms is recursive, capturing, by means of the obvious simplification of duplicative conjunctions, all the desired unrepetitive formula of the original as theorems, and yet involves no formula suggesting the existence of any objects beyond those mentioned in the unobjectionable theorems themselves.

Craig's technique is a logician's trick no doubt, but it is not obviously less

---

[14]A class is recursively enumerable just in case there is a recursive function such that each member is the value returned by the function for some natural number given as the argument. I follow the convention of applying "recursively enumerable" without qualification only to classes that are recursively enumerable and not recursive, though strictly speaking the latter are a proper sub-set of the former.

[15]Recall that we are granting this distinction to the candidate strong empiricist for the sake of the reductio.

devastating to third strategy arguments for that. A lingering suspicion, which
Craig considers in his second paper, is that this device is merely superficial,
failing to avoid dependence upon objectionable terms and the formulae contain-
ing them because of the continued reliance upon the source theory to guide the
choice of axioms. Craig's response to this criticism warrants full quotation:

> This criticism can be answered as follows: The method . . . of
> testing whether or not a given conjunction $A\&A\&A \ldots \&A$ is an
> axiom of $S*$ [the target theory] can be modified so as to bypass
> the actual construction of proofs in S [the source theory]. Instead,
> we can test by a purely numerical method. Details of this method
> do not matter here. We must specify, among other things, which
> numbers are Gödel numbers of axioms, which numbers are Gödel
> numbers of expressions regarded as auxiliary [i.e. objectionable] etc.
> What matter is that in order to specify these Gödel numbers we
> do not have to write out the axioms, the expressions regarded as
> auxiliary, etc.; we only have to use *names* for them. Now names for
> expressions regarded as auxiliary seem no more objectionable than
> names for other expressions. In either case, we are naming the same
> kind of object." [16]

While this defence clears Craig's technique on the charge of tacit reliance
upon the objects of the source theory, it does not meet the requirements of
strong empiricism. The described technique avoids the use of objectionable ex-
pressions but it relies upon the use of numbers, in particular Gödel numbers of
expressions, and numbers are exemplary unobserved objects. The target *theory*
need not carry commitment to numbers or any other unobservational objects,
but the *theorist* must be able to calculate Gödel numbers of formulae and encode
them in repetitious conjunctions[17]. Seen in this light Craig's technique shows
itself as particularly subtle form of partial pythagoreanism; a way of compress-
ing unwanted ontological commitments into numbers, and particularly subtle
because no trace is left of the objectionable objects in the target theory; instead
unwanted ontological commitments disappear into the calculations required to
determine which conjunctions are axioms of the target theory.

Craig's theorem shows our initial statement of the third strategy for refuting
strong empiricism to be mis-targeted. To refute strong empiricism it is sufficient
to show that theorists must be committed to the existence of other objects —
that no theorist can believe empirical theories generating an recursively enu-

---

[16] (Craig, 1956, p. 50)

[17] Nor is there any advantage in sticking to expressions, for the longest of them are too long
to count as directly observable in any case, even given an acceptable alternative to identifying
formulae with classes of inscriptions.

merable class of predictions without reliance upon unobserved objects — even if this or that one of their theories carries no such commitment.

To put the matter this way appears revisionary. It seems to require abandoning the quantificational standard of ontological commitment and returning to the notion of ontological commitments as something in the mind of the theorist. But this is not quite so, for our ontological standard applies directly to the overt quantificational formulae manifesting the derivation of predictions; when such overt pronouncements are lacking the standard can be applied only insofar as interpretation supplies them. A theorist may avoid any overt commitment to unobserved objects if she is willing to leave the derivation of her predictions inexplicit (indeed a theorist capable of *deciding* the membership of recursively enumerable functions would not need to make recursive chains of proof at all) but the quantificational standard of ontological commitment might yet be meaningfully applied by attributing the posits made in one or another quantificational theory capable of generating just the requisite predictions (the scope for free invention in such exegesis is striking, so too the need to relativise ontological commitments to interpretations).

The objection to the products of Craig's technique is not merely that something is left out. No theory contains an open sentence true of just those open sentences which are false of themselves, but a theory can be rich enough to capture a process by which its theorems are discernible, rich enough to describe its own proof procedure. In a sense so can Craigian theories, for if a source theory contains a term interpreted as attributing theoremhood, and includes among its theorems all and only the true applications of such a term, then Craig's technique may be applied to produce a theory counting suitably repetitious versions of such theorems among its axioms while avoiding the need to count sequences of formulae among its postulates or for a predicate true of those sequences constituting proofs.

The charge is that Craig's proposed technique merely defers ontological commitment. It allows postulates of any sort to be supplanted by Gödel numbers or tuples of symbols, and defers even the postulation of even those to the level of a theory describing the proof procedure of the resulting theory, deferral that can be indefinitely repeated. To make this charge stick requires an argument showing that such deferral must stop at some point, that the derivation of theorems (including, in the case of empirical theories, predictions) is left inexplicit in the products of Craig's technique and successfully described by other theories; and even that is not enough, finding Craig's technique wanting does not show that other techniques for dispensing with unobserved objects are equally unsuccessful. We require a positive argument for the indispensability of hidden objects in supporting complex empirical contents.

1.1.2   Hidden objects and computational complexity

Computers store information beyond that displayed on their monitors. So obvious is this truth that the evidence for it is little noted though not hard to discern. The images displayed in the structured rectangles dubbed 'windows' often persist through being covered and uncovered, hence the ready applicability of the terms 'top', 'bottom', 'over' and 'under' to such windows, hence indeed the applicability of the term 'window' itself. We often predict that the image displayed within a window just before it is covered will be redisplayed when it is uncovered, and explain this by means of the computer's internal storage of information required to generate the requisite image.

One might think this enough to establish the existence of (occasionally) hidden objects behind our perceptions. Balls pass behind pillars and re-emerge on the other side, facades reliably appear and change shape as we circle around buildings, such is the stuff that leads observers following the system of Carnap's Aufbau to attribute colours to points of space-time to which they have no line of sight. But such assurance of the existence of hidden objects is only forthcoming on the presumption that causation is always local — that objects are directly causally affected only their immediately adjacent surroundings at immediately preceeding times. If a steadily disappearing circle may directly cause the later appearance of a qualitatively identical circle on the other side of some differently coloured patch of the visual field there is, in that respect at least, no reason to postulate an enduring ball passing behind an enduring pillar.

So too, to continue the analogy with computers, if we are prepared to envisage circuitry connecting memory locations directly displayed[18] at non-consecutive times (together with the settings of switches under user control) much of the argument for internal storage goes by the by. So firmly embedded is our belief in locality that such permissiveness seems to allow a kind of de facto storage of hidden values, but all of these values we are imagining being transmitted by connecting circuitry are directly displayed at some time or other. Rejecting locality, showing that some observed events are directly caused by (or at least predictable solely on the basis of) events observed some time ago bolsters confidence in strong empiricism rather than refuting it. An empiricist might explain the postulation of hidden objects as a result of our preference for

---

[18]Some computer architectures do directly display an array of memory on screen. Starting at some specific memory location each successive byte or word of storage determines the colour displayed in each successive pixel. Most architectures are not like this, and are in any case lost in the mists of device drivers and graphics packages to high-level programmers. Let us concentrate upon the simplest architecture, the advantages of other arrangements are merely pragmatic. Even in the simplest case however there is a mass of circuitry standing between the values stored in directly displayed memory and the beam of electrons striking our screens. "Direct" in this case means merely that there is a one to one correspondence between the tuple of values stored in directly displayed memory and the images displayed onscreen.

local causation and, while admitting the merit of that preference, still maintain that such postulates are in principle omissable without loss of empirical content.

To find an argument for hidden objects that does not rely upon locality let us imagine not one but a sequence of screens, each consecutive screen displaying something of the state of a computer at successive times, and consider what reason there could be to suppose that in addition to all the values displayed at one time or another there was some further hidden storage. To put the matter another way, let us determine if there are any calculations which cannot be performed by machines without hidden storage (we shall call machines "transparent").

It is readily seen that any computation performed by a machine with hidden storage may also be performed by some transparent. Given any machine with hidden storage a transparent machine capable of performing the same computations may be constructed by simply connecting the hidden storage to the appropriate number of extra pixels (or other flashing lights) without disturbing the existing connections. The result would resemble the mock-ups of computers known from 1950's television and film — a somewhat restricted display of useful information surrounded by a mass of lights flashing on and off with little readily apparent pattern. Clearly such a machine could perform any computation performable by the original machine on which it was based. Given any particular input it would, eventually, display the same output on those display elements that the two had in common, while displaying traces of its computation on the additional display elements. (Though, if desired it could be easily arranged for all the additional display elements to be blank at the completion of the computation.) To produce a machine in which every step is displayed permanently on its own display is almost as simple: start with at least as many copies of the original transparent machine as there are steps in the computation in question, arrange them in sequential order, in each machine disconnect the outputs of each circuit leading to a memory location and connect them to the corresponding memory locations in the subsequent machine (if there is one).

While this demonstrates that there is no computation that cannot be performed by a timeless transparent machine, this should not deceive us into thinking that the two types of machine are equivalent. In considering computational power we focussed solely on input and output, and not on the intermediate steps by which the output was computed, and it is in the latter that the difference between the two types of machine lies. Like well-trained students in maths exams, transparent machines show their working. In contrast a machine with hidden storage may unchangingly display just its input on all screens up until the last which displays the output.

To see that the need to display intermediate calculations is not merely an

artefact of a simplistic method for constructing timeless transparent machines we must consider the limitations upon the circuitry connecting memory locations. In the case of logical systems we required each system's class of fundamental inferential steps to be recursive so that proofs could be decisively checked[19] and much the same restriction applies to the circuitry mediating the steps by which our computers proceed. Requiring each of these circuits to compute only recursive functions is justifiable on the grounds that we require each circuit to return a definite result after some finite period of time. But such a justification relies upon a hypothesis as to the computational power of the physical systems by which circuitry may be implemented. While the hypothesis that physical systems can compute only recursive functions (between non-analog values such as the sentences of our theories or the values displayed in our pixels) is tremendously plausible, it is far stronger than we require. We wish our computer circuitry to be checkable, at some remove anyway, and if our decision capacity is restricted to performing recursive functions then so too is that of the machinery we can check or, one would think, construct. Though this too is plausible, it is yet stronger than required. Our first interest is in the theories humans rely upon to derive predictions and not the tools we may somehow be able to construct or discover. It is our beliefs as to the limits of humans computational capacity that lead us to the quantificational standard of postulation, and it is by that standard that we must judge the need for postulating hidden objects. The restriction on the capacity of the circuitry connecting various memory locations in our machines to the computation of recursive functions is needed to maintain the analogy with our own inferential capacity, and in the end based upon a purely psychological thesis as to our own computational capacity.

While the computational complexity of our circuitry is limited to the calculation of recursive functions, the power of our theories and our computer programs is not; both we and they may generate(/accept) just the members of any recursively enumerable class as output(/input). As we noted earlier, systems whose capacity is exhausted in determining membership of recursive classes can, at best, only enumerate the members of recursively enumerable classes. What this does not make apparent is that to conduct such enumeration, to perform a check which is guaranteed to turn up the instance one is looking for if it exists, against a endless (and unhelpfully disordered) list of members one needs to keep track of how far along the list one has proceeded. Though the requirement of keeping track of how many items one has checked so far seems trivial this is only because relentless enumeration is a trivial method of checking for membership in a recursively enumerable class. Theorems are recursively enumerable and there

---

[19]Discussion of why this restriction is necessary for proofs to be decisively checked was deferred till section 4.6.

is no more certain method for determining theoremhood than an exhaustive enumeration of proofs, requiring of course that one keep track of which proofs on has checked so far (which given that one may arrange to inspect proofs in alphabetic order may be boiled down to keeping track of how many have been inspected so far). In practice, of course, no-one follows such a procedure, we follow heuristics which guide our choice of which proof to inspect next, we try to get from here —the initial premises— to there —the conclusion we wish to demonstrate— by means of whatever intermediate steps seem to lead in the right direction; with each step we take we further restrict the range of proofs left in consideration. What appeared as keeping track of how many proofs had been enumerated, now appears as keeping track of all the steps one has taken so far, and all the branches that one has abandoned as unproductive. The same memory our heuristics rely upon also underpins the completeness of our proof procedures and gives substance to our claims of the adequacy of our theories.

Clearly memory allows circuitry capable of computing only recursive functions to perform proof procedures, the argument for its necessity is straightforward. Computational procedures that do not use intermediate storage cannot vary in the time they take to perform from one input to another. The signal takes so long to pass along the wire, and there are no intermediate places in which values may be held for further computation[20]. Therefore any system which can discern membership in a recursively enumerable class without the use of intermediate storage, in a single step as it were, could also decide membership in that recursively enumerable class since only inputs which were not members of the class in question would not be recognised as members in the same finite period of time it took to perform. Such a system would have a computational capacity beyond the merely recursive. Memory is thus necessary and sufficient for the performance of proof procedures by systems which are, in their individual steps, restricted to the computation of recursive functions.

1.1.3   Another view of hidden objects

As well as viewing recursive and recursively enumerable classes, respectively, as ordered and disordered lists[21] we may view them as classes of objects that may be reached by recursive combinations[22] of some finite number of fundamental

---

[20]Well, to complicate the argument as much as possible, circuits with indefinitely many inputs may vary in the time it takes to produce its answer depending upon the number of inputs, but not their content.

[21]A list is disordered if you cannot guarantee that a desired entry is not yet to come up, no matter how many one has inspected so far. An alphabetised telephone directory is not disordered, since upon scanning all the surnames up to "Beatrice" without avail, one can tell that the surname "Alphonse" is not to be found further on.

[22]Paralleling the definition of recursive functions (see footnote 13 in §1.1.1) a recursive combination of operations is a combination produced by applying an operation to the results

operations. (A list is merely that special case in which the only operation transforms entries into their successors.) This approach is often applied wherever it is wished to provide clear means for distinguishing the members of infinitely large classes. In Wittgenstein's *Tractatus Logico-Philosophicus* it is applied to the task of delimiting the limits of language. Rather than giving a method for listing coherent assertions, or distinguishing coherent from incoherent meanings, Wittgenstein attempted to specifying the distinguishing characteristics of the fundamental elements (atomic sentences) and the means by which they might be combined (logical operations as specified in proclamation 6).

On this view the distinction between recursive and recursively enumerable classes is that for the latter there is in general no maximum length on the chain of operations required to reach any given object. In particular there is, again in general for there may be many exceptions, no way of telling of two given objects which requires the longer sequence of operations to reach other than brute inspection of all such sequences. To imagine a case in which this is not true, consider a class of sentences generated by operations which all increase the length of the sentences they were applied to. Of this class we could tell that A is generated by a shorter chain of operations than $A\&A$, if it is generated at all, since it is shorter. More significantly having determined that all sequences of n operations generate sentences longer than A and that no sequence of less than n operations generates A, we might justifiably conclude that A was not a member of the class in question without needing to look any further. This is not the case with theorems of quantificational calculus, some very long theorems have very short proofs indeed, and some short theorems can be derived only by means of extraordinarily long proofs. Since there is no other way of readily ordering theorems by the length of the proofs by which they are derived other than by inspecting all such proofs, the class of theorems is recursively enumerable.

Consider the range of predictions that follow from spotting a chair in the room next door. One expects that if the room remains unshaken and no-one enters or leaves it, then the chair will next be spotted in the same location; that chair-detectors will ping when pointed in the right direction; that observant people in the same room will usually report the presence of a chair if asked; that gossipy chair-spotters will pass on news of the chairs presence. The point is not merely that that the range of empirical consequences is open ended, but that there is in general no rule correlating the complexity of a prediction and the machinations producing the observation which fulfils it.

The empirical claim upon which our argument rests presses itself all too

---

of one or more other operations, repeatedly applying an operation to its own results until its first operand is zero, or determining the operand which minimizes the results of some operation.

forcefully on the minds of folk seeking to understand the world around them: Life presents us with no simple path to truth. In general there is no guarantee how long even the simplest of truths may take to discern. This also manifests itself in the need for ingenuity in devising experimental tests for empirical hypotheses; if the consequences of hypotheses (against a background theory) were recursive there would be an algorithm for experimental design. As it is, in both mathematics and science we must proceed more or less blindly and rely upon intuition and luck (let us flatter ourselves and call it insight) to find decisive results. But if this is so then it may require arbitrarily large amounts of memory for storing intermediate states through which such derivations (and hence their corresponding etiologies) proceed to determine the observable implications of our beliefs and hypotheses. Our statements about unobserved objects *are* that storage.

According to some apocrypha, Wittgenstein thought the worst mistake in his *Tractatus* was treating quantification as extended conjunction or disjunction. Whether or not the ascription is true, the error is prodigious. Quantification supplies access to the unlimited number of objects we may need to postulate in order to store intermediate steps in our derivation of predictions; lifting us to the level of the recursively enumerable by means of the postulation of objects.

Chapter 2

Weak empiricism

Weak empiricism is marked by commitment to the empiricist principle of justification, viz. all justified beliefs are justified by experience. Having observed the difficulties strong empiricists faced in postulating objects of immediate perception, it is natural to suspect that weak empiricists face similar difficulties in explicating the notion of experience, but happily this is not so. For having given up on finding unmistakable entities to provide certain foundations for our scientific theorising, weak empiricists may openly avail themselves of all that our fallible science has to reveal about the nature of experience. The doctrine of justification is itself an a posteriori claim, to be further refined as science progresses. If, according to current best science, observations of objects outside our own bodies are causal effects of particles impinging upon our skin or retinas, then in light of this a weak empiricist may say that the only evidence for our beliefs about external objects lies in such causal effects. The emphasis thus shifts from experiences to stimulations.

2.1 From experience to stimulation: the rejection of phenomenology

> First-person introspective access provides a rich and *essential* source of insight into our conscious mental life, but it is neither sufficient in itself nor even especially helpful unless used in a trained and disciplined way.
> Stanford Encyclopaedia of Philosophy entry on Consciousness[1]

> [Pareto's] Manuale is distinguished by the original idea of treating the laws of demand and supply, or rather the "curves of indif-

---

[1](Gulick, 2008)

ference" from which those may be deduced, as objective, capable of
being ascertained by external observation without the psychological
knowledge obtained through sympathy. The conception has been
criticised . . . as a needless abandonment of one large source of
information.

Francis Y. Edgeworth, *Palgrave's Dictionary of Political Economy*[2]

To focus upon stimulations, rather than experiences, as our source of knowledge seems dehumanising to some and its acceptance marks a difference between weak empiricism and phenomenology. Phenomenologists and their fellow travellers claim that natural science gives only a partial picture of the world — in taking it to be capable of delivering an exhaustive description of the experiencing subject we do violence to such subjects, treating them as less than they truly are. But wherein does such violence lie? To be sure, there have been theories which have been too limited, behaviourism is an infamous recent example, but the objection is not to this or that particular theory but to any product of scientific methodology.

The most substantial basis for the charge is that natural science only gives account of experience from the third person perspective[3]. As Laing puts it "Natural science is concerned only with the observer's experience of things. Never the way things *experience us*. That is not to say that things do not react to us, and to each other."[4]. If only natural science abided within this limitation and left the study of experience to phenomenologists, the lament goes, no harm would be done. But, of course many natural psychological theories *do* attribute experiences to objects. The objection is that such theories can only treat of experience insofar as it affects how objects "react to us and to each other", and to treat this as all there is to experience is mistaken and dehumanising. Instead, a distinct "science of being" is required, based upon the raw data of experience in a way that natural science fails to be, and hence able to explicate "the essential meaning of being a cognizable object or, what comes to the same thing, of being an object at all"[5].

If there is scope for such a science of being, it must be based upon data beyond the reach of the more prosaic natural sciences [6] and since natural science has access to all the public data, the data for this science of being must be

---

[2](Higgs et al., 1963)

[3]This criticism is made, repeatedly and at length, in the work of Laing, Husserl and Nagel among many others.

[4](Laing, 1967, p. 17)

[5](Husserl, 1964, Lecture 1, p.17)

[6]Husserl and other phenomenologists wish to subject this data to a transformation or "eidetic reduction" whose nature I do not investigate. Our concern is only with the raw data for this transformation; if this is not to be accepted, the nature of the consequent transformation is irrelevant.

private. The question is thus whether there can be a science based upon private data. Our consideration shall be focused upon whether there could be a science based upon private introspection, but since we make no special assumptions concerning introspection other than its privacy, the arguments we shall consider apply equally well to any other private sources of evidence. The fixation upon introspection serves to match the usual focus of phenomenology and will also allow us to see why naturalistic psychologists continue to exclude introspective observations even though, being good materialists, they do not believe it to be a source of irredeemably private information.

Prima facie the attractiveness of a science based upon introspective observation is palpable. Introspection is a source of knowledge of our own psychological states and appears to lend itself to codification and the development and testing of predictions as well as any other source of empirical evidence. Behaviourists who deny the existence of internal processes may balk at such a generous attitude to introspection, but they have trouble enough accounting for the data they do acknowledge[7].

Plausible arguments against the possibility of a science of being based upon the private observation of matters beyond the reach of natural science fall into two main categories. Instances of the first kind are intended to show merely that introspection is not private, that the matters revealed through introspection (or at least the projectible laws that govern them) can be discerned by other means; arguments of the second kind are intended to show that objects (or aspects of objects) observed solely privately cannot be the subject of collaborative public theorising.

Materialism supplies a very direct argument of the first kind. One of the hallmarks of material objects is that they can be observed by more than one person. Therefore if the objects and processes observed through introspection are (implemented as) material objects and processes then they cannot be irredeemably private. The knowledge gained through introspection might be gained through other means, hence natural sciences may scout introspective observation without thereby limiting their scope. While this argument is sound and valid, it fails as a criticism of phenomenology. Phenomenologists grant that the privacy of the (aspects of) the objects they observe is incompatible with materialism, but conclude instead that materialism is false, its acceptance a symptom of the very fixation upon naturalistic public science they abhor. A convincing argument against phenomenalism must come from more compelling premises.

It is difficult to see how any better argument of the first kind could be produced. There is no in principle argument against private observations: some of

---

[7]The refutation of reductive empiricism offered above (cf. §1.1) also serves *mutatis mutandis* to refute reductive behaviourism.

our observations are private, and all of them could be if we were sufficiently far apart from one another; and empirical studies seem irrelevant — phenomenologists may decry any empirical study revealing projectible correlations between states of the brain and introspective observations as failing to reveal how those states feel, unable to exclude inverted spectra or other inter-personal switchings of qualia. Thus it is the second type of argument that must bear the burden in making a case against phenomenology.

The most common argument for excluding introspective observations (and most widely criticised by phenomenologists) is of this second type and it is that introspective observations cannot be double-checked. The reading on a dial, the width of a band of shadow, or a strong odour of ether, can all be verified by more than one person. Not so, we are hypothesising, for the states revealed by introspection. While one's private experiments may be repeated, no one else is capable of checking their results.

Phenomenologists have contended that double-checking is misplaced in introspection: that our own experiences are what we know best of all and doubting them reflects a profound alienation that is both cause and consequence of an inappropriate application of naturalistic science [8]. But this defence of introspection is misplaced, based upon the illusion that natural science is the product of a single corporate entity. Scientists are not part of some mass hive mind. Each has access to only her own observations to garner the confidence to hold her favoured theory. While hearing other scientists confirm one's observation certainly increases one's confidence, it is not a matter of directly comparing the observations of two people and seeing that they are in agreement, but of two aspects of one persons experience, the initial observation and then hearing a confirming report, confirming each other. Each individual's knowledge of the world is based solely upon their own experiences the situation is not dramatically different when it comes to introspection. Insofar as empirical researchers were interested in discovering general truths, their work might proceed perfectly well (in principle) with each performing experiments in their own private laboratory, provided there was a common language for them to communicate their experimental procedures and results with each other. But here we come to the crux of the matter — can there be a language for reporting observations of otherwise private laboratories?

---

[8] Laing (1967) emphasises the neurotic nature of subjecting first-person experience to skeptical analysis at great length. A more considered and less strident version of the same theme can be found in Husserl (1964). A residual affection for strong empiricism can be sensed in such defences of introspection. The old yearning for certainty has, in the face of a disappointing world, been safely redirected to a custom-built realm beyond that of ordinary science; the complained of alienation caused by a loss of faith in a pleasing illusion.

Ordinarily languages are taught in triangular epistemic situations — where the teacher is aware of the cues to which her student is exposed, and attempts to shape his responses by re-inforcing the desired pattern of assent and dissent, or, in the other direction, the student is aware of the cues to which his teacher is exposed and models his responses upon hers. Such epistemic triangles are unachievable in the case of irredeemably private observations such as (ex hypothesi) introspection, but language learning is possible without such mutual access to the evidence.

While epistemic triangles are well suited to producing a ready supply of instances from which the student may project a general hypothesis as to which conditions warrant assent (or denial or neither) to some sentence and then tune that hypothesis until he is in conformity with his linguistic compatriots, they are not the only circumstances in which such exemplary training sets can be produced. If the student and teacher are both subject to the same cues reliably enough, and the student is aware of the teachers responses then a drive towards mimicry may be enough to instil the same response in him. Upon a sufficient basis of such co-incidentally learnt observation terms a sophisticated theoretical structure might be collaboratively built.

Even the co-incidence of like cues between students and teachers may be foregone so long as there are other public cues to guide the student's learning. If both teacher and student share distinctive publicly observable responses to encounters with certain types of private cues or reliably have such encounters under distinctive publicly observable conditions (or some complex combination of the two) then an astute student might calibrate his dispositions to assent and dissent from sentences against those publicly observable conditions. So it is that the terminology of introspective reports is to be learnt, if it is to be learnt from experience at all. If feeling pain is a private state one might yet learn to assert "I am in pain" from observing that one's teachers assert "I am in pain" while in publicly observable circumstances (and producing public behaviour) like those one suffers (and produces) when introspectively aware of pain.

The problem for phenomenologists is that it is that a mere chance correlation between private and public cues is not enough to found the knowledge of a language, or at least not the knowledge that others are speaking a language, and a correlation that is discernibly more than mere chance breaches the privacy of introspection One can imagine someone learning a language by chance; acquiring the relevant dispositions to linguistic behaviour as a result of imitating trees creaking "Lo! trees" as they are passed, the wind whispering "It is windy" as it blows etc., though such a run of luck is preposterous such learning still squares with our empiricist requirement that language be learnt from experience. If such a student reflected upon the encounters which lead to his learning language,

and discerned that they were a freakish series of flukes, he would still retain his ability to speak his language, but he would no longer have any reason to take such noises as assertions, nor to suppose that the corresponding assertions were true.

For those who insist on the privacy of psychological states, it is, so far as can be told from the public evidence, a kind of co-incidence that certain types of public bodily events are accompanied by private sensations. The situation is not quite the same as the learning by chance considered a moment ago, there is a natural (in part evolutionary) explanation of why there are English speaking people and why they are apt to cry "Ow. Stop it!" when sharply poked, but that explanation gives us no reason to presume that those reports (taken as reports of private sensations) are true. So long as those states are private there can be no reason to take other people's assertions concerning their own experiences of such states seriously. Though researchers so isolated might be able to communicate with each other, in the barest sense of being able to understand each other's utterances, so long as there was a better explanation of the noises others made than that they were assertions, and in particular reports of observations, there could be no collaboration — there would be no good reason to suppose the reports to be true. Each researcher, insofar as she wished to stick to justified beliefs, could only believe those claims she had verified herself, any genuine collaboration would be impossible

The problem is not that credible reports of introspective states would themselves constitute a breach of privacy, or at least not just that. Phenomenologists' insistence upon the special access provided by introspection would be satisfied if confidence in the reports of credible persons were the only way in which knowledge of the states introspected by others could be gained. That taking others as intentional beings, capable of generating utterances with content, of responding to reasons as well as causes, is essential to gaining knowledge of the states they introspect is perfectly compatible with phenomenologists' criticism of natural science so long as intentionality remains elusively unnatural. (Hence the sympathy with which phenomenologists tend to regard Brentano.) Our objection is not that credible reports constitute a breach of privacy, but rather that no such narrow breach is possible. The problem is (to adopt a Socratic device) that we do not have good reason to believe reports because they are credible, but rather that reports are credible because we have good reason to believe them.

A report is credible when we have good reason to believe that there is a reliable knowledge supporting (hence projectible) connection between the facts reported and the issuing of such reports. Typically such reliability is achieved by the epistemically accessible mediation of a process between the observation of evidence for the truth of the report in question and the issuing of such re-

ports, but their very accessibility places such processes and their antecedent observations squarely in the realm accessible by natural science. It may be that there is no such mediating process, but in that case the reliable connection must be fundamental, not the object of some special science, but an additional connection to be recognised by fundamental science, achieved by means of relations between (objects instantiating) properties every bit as fundamental as charge, spin, and mass are according to our current fundamental theories. Such lack of implementation would leave Brentano's thesis of the autonomy of the intentional untouched, but would rely upon epistemic access to the objects being reported incompatible with the special nature of a phenomenological science. Either the public vocable blasts (or inscriptions etc.,) are themselves good evidence (under some description) that the reports they are used to make are true by virtue of a reliable connection between them and their subject matter falling within the scope of some non-intentional natural science, or, failing that, there are some predictions of publicly observable phenomena captured by an intentional theory postulating those entities (and attributing to them those properties) we have hypothesised to be revealed by introspection alone, or there is no good reason to take reports of introspective states to be true. In breaking through epistemic barriers language never escapes alone.

This conclusion serves not so much to show that introspective observations can be excluded without limiting the scope of our public sciences, as that such private observations could never warrant inclusion in the first place. Our capacity to communicate our introspective observations suggests they are not so private as we have been supposing. In fixing upon such absolute privacy we have managed to refute only the most extreme and implausible sort of phenomenology, and have utterly occluded materialistic psychologists' reasons for excluding introspective observations from their evidential base.

A more moderate phenomenalism might concede that that there is no reason to find other people's reports of introspective observations credible so far as the purely *public* evidence is concerned, but maintain that the combination of one's private observations *together* with the public data might warrant such credence. On such a view the acknowledgement of one's own introspective observations would be an essential step towards recognising others as experiencing selves. The form of reasoning involved is known as "generalising from one's own case" and has come in for wide abuse as being too reliant upon an isolated case (oneself) to warrant the confidence with which we attribute states such as those we introspect in ourselves to those around us. The merits of this form of reasoning do not concern us directly however, for in making this concession to publicity of psychological states phenomenologists have already given the game away, and placed the entities and states observed through introspection, or at least the laws

constraining them, squarely in the public arena. The claim that belief in the
psychological states of others (or those aspects of them revealed solely through
introspection) relies for its justification upon some combination of the principles
that like causes have like effects (applicable if one is an outright epiphenomenal-
ist concerning the entities revealed by introspection, holding them to be caused
by physical events but themselves utterly without effect) or that like effects have
like causes (applicable if one is only a property epiphenomenalist, or otherwise
believes that the states revealed through introspection do have causal effects
and it is only some aspects of them which are, apart from introspective obser-
vation, without effect) implies that one can learn the language for reporting
introspective observations (or, more to the point, justify homophonic transla-
tion of others reports of their introspective observations) only insofar as one
can match up, however indirectly, one's own dispositions to behave with those
of one's supposed linguistic compatriots.

On this view of the learning of the language for making introspective reports
one begins, with one's teachers, by appealing to the psychological states of the
organisms in order to explain the behaviour (taken simply as movements) of
their bodies. In this there is nothing distinctive about psychology, physicists
also appeal to states of unseen entities to explain the phenomena they observe
— and this is one way of coming to understand what counts as evidence for the
correct application of these theoretical terms. Thus our psychological terms are
learnt first and foremost as theoretical terms, connected only indirectly with the
observations of organism's behaviour which they serve to predict and explain.

The matter does not stop here, for each of us has a more direct and (de
facto) private access to their own psychological states through introspection.
This introspective access is in addition to observation of our own behaviour,
and it is the latter that provides the entering wedge for learning to apply our
psychological terms on the basis of our introspective awareness alone. Having
mastered (more or less) the use of psychological terms on the basis of behavioural
cues, we can apply them to ourselves on the basis of our awareness of our own
behaviour and then correlate them with our introspective observations. The
method is a sort of introjection or projecting within; one starts at the surface
and works in. Having achieved a modicum of success at such introjection,
we can surpass the available behavioural cues, talking of covert pain, hunger,
remembered dreams and so on. But still we are warranted in this (have good
reason to believe we are using our terms as our peers use them) only insofar as we
can relate them to public touchstones of behaviour, physiology, or stimulation.
Introspection may give access to unexercised dispositions, but it is their public
exercise that provides the common arbitration. We all learn that hunger is
sated by food; if confronted with a feeling phenomenologically like hunger but

not sated by food, one might take the extension of "hungry" to include those having similar experiences, or perhaps deny that such feeling is hunger at all, or admit of introspective error. But if this decision is too far out of step with one's peers, they may well respond by abandoning homophonic translation. If such choices are beyond our peers capacity to detect, there is no warrant for assuming a common use of terms (recall that we are not here considering whether there could be a private language, but only whether it could be used for collaboration in the development of scientific theories).

We can now see why introspection is peculiarly unsuited to provide scientific data, combining as it does the theoretically laden inheritance of the teaching process with the immediacy of direct observation. In general, the more indirect a term's connection with experience[9] the more free one feels in re-jigging its connections with experience, and the more freely one accepts the re-jigging of others. So it is that "phlogiston" has been surpassed in favour of "mean molecular kinetic energy" but "heat" remains. This is not an absolute rule, for the sentences we use to report the most superficial of observations are also subject to change from time to time, but the change is much harder to push through. Such resistance serves a purpose, for without comparatively fixed connections between observations and the sentences used to report them, more theoretical claims would fail to gain any predictive grip on our observations. Our feeling tends to be that the deeper interstices of the world are harder to discern, and our speculations about them more error prone, though this is simply another reflection of which parts of our theories we usually find easiest to modify in the face of recalcitrant experience. The sentences we use to report our superficial observations, on the other hand, generally have an inflexibly tenacious connection with experience. So it is that "sunrise" has survived the Copernican revolution for some hundreds of years.

When those who taught us our psychological vocabulary — including terms such as "pain", "thinks", "believes", "greedy" and "weak-willed" — demonstrated correct use by applying such terms to ourselves or others, they relied upon their observations of our behaviour to guide them. Their use was theoretical, however due to the odd asymmetry of the situation, we have been taught to apply such terms to ourselves on the basis of our introspective awareness. In our hands, as applied to ourselves such terms are observational. Thus our reports of introspective observations are both vividly immediate and highly theoretical; vivid for the introspecting learner and highly theoretical for her teacher. It thus has the worst possible combination of immediacy and theory, combining

---

[9]We shall have to rely on this vague measure of observationality until we have vindicated the shift from vague experience to precise stimulation. Even after that we shall leave discussion of the notion until we turn to criticism of Quine's version of observationality as given in *Word and Object*. Until then Quine's measure will do as a first approximation.

the highly theoretically laden inheritance of the teaching process with the immediacy and inflexibility of observation. No wonder common-sense theories of the mind are so slow to change. Allowing introspective observation to count as scientific evidence has the unfortunate tendency to freeze psychological theories in place. In the extreme, when the knowledge that one is conscious, angry, upset, devout, desiring food or whatever else is revealed by introspection is put beyond question by being held to be more directly than anything else then one's psychological theory reaches paralysis.

These considerations do more than establish the positive motive for excluding introspective observations from the evidence for scientific theories, they also show such exclusion to be innocuous. Since it is only by projection (or introjection) from the publicly accessible data that the terminology for reporting introspective observations is learnt, the exclusion of introspective observations, though it requires the exclusion of sundry particular observations, puts no psychological laws beyond epistemic reach. Monitors at mission control *can* learn about distant environment from astronaut's reports by relying upon their observations of the environments that prompted the astronaut's responses before they left, together with their knowledge of how these responses generalise to novel situations and how translation through space affects both the astronaut's sensory sensitivity and the properties of the objects she engages with (i.e. usually not very much, since space-time is generally only slightly curved). The principles thus relied upon may be flexible, and may be modified on the basis of the astronaut's reports, but in extreme cases where the astronaut's observations (or their dispositions to respond to them) are contrary to our most basic expectations, then we are at a loss to understand them. In those cases where our understanding survives the disappointment of our expectations it is because there is a clear way of reforming our guiding theories in the face of the recalcitrant evidence of the astronaut's reports themselves. Likewise, in the case of introspection we can justifiably rely upon only those reports where we can project laws connecting the dispositions to issue or assent to such reports with the public evidence that exercises them (though we can project them into particular cases where there is no such public evidence).

We rely on publicly available data to learn the terminology we use to report introspective observation. Although in eschewing such reports some data is left out, no projectible laws are omitted, since it is only by reliance upon publicly establishable laws that introspective reports serve to communicate our observations. Although there may be truly private phenomena, there can be no co-operative study of them; no co-operative or communicable "science of being". Thus nothing significantly affecting the development of (public) psychological theories is excluded by barring introspective observation as a source of data.

This is not to say that introspection should not play any part in guiding our scientific enquiries; like evidence acquired illegally by the police, introspection may properly spark an investigation or suggest its direction; It may guide us in the framing of hypotheses and the search for evidence for them, but its use as a source of that evidence reveals no principles or laws that cannot be discovered by other means and tends to give the theoretical presumptions guiding the teaching of the language in which introspective evidence is described the inflexibility of observation statements.

2.1.1 Evidence naturalised

The justification of the exclusion of introspective evidence clears the last obstacle preventing a purely natural treatment of the evidence upon which our empirical theories are based. The first was side-stepped when the quest for certainty was abandoned and empiricists could allow themselves to rely unashamedly upon scientific findings concerning the causes and nature of experience. Still the fear persisted that some aspect of experience inaccessible in principle to the methods of natural science might be left out of an account of experience based purely upon natural science. Having seen that fear to be misplaced (or rather that no theory capable of collaborative development can do better) we are free to make unfettered use of scientific discoveries in our attempts to explicate the evidential basis for our empirical theories. The immediate use for such knowledge is to sidestep the difficulty of explicating the nature of experience.

We have seen how difficult the explication of "experience" was for strong empiricists, hamstrung as they were by the requirements of their epistemology. Even without the constraints of indubitability, immediacy and theory-neutrality however, weak empiricists fare little better. Our folk theories are still recovering from the influence of Descartes and the British (strong) empiricists, and our psychology is not in a position to render any unified account of the matter. It may be that the notion of experiences (entities which experiencers have or encounter) will never be revived in serious theories of the mind. Fortunately weak empiricism is not tied to any such notion and can survive the rejection of sense-data or experiences from one's ontology altogether.

Rather than trying to give an account of experience (which no-one is in a position to do) weak empiricists can avoid the issue by using a fragment of knowledge concerning our experiences that scientists have already uncovered; viz., that our observations of the external world are causal effects of particles impinging on our sensory surfaces. In light of this the empiricist doctrine of justification may be rendered as the claim that stimulations are the only justification of our knowledge of the external world.

In shifting to this doctrine, weak empiricists are not identifying experiences with stimulations (or anything else), nor claiming that stimulations comprise some mysterious indubitable 'given', nor holding that being caused by stimulation is either a necessary or sufficient condition of being an experience. None of these claims that strong empiricists were forced to make concerning their foundational sense-data need be made with regard to stimulations. Instead weak empiricists are simply relying upon a convenient bottle neck in the flow of information from the world to its observers. We are warranted in believing there is such a bottle-neck, by the conjunction of both empiricism and our scientific doctrines regarding the causes of experience. If some extra-sensory power, perhaps telepathy or clairvoyance, unreliant on the mediation of particles or propagating fields were found to affect experience, then weak empiricism would suggest shifting attention from our sensory surfaces to some other locus, or reverting to the unrefined claim that experience is the only source of knowledge.

The shift to stimulations is made without claiming that all experiences are the result of the stimulation of sensory organs, nor that all stimulation affects experience, nor that all experience is attended to and so results in belief. Empiricists have no reason to contest the existence of introspection or hallucination; only to maintain that such internal disturbances are not in general a source of knowledge of the *external* world. Nor is there reason to deny that nerves may tire and for this or some other reason be temporarily insensitive to the particles impinging on them. An empiricist may gladly concede that in considering stimulations generally to be a source of knowledge she is taking a conservative approach and admitting much that does not in fact produce knowledge of any kind. All that is maintained is that stimulations are our only evidence of the external world.

2.2   Stimuli

The prime motive for shifting the focus of our empiricism from experiences to stimulations was to avoid the difficulty of giving a naturalistic account of experiences. Having made this move we are squarely confronted with the problem of giving an account of stimuli.

Intuitively and unreflectively one tends to suppose that stimuli are types of patterns instantiated at observant creatures' surfaces. After all, stimuli are inputs and inputs happen where the world touches us. This intuition is reflected in Quine's focus on patterns of retinal irradiation[10], and before him in Berkeley's doctrine that the visual field is two dimensional. But the reasons for fixing on

---

[10]See (Quine, 1960)

bodily surfaces as the locus of stimulation are not absolutely compelling.

2.2.1   Looking under the skin

Initially at least, the reasons against choosing a tighter (i.e. subcutaneous) boundary are based upon a reasonable acknowledgement of ignorance. There is a fear of cutting too close to the bone and not being left with an experiencing subject at all. Descartes' focus on the pineal gland as the seat of the soul provides an extreme example. We now know that too much of our soul's work goes on outside the pineal gland for our empiricist doctrine to apply to its boundaries. But perhaps we still might move somewhat inside the boundaries of our skin to the receptor cells at the periphery of our central nervous system. Even granting that brains in vats (or skulls) are experiencing subjects and so come under the purview of our empiricist principle, we are currently ignorant of the subtle causal effects our bodies have on our nervous systems unmediated by the firing of its peripheral receptors. Perhaps factors such as temperature and pressure have a direct effect on centrally located cells and a vat would have to be much more sophisticated in providing input to its resident brain then merely mildly electrocuting peripheral ganglia to successfully emulate the range of stimulations to which humans are usually subject.

Behaviour not the only evidence

Many philosophers have been tempted to suppose that there are more substantial reasons for refusing to draw such tight boundaries. Wittgenstein succumbed to this temptation when he proclaimed "Only of what behaves like a human being can one say that it *has* pains"[11]. If behaviour were the only public manifestation of psychological states and activities, that would be a conclusive reason for locating stimuli at bodily surfaces. Our motivating epigram is that experience is the only source of evidence and "experience" is itself a term of psychological theory on par with pain, justifiably ascribable only to "what behaves like a human being" hence only to bodies, not nervous systems.

The naturalistically inclined are, naturally enough, suspicious of this claim that behaviour is the only public manifestation of psychological states. If materialism is true then psychological states and processes are comprised of physical states and processes which are epistemically accessible by means other than their effects upon behaviour.

---

[11](Wittgenstein, 1953, No. 283)

Mental states need not have behavioural criteria

Even granting materialism, Wittgenstein might yet defend his claim on the semantic grounds that behaviour is criterial for psychological terms, whereas other
evidence for psychological terms (e.g. neural firing) are mere symptoms. The
latter relying for their epistemic worth upon correlation with the former (i.e.
it is only because neural firing pattern X is reliably correlated with behaving
angrily that we can take detection of pattern X as evidence of anger). Wittgenstein's supposition that meaningful terms require criteria of correct application
together with the obvious truth that we learn to apply psychological terms on
the basis of observed behaviour, imply that behaviour is criterial for psychological (and other inner) processes; hence our empiricist principles only bear on
behaving bodies (and not nervous systems nor brains in vats), and no tighter
boundary than that of our bodies the locus of stimulation.

Wittgenstein's criterial semantics is however an unhappy halfway house between strong and weak empiricism. It is like weak empiricism in conceding that
the observable consequences of some events may be open ended, but like strong
empiricism in insisting that some decisively recognisable sub-class of these consequences are criteria. It is not absolutely clear that the class of criteria be
recursively enumerable, Wittgenstein is typically elusive on the point, but in
his work it is fairly clear that criteria are supposed to be decisive. If Y is a
criteria for X and we can all see that Y obtains then there is no more work to be
done ("Here we strike rock-bottom" is the aphorism that springs to mind), further disagreement is a sign of "grammatical" differences; but as we saw earlier[12],
the determination of membership in recursively enumerable classes is marked
by the need for mediate calculations.

Having adopted weak empiricism, acknowledged that some truths have diffuse observable consequences, examples of objects whose presence is without
criteria leap to the mind: alpha particles, dreams and electrons. Even if criteria
may be open-ended the rationale for criterial semantics is demolished. Once
conceded that the observable consequences of some theoretical fact are diffuse
and dependent upon other contingencies for their manifestation there is no reason to suppose that those consequence can be divided into the criterial and the
merely symptomatic.

Functionalist attitudes towards behaviour

At this point it is tempting to adopt a functionalist account of psychological
states. For functionalism does not require that those states have criteria but
only that they be apt to play a role in the causation of behaviour. It is clear,

---

[12]cf. §1.1.2

after all, that usually we attribute psychological states because such attributions are parts of theories which predict and explain behaviour. Functionalism need not be taken as implying that behaviour is the only evidence for these states but it does imply that it plays a special role. Whatever it is that has the functional role of anger will, if it physical, have effects other than the purely behavioural, but it is the behavioural effects, or at least the aptitude to cause such behavioural effects, that identify it as the state of being angry.

An extreme functionalist might insist that the mere capacity to have the requisite causal relations with behaviour is not sufficient to identify a psychological state. On such a view the dispositional connection to behaviour must be realised. Like Wittgenstein, such functionalists do not count brains in vats as capable of instantiating psychological states because such states are disconnected from behaviour. Such a view has much the same defects as the criterial semantics just dismissed, it involves treating some evidence of psyches, if not psychological states, as criterial and others as merely symptomatic. (Another aspect of this defect is that it implies the validity of the argument "I think therefore my body exists".)

A more moderate and plausible functionalism takes the mere aptitude for having such behavioural effects as a sufficient condition for being a psychological state. On this view, disembodied brains are to be admitted as fellow sentients because they would cause behaviour in accordance with our psychological theories if suitably ensconced. While it is the aptitude to having the appropriate disposition to cause behaviour that still qualifies systems of states as mental states a functionalist of this stripe is happy to find his behavioural evidence indirectly, and may rely, for instance, upon neurological findings to provide assurance of the correct behavioural dispositions.

Such moderate functionalism still provides good reason not to take subcutaneous surfaces as loci of stimulation. If psychological states are those coming "near enough" to realising our psychological theories, then there is an unavoidable vagueness concerning how much of an experiencing subject may be removed before it is destroyed — just where the line is between a brain in a vat and an artificially intelligent vat. Another reason is that behaviour remains the touchstone of psychological states — to show that a vat, or any sub-region of our bodies, contains an experiencing subject requires showing that those contents could be suitably connected to a body in which the requisite behavioural dispositions could be exercised. In taking a more restrictive boundary we double the work required since we have to build out again to a bodily surface in order to establish that there is an experiencing subject to be stimulated. Such labour might be required in cases where the normal bodily carriage is missing, but in such cases we can identify the stimulations as those impingements upon a hy-

pothetically reconstructed bodily surface that *would* have the same affect upon the fragment in question. Outside such exceptional cases however, nothing is accomplished by drawing a tighter boundary. Both to avoid such difficult reconstruction and to ensure that our chosen loci of stimulation contain experiencing subjects, we are well advised to seek our stimuli at or above subjects' bodily surfaces.

Neurological Essences and Expert Advice

One might still reasonably wonder if functionalism's appeal is a product of ignorance. If mental states might have some physical essence currently unknown to us, our reasons for fixing stimuli at or above bodily surfaces merely a product of ignorance. Recently psychiatrists have proposed a neurological basis for schizophrenia. The existence of a neurological implementation of schizophrenia is no novelty, their claim is stronger. It is contended that people are schizophrenic just in case they suffer from a particular neurological defect regardless of their behaviour; in more dramatic philosophical terms, that schizophrenia has a neurological essence. This would imply, among other things, that sentient creatures lacking the requisite physiology are incapable of being schizophrenic.

Hilary Putnam has proposed that there are several examples of kinds with hidden essences. Water is his most famous example, its hidden essence is supposed to be being H_2O; gold is another, its hidden essence is supposed to be its atomic structure. Though these essences were unknown through much of human history still these terms — "water" as used by 17th century Englishmen, "$\chi\rho\upsilon\sigma\circ\varsigma$" (the Greek term for gold) as used by Archimedes — were true of just those objects possessing the relevant essences, *even if* other substances were indistinguishable from water or gold by the experts of the day. The natural question to ask is whether we might be in a similar situation with regard to mental states — if they might have hidden essences beyond our current knowledge of which our behavioural standards are only partial fallible indicators.

Putnam gives his reasons for supposing that mental states do not have hidden essences in "Psychological Predicates"(1967), appealing to neurological equipotentiality, evolutionary convergence and the possibility of artificial intelligence to support the claim that creatures with radically different physiologies may be in identical psychological states. An argument based upon examples like these cannot show that functionalism is true (nor is it intended to) but only that if functionalism[13] is true then empirical considerations suggest that psychological

---

[13]In discussing Putnam one should be careful to distinguish his narrow psychological functionalism, which applies to states such as being angry, or being in pain but not contentive states like believing today is Tuesday, or hoping it will rain, and the more popular version of

states will not have physical essences.

While the empirical case for there being no common physical structure implementing (functionally individuated) psychological states is convincing, our current objective is to determine whether functionalism is susceptible to being overthrown by the discovery of psychological essences. If so, there may be cases where the requisite essences are atypically connected with behaviour (if connected at all) and yet still count as beliefs (just as albino tigers are still tigers), and organisms lacking the requisite internal structure whose behaviour is in accordance with the predictions of our psychological theories but lacking the appropriate internal structure are not believers (just as XYZ is not water).

Another reason for considering Putnam's explanation of his examples is his suggestion that they involve the linguistic division of labour[14]. On this view Archimedes "borrows" the term "χρυσος" from later experts who can distinguish gold from the ersatz-gold which passes Archimedes' unsophisticated tests. The division of linguistic labour is a tremendously exciting idea because it suggests that the storage units of our theories are total societies or linguistic communities. So far we have been taking stimuli as the basic evidence for our empirical theories, but if theories are held by societies then our empiricism would enjoin us to take as basic evidence for a theory not the stimuli of an individual, but the collective stimuli to which whole societies are exposed.

We are confronted by two questions: is linguistic labour divided as Putnam suggests, and then, whether through linguistic borrowing or other means, if psychological terms might be true of just those organisms instantiating hidden non-behavioural essences. Let us deal with them in order.

Part of Putnam's claim is that the linguistic division of labour relies upon special social roles and conventions. It is supposed to be something like the holding of positions or jobs at a factory: "in this 'factory' some people have the 'job' of wearing gold wedding rings, other people have the 'job' of selling gold wedding rings, still other people have the 'job' of *telling whether or not something is really gold*."[15] Like a factory, having a job is supposed to be something more than merely doing the job.

On closer inspection however this reliance upon social roles and conventions disappears. It turns out that the linguistic division of labour involved in Archimedes "borrowing" of χρυσος relies, in the first instance, upon experts being able to produce convincing cases rather than expert authority. "If we had performed the experiments with Archimedes watching, he might not have known the theory, but he would have been able to check the empirical regular-

---

functionalism which applies to both.
[14](Putnam, 1979, see p.146)
[15](Putnam, 1979, p. 144)

ity that 'X behaves differently from the rest of the stuff I classify as $\chi\rho\upsilon\sigma\sigma\varsigma$ in several respects.' Eventually he would have concluded that 'X may not be gold'."[16] Notice that the key feature is that we, or experts, are able to produce observations that would have convinced Archimedes (or at least have suggested to him) that his sample of twin-gold was not gold. If, for instance, he would have reacted by saying that our results were irrelevant to whether or not the sample was $\chi\rho\upsilon\sigma\sigma\varsigma$ then this surely would establish that he was not borrowing the term gold from current experts and we should not assess the correctness of his application of the term "$\chi\rho\upsilon\sigma\sigma\varsigma$" by comparison to the extension of "gold".

The ability to produce convincing evidence is no social phenomena. It doesn't rely upon social laws, or social conventions at all. If the relevant conditions are strikingly obvious over the next hill, it can be children who know the area rather than trained scientists who qualify as experts. Experts may understand wider ramifications of their evidence than laymen, but the borrowing depended upon the evidence being able to sway the borrower, not the lender. Cases whose significance is understood only by experts do not serve to divide linguistic labour.

On closer consideration we can see that the experts are completely superfluous in this "borrowing" of terms; the cases stand by themselves. $\chi\rho\upsilon\sigma\sigma\varsigma$ as used by Archimedes is fixed by the fact that *he* would conclude that "X is not gold" when it behaves differently to his standard samples under certain circumstances regardless of whether any expert then *or later* knows how to bring those circumstances about. What the expert knows that Archimedes doesn't, is how to perform these convincing tests, and how to describe them. That doesn't reflect any special distribution of linguistic labour, but merely the commonplace truth that one can have expectations one is incapable of specifying in words, and that expert theorists may be able to explain those expectations more lucidly.

So much then for the division of linguistic labour in the cases such as Archimedes' use of "$\chi\rho\upsilon\sigma\sigma\varsigma$" where putative borrowers have some rough and ready test for identifying the substances denoted by these terms, and the linguistic division of labour was called upon to explain how passing these rough and ready tests was not a sufficient condition of being in the particular term's extension. As we noted however the supposed borrowers were self-sufficient, though they were unable to *perform or specify* more rigorous tests they were able to *recognise* more rigorous tests for these substances *as* tests for these substances. But Putnam has a more extreme type of case on offer in which putative borrowers use a term whose extension they are unable to determine at all except by asking expert advice.

Like many others, Putnam is unable to identify elms. Despite this the term

---

[16](Putnam, 1979, p. 156)

"elm" as used by Putnam still manages to be true of just elms. Putnam's explanation of this is that the linguistic community of which he is a part includes some experts who are able to identify elms, and the rest of the community borrows the term "elm" from these experts by a social convention of deferring to these experts' elm judgements. Much of this account is unexceptionable, clearly the term "elm" as used by Putnam is true of just elms, and clearly this is dependent upon Putnam having expert compatriots who can call an elm "an elm". The more questionable (and more exciting) part of Putnam's explanation is that this borrowing of terms depends upon linguistic affiliation and social convention. This would be exciting because it implies that we cannot determine an individual's beliefs in isolation from the rest of their community.

As it turns out however, deference just doesn't do the job. Most of us have at one time or another gone along with a supposed expert's assessment of a situation despite private misgivings. One of the famous hazards of employment (and particular soldiering) is being coerced into going along with the claims of experts in the full knowledge that they are wrong. No deference to a mistaken expert elm assayer (nor even a convention of such deference) is capable of making "elm" true of a beech. Nor is this because there are more authoritative elm assayers to over-rule the original expert's mistake. In a suitably unhappy situation, the views of the people to whom there is a convention of deference may be so entrenched as to be practically immune to contrary evidence (such evidence being explained away as hallucination, or communist/counter-revolutionary/liberal/right-wing dis-information). Such blind pig-headedness is widely suspected of being common in the authorities of the medieval Catholic church, the Soviet government at its peak, and the current American administration, and while such suspicions might be misplaced, they are surely comprehensible.

In the other direction, deference isn't a necessary condition for the "borrowing" of terms. One can borrow terms from those one publicly repudiates so long as one believes them, however secretly, to be true. One can defer to Lysenko as far as humanly possible and scorn Mendel as bourgeois, but so long it is Mendel who one believe to be correct it is he from whom one's terms are borrowed. This point is so obvious once pointed out, and semantic theories so dubious by contrast, that it is difficult to argue for it. But we may take what comfort we can in the support of well-known plausible semantic theories: Fodor's semantics of nomological locking implies that it is Mendel from whom one's terms are borrowed since it is his assertions to which having the requisite formulae in one's belief box are locking; Lewis's account of the introduction of theoretical terms carries the same implication since, ceteris paribus, it is the things Mendel declares to be P that are the nearest and near enough realisers of one's P-theory; brute causal connection theories likewise — Mendel's assertions

mediate a causal relation between the P instantiation and P-assertion.

The dependence upon belief rather than overt deference is in striking difference to genuinely distributed labour found in large companies. Imagine a company in which one of the positions is that of Elm Assayer and that the declarations made by Elm Assayer have a substantial effect on the position holders within the company. Such declarations might lead others to build protective barriers around supposed elms while mere beeches are reduced to woodchips. Importantly these other employees may be acting merely because of their knowledge of the Elm Assayers role, and not because of any belief about the Elm Assayers ability to assay elms; indeed they might not even know that much, instead being conditioned (after years of in-house training) to simply accept declarations on the topic of elms made on Elm Assayer stationary. Here we have intellectual labour divided to the profit, at least potentially, of all. However such gains are made, if at all, precisely because knowledge is not shared. Neither company directors nor mercurial shareholders care one whit whether barrier builders believe that the Elm Assayer can identify elms long as they obey the company's conventions regarding deference to the Elm Assayer on the topic of elms. By the same token, they care not one whit if the barrier builders acquire beliefs about either elms or the term 'elm' by means of such deference. But here we can see that though social convention and the 'borrowing' of terms are not mutually exclusive, each renders the other pointless, at least for company purposes. To the extent that the builders defer to the Elm Assayer because of social convention they do not acquire the ability to use the term elm any more than they would by deferring to someone incapable of identifying elms. If, on the other hand, they believe in the Elm Assayer's ability to recognise elms (whether because of consideration of the corporation's operations or otherwise) no special convention is required. The office of Elm Assayer may persist since time will be saved if the builders do not have to find their own elm-recogniser, but no special convention of deference is required, only the belief that the Elm Assayer knows his elms. If the belief is sufficiently wide-spread the company can save on in-house training and drop the Elm Assayer's special stationary.

Kant claimed the commandment to love thy neighbour as oneself must be taken as a commandment to *act* as though one loved one's neighbour because love is not under the control of the will. The commandment's interpretation is as may be, but a superficial sham will not serve to divide linguistic labour. The kind of deference required isn't just going along with the declarations of the designated experts, actual belief is required. But of course, unless things have gone badly wrong, most beliefs are even less under the control of the will than love. Relevant evidence can be sought out or avoided, but that is like being able to control one's pulse rate by exercising heavily or lazing about, and

not to be confused with being able to choose one's beliefs. Beliefs, in the truth seeking intellect, are beyond the scope of direct choice, hence beyond the scope of convention. Though horrible scenarios in which beliefs are formed in response to social pressures are imaginable (see the end of *1984*) as yet our society is not so intellectually bankrupt as to command such control over our beliefs that a conventional underpinning of the borrowing of terminology is possible.

This does not show that there is no scope for convention at all in such borrowing, only that it plays at most a subsidiary and in principle dispensable role. We might adopt a convention of allowing only recognised experts to wear white lab coats, but the linguistic labour they perform is shared with us because we believe white lab coat wearers to be reliable spotters of whatever it is they are supposed to be experts about. This does not make linguistic "borrowing" conventional any more than the traffic-signal convention that red means stop makes not driving into cross-traffic conventional. Societies in which convention is elevated to the point of determining who to believe (rather than marking who is believable) can end up planting wheat in snow.

The application of all this to the question of whether current psychological predicates might be true of just those psyches instantiating the corresponding physical essences is fairly straightforward. None of the points considered show that mental states cannot have physical essences, nor that the current best tests for the instantiation of any psychological state are infallible. However, they do show that if functionalism is true *now*, and the only evidence relevant to the extension of psychological predicates *as currently used* is behavioural evidence (or evidence of behavioural dispositions), then that behavioural evidence is not subject to being overthrown by the discovery of any physical substrate.

Archimedes did not in any useful sense borrow his term "χρυσος" from modern chemists or other experts. He did however, take as evidence of being gold the behaviour of substances under situations he was unable to produce or describe. In parallel manner, the extensions of our psychological terms may depend upon behavioural dispositions of our subjects even where we do not know how to produce the conditions which exercise them, or know how to describe those conditions or know how to describe the "criterial" response (or even that there are such conditions). In this way even our current best tests of psychological state may be incorrect. However, if the the only evidence that we currently take as being relevant to the extension of psychological predicates is behavioural, then no non-behavioural tests for mental states can over-ride our current standards. If functionalism is correct then we *now* do not count any non-behavioural evidence as relevant to the possession of psychological states and hence cannot "borrow" any future expertise on that basis. The extensions of our psychological terms cannot depend upon neurological matters if *we, now,*

do not recognise such matters as relevant.

These considerations do not show that functionally individuated psychological states cannot have physical essences, only that behavioural evidence cannot be trumped by physical/neurological/other non-behavioural evidence concerning the substrate implementing those states. If functionalism is true and someone is in a state that has all the appropriate connections to behaviour and stimulation (including those we may currently be unable to specify) then they are angry (sad, hoping that today is Tuesday etc.) no matter how atypical their physical implementation of that psychological state and *no matter how psychology develops in future*. If a psychological state was realizable by only a finite number of physical arrangements then being so arranged would be essential for being in that psychological state. In shifting to a functionalism which acknowledges the material implementation of psychological functions, psychologists do not expose their psychological theories to neurological evidence, except in the unlikely event that type physicalism is vindicated and the psychological and neurological evidence is guaranteed to always coincide.

Why is it that none of this applies to schizophrenia? Scientific practice tends to trump philosophical speculation with good reason, so if the arguments just given imply that psychiatrists must be mistaken in thinking that schizophrenia has a physical essence then that is good reason to think our arguments must be flawed. Fortunately, in this case drugs really are an answer. Psychiatrists, unlike psychologists, are not restricted to talking to their patients, the taxonomy deployed by psychiatrists can take into account the effects of drugs and invasive surgery, treatments that involve more than the mere application of stimuli. Hence the taxonomy of psychiatry does not have to correspond to that of psychology. In particular, it turns out that "schizophrenic", in the sense in which it has a neurological essence rather than just being a constellation of behavioural dispositions, isn't a psychological term at all (or not at least if functionalism provides a correct analysis of current psychological terminology).

Though everything we currently know about psychology and the role and nature of psychological explanations suggests otherwise, a similar turn might effect all of psychology. Existing psychological terminology might wither away just as the humours have. All that the replacement psychology might have in common with the present is that it seeks to explain behaviour, but it might not limit itself to such explanation. Radical neurological reductionists such as the Churchlands have been arguing that this is precisely the course that psychology should take and, at least at first glance, many cognitive scientists agree.

While we must remain open to the possibility of such radical revisions, that is not reason against building upon our best current empirical theories — we gave up on the hope of a first epistemology when we gave up on strong empiricism.

Insofar as we take beliefs to play an important role in explaining our behaviour, take functionalist psychology as giving a correct account of the nature of beliefs, and seek to delineate the evidence which justifies our beliefs, we are well advised to stop short at our sensory surfaces and take our stimuli there. The further inwards one goes, the further outwards one has to reconstruct to be sure of the presence of an experiencing subject.

2.2.2 Looking at sensory surfaces: Problems with inter-subjective identification of stimuli

> Out of timber so crooked as that from which man is made nothing entirely straight can be built.
>
> (*attrib.* Immanuel Kant)

As we have just seen, functionalism gives good, though not insuperable, reason to delve no deeper than our surfaces in search of stimulations. It appears however, that no stimulatory loci and no version of stimuli can do the work required of them. The difficulty arises from the confluence of two constraints.

Our objective is to delineate the scope for philosophical contribution to our common pursuit of knowledge; what it is exactly that philosophers can hope to contribute to our theories of the world. We are of course only interested in public theories, theories which can be developed, tested and believed by more than one person, otherwise there can be no hope of collaborative work nor any communication of the results. One clear constraint upon the identity of theories (and hence upon what is required for two people to believe the same theory) is identity of empirical content. If one person's theory is falsified, and another's substantiated by the same evidence then they cannot be the same theory; by the same token, one piece of evidence cannot both confirm and infirm the same theory. It appears that making sense of this constraint depends upon making sense of the notion of two people being exposed to the same evidence.

If we are to take stimuli as the basic evidence upon which our theories are built then this stricture must apply to them. Different people starting with the same beliefs (or at least the same relevant beliefs) and exposed to the same stimuli must have equally good reason for drawing the same conclusions. If the principle that the substitution of identicals preserves truth is to apply to stimuli while respecting their connections with theories, then stimuli carrying different evidential import cannot be identified. This is not to suggest that each sentence need have it own proprietary range of stimuli verifying it; the considerations of §1.1 have led us to take stimuli as evidence of whole theories and not individual sentences. Nor that each individual stimulus is associated

with a range of theories for which it is evidence; presumably most stimuli, taken as types of momentary events, are compatible with most plausible theories. The constraint is that evidence is composed of stimuli, and these evidential atoms must be individuated with respect for their evidential import: if stimuli sequence E is composed of stimuli $S_1, S_2 \ldots S_n$, and is evidence for some theory T_1 and against theory T_2, and S_i' is an identical stimuli to S_i for some i greater than zero and less than $n + 1$ then substituting S_i' for S_i in E produces a piece of evidence which like E is evidence for T_1 and against T_2.

In Quine's *Word and Object*, this principle is tacitly relied upon in the definition of "observation sentence". Quine defines the stimulus meaning of a sentence for a speaker as the ordered pair of the class of the stimuli disposing the speaker to assent to that sentence, and the class of stimuli disposing the speaker to dissent from that sentence, and goes on to stipulate that "an occasion sentence may be said to be more observational the more nearly its stimulus meanings for different speakers tend to co-incide"[17]. Uniformity of stimulus meanings is a plausible mark of competent speakers taking current stimulation as sufficient evidence for the truth of the sentence in question only on the assumption that stimuli keep their evidential import unchanged across different subjects.

The next constraint is that stimuli be individuated only by properties that can be distinguished by the subjects exposed to them. Among other things this requires stimuli to be individuated solely by their causal powers[18] and not, for instance, by their etiologies. A perfect counterfeit ten dollar note and a real ten dollar note, for example, must be taken as giving rise *ceteris paribus* to the same stimulus; indeed the rationale for counterfeiting depends upon it. Any two patterns of events at the chosen stimulatory loci having the same causal powers must be instances of the same stimuli. This restriction on stimulus identity avails of direct test when relativised to a given population: any two patterns of events at a stimulatory locus count as distinct stimuli just in case some training regimen would induce some member in the relevant population to differentiate between them.

It is readily seen that taking our bodily surfaces as stimulatory loci cannot give rise to stimuli satisfying both these constraints. As we have already observed, many theories have empirical implications that are only recursively enumerable, the list of potentially supporting evidence is open-ended in a way that resists simple generalization. As Fodor puts it, one can tell that a dog is present by seeing dog-shaped sights, hearing woof-shaped sounds, hearing dog detectors go off, listening to dog reporting gossip, seeing the merest ripple in

---

[17](Quine, 1960, p.42)

[18]see (Fodor and Society, 1987) where Fodor argues that all scientific taxonomies must depend solely on causal powers

dog infested waters, by means of radar, sonar, telescopes, microscopes, hearing aids, bifocal lenses, and other apparati[19]; and just as the list of apparatus by which dogs can be detected is open-ended, so too is the list of sense organs by which dogs may be sensed. Just as there is no non-relational physical property possessed just by dog-detectors, there is none possessed just by dog-detecting sensory organs.

Though this problem for the individuation of stimuli is rendered most obvious in the consideration of creatures with wildly different sensory apparatus (What pressure on a human frame corresponds to a five newton force perpendicular to the upper carapace of a Martian or a servitor-5000? How are our visual stimuli to be matched against those of creatures with compound eyes or differently shaped retina?) no fantasy is needed to give rise to it. Ordinary variations in the human physique present enough of a problem just as they are. Whereabouts on John's and Mary's arms do two pokes have to be for them to count (ceteris paribus) as instances of the same stimulation? We are neither all of the same size nor even of the same proportions. Even between the sharpest of sight, retina vary in curvature, size, and positioning of the blind spot. Such variety is not merely superficial, but is found at every level within our nervous systems. If we are brains in vats, the vats are calibrated to our idiosyncratic nervous systems.

2.2.3 Looking above the skin

The problems in individuating stimuli are due, in part, to the variations in our bodily proportions, hence it is natural to wonder if they might be circumvented by choosing more uniform surfaces above them. Fixing our stimuli as occurring at the surface of spheres centred subjects' centres of mass for example, would allow meaningful talk of impingements on different subjects' stimulatory loci occurring at corresponding points.

To object that such stimuli do not mark the boundary between experiencing subject and experienced world is beside the point. The shift to weak empiricism freed us of he need to demarcate that boundary. The orthodox fixation upon events at our sensory surfaces is justified by the combination of the empiricist principle of justification and the a posteriori discovery that the information bearing causes of experience are events at our sensory surfaces. Instead of attempting to determine what, if anything, is "given in experience" (for perhaps there is no such theory neutral given at all), we are free to seek a more amenable bottle-neck in the flow of information to us from the external world.

A fixation on more removed surfaces is justified by these two principles

---

[19]Paraphrased liberally from (Fodor, 1998, p. 77-8)

together with the physical principle of locality. This last principle assures us that the effects of the more distant world upon our experiences are mediated by physical entities of some kind (particles, waves, forces, fields or what have you) passing through a closed surface above our skins. Though this principle marks another point at which an empiricism based upon such stimuli would be open to refutation, that does not render it unacceptable.

Fixing upon such removed surfaces restricts the range of evidence under consideration to some extent, and hence the subject matter of the theories based upon that evidence. A surface three centimetres above the surface of one's body is a perfectly good bottle-neck in the flow of information from much of the world but it needlessly chops off consideration of knowledge of one's own success in shaving and other proximate affairs. Even fixing on our bodily surfaces had something of this, consider the painful knowledge we each have of our internal injuries. The exclusion of our knowledge of affairs at such close quarters is undesirable, but we may be confident that it does not prejudice our current investigation to the same extent that we are confident that the scope for philosophical contributions to knowledge is not a product of our diminutive size or the variations in our bodily proportions.

Surfaces cued to bodily proportions would vary as unhelpfully as our bodies themselves, and the injunction against subcutaneous loci forces a choice of stimulatory loci so over-sized that no part of any contained subject can project through them. Stimuli then are patterns of physical entities passing through such uniform subject centred surfaces.

Unfortunately such stimuli cannot be individuated in a fashion respecting their evidential import; for though such loci allow us to individuate stimuli without the need to match the postures and proportions of their subjects they also leave us ignorant of those subject's machinations. A subject who catches a ball and throws it in another direction is exposed to stimuli which are compatible with very different theories to one who witnesses a ball fly towards her, float about her body and sail off in another direction. Even more commonly, the directions of a subject's wanderings are usually an effect of the forces they exert on the objects around them, and to ignore those forces leaves us with no way to distinguish the stimuli of subjects who are merely floating about rather than exerting themselves.

It might be thought that this last difficulty might be met by taking the forces subjects exert on their surroundings into account in the individuation of stimuli. If subjects were in micro-gravity, and making contact with only a small number of middle-sized objects, this project might succeed. But we are all in a strong gravity field; the more heavy-set among us have very immediate reason to believe our supports to have greater bearing capacity than the light-weights.

Equally, we are all surrounded by an atmosphere; the wisp of smoke that brushes over your body has no ready counterpart in any wisp of smoke blowing across mine, we must rely upon our observations of each other or third parties to coordinate our theories about such matters. Similarly the pressures we feel when immersed in a flowing stream — evidence of the speed and turbulence of its flow — have no ready comparison between one body and another; we must rely upon standardised tests to calibrate our theories and make sure we are expressing them in the same language. Even air as it moves around us, or as we move it around by our gestures, where it is not inconsequential, can give rise to stimuli which are not directly comparable between individuals despite being evidence of the evidence of the air's viscosity and density.

Variations in our physical frames have proved to be an intractable confounding factor in our attempts to individuate stimuli. Our bodies are different from each other, each of us learns of the properties of her particular body, in part, through observing its effects upon its environment, and learns about our particular environment through observing its differing effects. Given this it would be disastrous for our stimuli to be identical.

Quine and Davidson's shift to distal stimuli

The lack of identical stimuli across subjects forces those who seek to build their accounts of translation upon a foundation of matching observations to shift their search for such foundations from proximal to distal stimuli — from types of events at or somewhat above sensory surfaces to more distant events. Just this shift can be seen in Quine's published position over the course of his lifetime (though the rationale for it is only dimly acknowledged) and it reaches full flower in the work of his student Donald Davidson.

Early in his career, in *Word and Object* (1960), Quine identified (visual) stimuli with patterns of ocular irradiation. By the time of *Roots of Reference* he took stimuli to be classes of sensory receptor nerves and abandoned attempts to identify the stimuli of different individuals. This is no minor change, it renders his original definition of "observation sentence" incoherent and, on his account, observation sentences are critical in connecting theories with their empirical content and in both staging and resolving empirical disputes. On the earlier, superseded version, highly observational sentences were those whose stimulus meanings varied little between competent speakers. On the later view (expressed in *The Roots of Reference* and *Pursuit of truth*) a sentence is observational if its stimulus meaning is unvarying for each individual and "its truth value, on any occasion, would be agreed to by just about any member of the

community witnessing the occasion"[20]. Here, grudgingly, is the shift to distal stimuli — witnessed occasions. So grudging that Quine quickly suggests the shift to distal stimuli is a matter of mere imprecision: "The definition speaks of joint witnessing. In a more precise statement, it would speak of witnesses subject to receptually similar impingements[21]"[22]. As we have seen however, no such more precise statement is available since no local individuation of stimuli can preserve their status as evidence. Quine's suggestion that the individuation of stimuli according to receptual similarity is a merely technical problem — difficult because of its complexity but not for any more principled reason — is painfully reminiscent of strong empiricists' excuses for the lack of successful reductions to sense-data, and incorrect for much the same reasons.

The grudgingness of Quine's shift to witnessed occasions as providing the evidence for the identification of observation sentences, together with the expressed hope for a more precise account of such occasions in terms of receptual similarity, explains the laxity with which he is willing to delineate just what counts as 'an occasion' and how much of it need be witnessed. Left unchecked however, distal stimuli expand to subsume the entire universe. This is in effect the position adopted by Donald Davidson. Davidson denounces the distinction between theory and empirical content as unintelligible, "It is itself a dogma of empiricism, the third dogma. The third, and perhaps the last, for if we give it up it is not clear that there is anything left to call empiricism."[23]. Once stimuli go fully distal there is no distinguishing them from the world our theories are supposed to be about, no distinguishing one small part of the great distal stimulus that is the world as evidence, and the rest as posited objects.

The difference between the two positions is striking. Quine's approach to translation, in brief, is to determine the empirical content of the foreign speaker's theory and thence to read into her utterances a consistent inferential structure supporting that content. (Hence the close link between the under-determination of science by evidence and the under-determination of translation.) Quine explicitly extends charity only to sentences universally held true under all stimuli, and the logical structure of a foreigner's theory, taking its empirical content to be discernible by virtue of the independently identification of observation sentences and the stimuli that prompt assent to them. This charity excludes the attribution of logical inconsistencies, leading him to reject the possibility of pre-logical people (susceptible of asserting to patent contradictions) as a myth.

Davidson on the other hand holds that there is no distinguishable class of observation sentences, and no empirical content to be isolated, or at very least

---

[20](Quine, 1990c, p. 39)
[21]i.e. similar classes of firing sensory receptor nerves
[22](Quine, 1990c, p. 41)
[23](Davidson, 1984b, p. 189)

that it cannot serve to guide translation. Instead charity is to be extended more widely; translations are better for assigning sentences "conditions of truth that actually obtain (in our own opinion) just when the speaker holds those sentences true."[24]. For Quine gross illogicality was mythical, for Davidson all gross error is mythical. In the absence of any separately determinable empirical content, we must presume upon the truth of a foreigner's theory to guide our understanding of it.

Simplicity and Charity in translation

On Davidson's view the basic evidence by which translations are justified is the dispositions of speakers to hold sentences true, correct translation realises the optimal compromise between simplicity and taking foreigners' assertions to be true. This appeal to charity and simplicity has the dubious virtue of leaving things in a vague enough state to make his account of translation very difficult to refute, but it leaves a range of questions without good answers.

The simplicity Davidson invokes is not that which guides choice between empirically distinct hypotheses. The merits of the scientific notion of simplicity are difficult to untangle, but plausible principles can be discerned: Occam's injunction against multiplying entities beyond necessity is one aspect of this preference for simplicity and is often well justified — mechanisms with fewer parts tend to be more reliable, less likely to come apart, and hence offer less chancy explanations of their consequences. Also, each extra concrete entity must, in the nature of things, leave traces of its existence other that those it is invoked to explain, and with more such it is more unlikely that these collateral effects should have been over-looked. A more general principle of apathy or least change is also justifiable on much the same grounds; gross modifications of a theory are apt to modify its empirical content concerning more than just the matter at hand and leave it at odds with previous observations. The notion of simplicity Davidson relies upon possesses none of these virtues however; for Davidson invokes simplicity to choose between theories which fit equally well with the basic evidence (conditions under which sentences are held true).

Divorced from such empirical cues for its evaluation, simplicity appears to boil down to a measure of how neatly a translation (or other theory) can be stated within the resources of some alternative lexicon. Fix the primitive terms of the target language and then a translation is simpler for delivering less convoluted logical combinations of those primitive terms as the equivalents of foreign sentences, or for requiring less computational resources in producing those translations. But such a standard of simplicity, relative to our language (if the target

---

[24](Davidson, 1984b, p. 196)

language of the translation is own native tongue), is too parochial a matter to
do the job required.

On Quine's account determination of the stimuli that are compatible with a
foreigner's theory provides the essential entering wedge upon which translation
is based; by contrast Davidson holds the determination of compatible stimuli
to be a superfluous afterthought. In particular, a non-negotiable part of David-
son's account, is that the stimuli compatible with a subject's assertions can only
be determined after a translation for his utterances has been found. Simplicity,
an essential goal of good translation, is thus involved in fixing the empirical con-
tent, the compatible stimuli sequences, of a subject's utterances. This requires
there to be alternative translations taking the subject as asserting empirically
distinct theories, and yet compatible with the same basic evidence — the same
dispositions to hold sentences true.

This requires (at least) two different interpretations of foreign sentences,
fitting all the evidence (i.e. taking the subject as having the same dispositions
to hold sentences true) yet differing over which stimuli sequences are compatible
with the foreigner's theory. Identical conditions give rise to identical stimuli
(within the one individual), hence translations which differ over their assessment
of the empirical content of the subject's theory must differ over which sentences
are true under some fixed conditions. Despite this, these alternative translations
must take the subject as holding the same sentences true under those conditions.
Hence simplicity must determine (to some extent[25]) whether or not the subject
is in error, and if so which of his assertions are mistaken.

Without an over-riding standard of simplicity, common between different
target languages there is no reason to suppose that translation establishes an
equivalence relation — that whenever the simplest translation of foreign utter-
ance S into Junglese is J and into urban is U, then J and U must be the simplest
translations of each other. At the extreme J and U may not even have the same
truth value, and hence cannot be translations or each other. Hence sentences
may be rendered into Junglese differently when translated directly and when
translated via a language associated with a different standard of simplicity.

There remains no point in maintaining allegiance to one language's associ-
ated standard of simplicity. Better to recognise the indeterminacy of translation
than insist that it can be resolved (to some extent) by recourse to a parochial
standard of simplicity. If disparate translations into Junglese and urban are
both optimally simple by their own standards there is no point served by dis-
missing the translation of some third tongue via one or the other as incorrect.
In the face of such indeterminacy we will of course choose the translation that is

---

[25]For, though there is no suggestion of this in Davidson's account, there may be errors
attributed on other grounds by both simple and complex translations alike.

most easily produced, or upon production most easily comprehended, or otherwise optimising costs in terms of whatever resources we are stretched to provide, but that can be properly recognised as a matter of convenience and need not lead to the dismissal of other translations as unacceptable.

Such relativism may be acceptable (contra Davidson) in the allocation of truth values to the translations of sentences of an empirically inadequate theory, but it is grotesquely misplaced in assessing the empirical content of that theory in the first place. There is only comedy in imagining a scientist who, in applying for further funding, says "Well yes, our experimental results were contrary to our predictions *in English*, but in Uzbekistani our theory was borne out.". It is absurd to think that the empirical content of our own theories, the range of stimuli which do not give us reason to re-evaluate our hypotheses, is determinate only relative to a translation into a further tongue. Not the least problem is that such relativity appears to lead to an infinite regress. Presumably the foreigners are in much the same position, the empirical contents of their theory — including their theory concerning the translation of our utterances — subject to much the same indeterminacy; for we are not to suppose that reliance upon simplicity is the result of any peculiar flaw in the language from which translation is sought.

A somewhat similar allegation is levelled at Quine's doctrine of ontological relativity. According to that doctrine there is no saying what the ontological commitments of someone's theory are except by re-interpreting it in the terms of some background theory, and the resulting commitments are relative to both the background and the interpretation. The ontological commitments of the background theory are themselves in turn only specifiable relative to some further theory. While this leaves scope for infinite regress, it is not, or at least not obviously, an unacceptable regression. The reference of some foreign term is adequately explained to someone (relative to their tongue, and the method of translation) when re-interpreted in a language they understand. (Which is to say when re-interpreted in terms they use in such a fashion that our strictures upon good translation apply to them in turn.) To take the Junglese term "gavagai" as referring to rabbits is simply to translate "gavagai" as "rabbit". In one's own case this amounts to resting with homophonic translation, taking one's own assertions at face value.

In the case of the relativity of empirical content however, it is not clear that there is any point at which empirical content can be stabilised The nearest parallel to taking assertions in one's own tongue at face value, is taking them to be true when one is disposed to assent to them, and while this is no doubt indispensable guide to translation, on Davidson's account it isn't sufficient to fix the empirical content of one's theory. Holding sentences true was supposed to be the fundamental evidence which all theories of translation, including the

unacceptably complex, were supposed to fit.

All these problems would be resolved by a suitably unparochial, trans-linguistic, standard of simplicity by which to judge between empirically equivalent translations. Such a standard would rule out the failure of transitivity of translation noted earlier; and stabilise the empirical content of our theories. Davidson owes us an account of simplicity which is neither empirical (i.e. is capable of deciding between empirically equivalent theories and translations) nor parochial but does not provide one. My suspicion is that there is no such notion to be had.

The other essential guide to good translation, according to Davidson, is charity. Charity, in translation, consists in taking foreigners' assertions as largely true. Unlike simplicity, charity does provide an objectively invariant guide to translation: if, of all the sentences foreigners are disposed to assert, one translation maps onto truths (by our lights) all those mapped onto truths by a second translation *and more* then the former is the better of the two translations. While charity thus establishes a partial ordering on translations, it has no immediate application in evaluating the comparative merits of translation schemes in which neither of the respective classes of assertions translated as truths contains those translated as truths by the other scheme. As Davidson notes, there is, in general, no merit in the idea that a translation is more charitable for translating *more* foreign assertions as truths since any minimally plausible translation rendering foreign terms as logical connectives will render countably many foreign assertions as true if any.

If the choice between such alternatives is guided solely by simplicity, then the criticisms based upon the lack of a universal standard of simplicity become even more pressing; but there is a suggestion in Davidson's works that it is more important to be charitable to some sentences than others. The examples of extending charity that he gives all concern immediately palpable situations. Charity requires translating "Es Regnet", as used by Karl, into English as "It is raining" (ceterus paribus) because Karl is disposed to assert it when it is raining in his vicinity, to translate a companion's term 'yawl' as 'ketch' if she asserts it of a passing ketch whose jigger is plainly visible to her. The impression that this choice of examples is not merely co-incidental is re-inforced by Davidson's declaration that we should extend charity "as far as possible, subject to considerations of simplicity, hunches about the effects of social conditioning, and of course our common-sense, or scientific, knowledge of explicable error."[26].

It is truistic that translations are less plausible for attributing less plausible errors to speakers. But what is not clear is how the application of this principle can get off the ground in the first place. There is, of course, rectifiable error

---

[26](Davidson, 1984b, p. 196)

— mistakes which speakers correct themselves given a review of the evidence — but those errors are so readily filtered out that they are hardly relevant to translation. What is at stake is whether or not to take your friend's adamant declaration that the repeatedly viewed boat moored just over there is a yawl, as evidence that he is using the word idiosyncratically or is somehow blind to the position of its jigger.

An experienced translator might rely upon prior experience, having translated the speech of one community and seen where its members are apt to err, she might suppose that speakers of other languages have much the same tendencies. But this relies upon the assumption that the translation of the first community imputed errors with good reason. This is not to suggest that which errors can be plausibly attributed to speakers must be determined prior to engaging in translation, but only that it cannot be determined afterwards; it must at least be part and parcel of the general project translation.

Intuitively translations that do attribute mistakes concerning obvious matters, whether it is raining, whether the jigger is fore or aft of the mast, whether seals are one's brothers etc., are suspect. One naturally supposes that the error may lie on the part of the translation rather than the speaker being translated. But it is far from clear that the application of charity to such assertions follows from Davidson's more general goals of maximising overall charity and simplicity. Indeed it is not clear if the application of charity in such cases (on which Davidson relies to give his more general appeal to charity much of its plausibility) is due to those principles at all. Reliance upon a prior knowledge of where errors may be plausibly attributed is misplaced in an account of radical translation (though the principle is still a good one for translators to follow, we are left without an explanation of how they manage to follow it) and it is not clear that Davidson has anything better to offer.

Each of the difficulties just raised is either largely resolved or avoided by Quine's account of translation. This is not to say much however, for any account of translation which, like Quine's, does not rely heavily upon either charity or simplicity will, of course, avoid the difficulties surrounding those notions. But Quine's account of translation does not merely avoid these problems, it also explains the appearance of dependence upon charity and simplicity. Just for the moment let us ignore Quine's critical dependence upon a suitably liberal standard for the inter-subjective identity of stimuli and focus instead on how his account fares in explaining these apparent dependencies.

Let us begin with a sketch of the Quinean program for developing a translation as it might be practised by a translator with unlimited resources. The first constraint upon translation is provided by the identification of the sentences which speakers of each of the languages hold true under all conditions, and,

more tentatively, the truth functional connectives. Together these two components fix an invariant framework, delineating the patterns of combinations of assent and dissent which mark the limits of proper language usage. Insofar as possible the translator is seeking to match these two frameworks, so that acceptable combinations of assent and dissent in one language are mapped onto acceptable combinations in the other. This extends beyond an injunction to take speakers as assenting to merely logical truths; the stricture is breached by a translation taking a speaker as assenting to "The vase is red" and dissenting from "The vase is coloured". Though "Either the vase is not red or it is coloured" is not a logical truth, it is held true by all competent English speakers and logically implies that if the first sentence is true so must the second. These connections establish the fundamental inferential pattern of the speakers of a language which must, so far as possible, be read into the verbal dispositions of those with whom a translation is sought.

The second constraint, equally and independently determinable from the evidence of speakers' dispositions to assent/dissent to sentences (according to Quine) is provided by the observation sentences of the two languages. These are the sentences assented/dissented to under ranges of stimuli that are pretty much *invariant* across each speech community. Stimulus meanings of these sentences provide the second constraint upon good translation; observation sentences of each language must be matched with observation sentences having pretty much the same stimulus meanings (i.e. to which speakers are disposed to assent and dissent under pretty much *the same* stimuli).

Together these two strictures fix both the inferential connections that connect sentences together into a coherent theory and the empirical content of the result. Allowing a speaker of one language, familiar with the acceptable patterns of assent of her own language, to make justifiable predictions as to which other sentences a foreign speaker will assent to upon witnessing a handful of his assertions, and hence which observations they are compatible with. Any mapping of sentences to sentences obeying these constraints is, by Quine's lights, a good translation regardless of its complexity and the charity it extends to foreigners' assertions in general.

These constraints go a long way to explaining the appearance of the extension of charity. In the first place sentences foreigners universally hold true must be translated as sentences which we likewise hold true; and secondly observation sentences, which since their stimulus meanings are the same as their English translations, will be translated as sentences that are by and large true in those conditions in which speakers are exposed to members of their stimulus meanings. Such charity is not another constraint upon translation, but simply the result of preserving the empirical content of the foreigners' theories (entailed by the two

Quinean constraints on translation). There is no sense in which charity is being optimised, no degrees of charity to be compared; judged by this residual sense of charity any translation preserving the empirical content of speakers theories is charitable and any failing to do so is uncharitable.

In practice a more substantial notion of charity, applied to more than observation sentences, will be relied upon to guide translation since translators have neither the time nor resources to follow a Baconian scheme of gathering all their evidence prior to advancing hypotheses on how to translate foreign sentences. Witnessing her subjects' happy alignment to their surroundings, and happy with the truth of her own theory, a translator will no doubt rely upon her own predictions and patterns of assertion to guide her in choosing which sentences to test for universal and unvarying assent, and which sentences to test for observationality. If she begins by translating observation sentences and then continues by presuming in general that foreigner's draw the same conclusions from their observations that she and other English speakers do, much time is apt to be saved. Such an approach will bias the result against those translations which, while compatible with the two strictures just outlined, take many foreign assertions to be false. It is this practice that gives rise to Quine's metaphor of "grafting exotic shoots on to the old familiar bush ... until only the exotic meets the eye"[27]. But as useful as such an approach may be, it is not essential. The approach can fail, if foreigners' expectations diverge sharply enough from the translators. In such cases the translator has only the two strictures and her own ingenuity to fall back upon. Even when it succeeds a translator can, in hindsight, see that there are many other empirically equivalent translations would have done as well. While charity can often expedite translation it is not essential.

Hopefully this comparison of Quine's and Davidson's account of translation has made clear Quine's dependence upon a inter-subjective standard of the individuation of stimuli which allows us to make sense of subjunctive conditionals of the form "English speaking subjects assent to ... when exposed to stimuli that dispose Junglese speakers to assent to ..." and frequently evaluate them as true. As we have seen, there is no such standard for individuating stimuli to be had.

Hopefully too, we can see that the lack of a suitable standard for identifying stimuli across individuals is the only reason for preferring Davidson's account to Quine's. That the woeful dependence upon the unhelpfully vague notions of charity and simplicity (let alone optimising the combination of both) is only tolerated because they give some account of how translations can be developed in the absence of trans-individual identical stimuli while sticking solely to available

---

[27](Quine, 1960, p. 70)

empirical evidence (dispositions to assent/dissent). If as empiricists we are
convinced that that fitting the evidence of verbal behaviour (or suitability for
causing verbal behaviour) is the only criteria of good translation, and committed
to the idea that translation must proceed by matching something about the
conditions of assent, and, finally, that there is no hope of matching stimuli
between individuals, then one must seek to match something else , and distal
stimuli — passing yachts, falling rain and the like — are the only plausible
alternative on offer.

Between them Quine and Davidson exemplify a classic pattern of over-reach
and skepticism. Quine supposes that the world is built to fit the needs of
his method of translation — that there are trans-individual identical stimuli.
Davidson, realising that there are none such to be had, rejects the entire no-
tion of empirical content, with it the distinction between content and scheme,
and announces the end of empiricism. The assumption common between them,
and upon which their dispute rests, is that translation can only proceed by
matching something about the conditions under which speakers assent/dissent
to sentences; if not the stimuli which prompt them, then more distant condi-
tions. If this assumption is false, and there is no need to identify the conditions
disposing speakers to assent/dissent to sentences across individuals, then there
is no reason not to take those conditions as proximal stimuli — patterns of
impingement at sensory surfaces or firings of sensory nerves unique to each in-
dividual — no need to rely upon charity or trade it against simplicity and, most
importantly, no need to further torture ourselves in search of the loci of suitably
invariant stimuli. In order to see that this is the case, that the need to match
stimuli is illusory, let us see how Quinean translation fares without it.

Translation without trans-individual identical stimuli

Without inter-subjectively identical stimuli Quinean methods of testing sub-
jects' dispositions to verbal behaviour under various stimuli are still sufficient to
determine the sentences to which all speakers assent under any stimulation and
the truth functional connectives. Hence the patterns of assent/dissent delimit-
ing competent usage can still be discerned. But these methods are insufficient
to distinguish observation sentences, still less to translating them. It appears
that translators have lost their entering wedge.

The problem is not that the stimuli prompting assent/dissent to any sentence
(for any particular individual) cannot be determined, but that every sentence[28]
has its stimulus meaning and it is impossible to tell, by superficial observations
of one person, whether this stimulus meaning is fixed merely by the speaker's

---

[28]Every *occasion* sentence, but we do well to ignore the Quinean distinction between stand-
ing and occasion sentences for the moment.

competence in his own language (and hence stable, and projectible into the future), or due in part to the speakers store of idiosyncratic collateral information, (and hence as susceptible to change as the beliefs that support it). If for example the subject is an English speaker and the sentence in question is something like "Red", then the stimulus meaning of the sentence is fairly fixed, and a translator can rely upon that in determining the empirical content of a speakers assertions. If, on the other hand, the sentence is more like "Bachelor", then although this sentence has its stimulus meaning (for a speaker at a time), that stimulus meaning is due, in part to the subject's beliefs about who is and is not married. Those beliefs can vary while the subject remains a competent English speaker, and hence a translator who relied upon the stimulus meaning of "Bachelor" to remain constant would be misguided as to the empirical content of a subject assenting to "John will be a bachelor again in July" (for example) since it is very unlikely that John would come to resemble any of the people whose impressions currently dispose the subject to assent to "Bachelor".

No doubt the stability of a sentence's stimulus meaning manifests itself in some hidden (de facto) physiological condition. Given a suitably sophisticated investigation of subjects' interiors, a translator might determine which stimulus meanings were stable across time and which not. While this may satisfy our requirement that the grounds of translation be empirically determinable, it is obviously not the evidence upon which translations are based. If observation sentences are to be identified it must be by other means.

The official Quinean method was to determine which sentences had constant stimulus meanings across all the speakers of the given linguistic community. That speakers would all assent to a given sentence when exposed to the same stimuli was supposed to be good evidence that the stimulus meaning did not depend upon idiosyncratic beliefs, and hence that the sentence would continue to have the same stimulus meaning for any one speaker despite any variation of his beliefs in future. While the principle is clearly sound, it is critically reliant upon identifying stimulus meanings across speakers.

Still, while Quine may have misdiagnosed the nature of these sentences, the sentences he called "observational" can still serve as the entering wedge for translation. Though they do not have subject invariant stimulus meanings, these sentences are still distinguished as argument settlers — they are the sentences to which "a scientist will tend to fall back when pressed by doubting colleagues." , "on which there is pretty sure to be firm agreement on the part of well-placed observers" [29] . While the unanimity of well-placed observers is not due to the uniformity of the stimulus meaning of the sentence in question, it remains true that the conditioning to assent to such sentences is firm enough to over-ride any

---

[29](Quine, 1960, p. 44)

differences of opinion. It is reasonable to hypothesise that they will continue
to do so, that the stimulus meaning of such a sentence for a given speaker will
continue to be stable despite any changes in his opinions[30].

A translator will thus look to those sentences whose stimulus meanings across
different speakers include stimuli that are apt to be caused by the same external
conditions. This is not back-sliding to our tedious attempts to identify stim-
ulatory loci above the skin, or otherwise attempt to work out what it would
be for two people to be in exactly the same conditions. We need not identify
conditions absolutely. If being eyes open in from of a well-lit cat disposes most
speakers to assent to "Cat" most of the time, that is enough. We need not con-
cern ourselves with whether two speakers are presented with exactly the same
conditions. Such argument settlers are the sentences upon which a translator
will rely to maintain the same stimulus meaning over time.

While this allows 'observation sentences' to be identified, if only tentatively,
it does not allow them to be translated or at least not immediately and directly.
Having identified the observation sentences of the source language, a Quinean
translator was supposed to match them with sentences of the target with the
same stimulus meanings (while respecting the constraints imposed by adherence
to the stimulus analytic sentences). Clearly no such direct approach can work in
the absence of any way of identifying the stimuli of speakers of the two languages

The saving grace is that even without translation of the 'observation sen-
tences', the empirical content of a foreigner's theory can be determined. We
can identify the stimuli compatible with the observation sentence he holds true
without translating them into English, and the fixed interconnections between
sentences are sufficient to determine which future assertions of observation sen-
tences are compatible with a speaker's current assertions. This empirical con-
tent comprises the second constraint upon good translation. Instead of seeking
a translation rendering the speaker as holding a theory which an English speaker
could hold if they expected to receive the same stimuli, we must translate the
speaker as holding a theory which implies that *he* will receive just the stimuli *he*
in his particular body expects. Rather than taking foreign speakers as propos-
ing theories which a competent English speaker could hold *in the some position*
(whatever that means), we must take foreign speakers as holding theories which
imply that *they* will receive just the stimuli that are compatible with the empir-
ical content of their theories. There is no need to match stimuli across speakers
at all.

Note that the preservation of empirical content implicitly extends charity

---

[30]There are concerns about identifying the stimuli between different temporal stages of the
same speaker, but to address them would merely add complexity without any gain in the
understanding of the principles involved.

to sentences closely tied to observation. Not merely because we relied upon uniformity of judgements to discern these sentences in the first place. That was merely a device to allow us to distinguish relatively stable dispositions to assent/dissent to sentences from those that varied with the background beliefs of the subject. If the subject's stimuli are such that they can only be caused by a passing ketch, (and he modifies his theory to fit with his observations) then we cannot take him as disposed to deny the translation of "Lo a ketch", since any theory in which that sentence is valued as false cannot be compatible with current observations. Likewise, if the subject's stimuli are such that they are incompatible (by our lights) with nearby rain, then the subject cannot be translated as assenting to "It rains". The application to present observations is only a special case, the same principle applies more generally to the subject's predictions. If the subject's predictions exclude all stimuli except those that can only be caused by nearby rain then he cannot be translated as expecting it to rain. But in the case of current observations the principle guarantees that the subject's assent/dissent cannot be taken as false. This is simply a result of the requirement that translation preserve evidence; i.e., that the same individualistic stimuli sequences should be compatible with a theory and any of its translations.

The charity involved is on the order of taking foreign speakers as able to call a spade "A spade" at close range in good lighting and does not extend to predictions of where spades will be found or what they will be used for, all of which can be grossly mistaken. Such charity does not even extend to the subject's assertions of the presence of a spade when spades are neither conspicuously present nor absent (i.e. when the subjects stimuli are compatible, by our lights with both the absence and presence of spades). Even within its bounds such charity is *not* a separate constraint upon translation, and translations can proceed, mercifully, without any disentangling of the hopelessly unclear notion of an optimal trade-off between simplicity and charity. Preserve empirical content and charity will falls where it ought.

By much the same token, the method of translation being advocated is not a kind of projection. Translators are not required to take foreign speakers as assenting to the correlates of the sentences that some competent English speaker would assent to if "projected" into the same situation. Even if we could make sense of such projection, the evidence may require the subject to be taken as making predictions that no English speaker would make, and as believing true sentences about the more subtle aspects of his immediate surroundings that no English speaker would believe. Still, there is an element of projectionism which may explain why projection has been so often taken as an essential component of translation; viz., that we must take subjects as holding as true combinations

of sentences that a competent English speaker (one assenting to all stimulus analytic sentences, dissenting from all stimulus contradictory ones, and assenting to the proper combinations of sentences containing truth functions and their component parts) could hold true if they were currently subject to the same stimuli *and had the same expectations for future stimulation.* To the extent that those future expectations can be determined independently projection ceases to be an additional guide to translation.

In practice both projection and charity will be relied upon more than is strictly warranted by the constraints imposed by the preservation of empirical content. In practice a translator will not first determine the empirical content of her subjects' theories and then seek to translate them. Such an approach would be a wanton waste of the inferential resources provided by the target language. To specify the inferential connections of a foreign tongue by means other than matching them against English sentences is a needless pedanticism that only a philosopher would dream of. In the field translators will mingle the two steps of determining empirical content and assigning translations. Hypothesising translations before subjects expectations are fully known, and then afterwards going on to test if the stimuli the translator expects subjects to find compatible with his theory are indeed held so by the subjects. In making such hypotheses the translator will no doubt rely upon her own opinion as to which stimuli the subject may be exposed to in future, such projection carries with it charity, and the translator is apt to overlook the capacity for alternative translations by sticking, as closely as the evidence will let her, to taking her subjects as assenting to those sentences she thinks they have good reason to hold true.

It is tempting, having spent so much ink and effort to reach this point to make a big deal out of this account of translation, but in a sense the modification to Quinean methods of translation is quite minor (though essential). Quine's method of identifying the empirical content of foreigners' theories survives pretty much intact, it is just that the empirical content it identifies is compatible stimulus sequences of the particular subject. What makes this work is that the world is an open world. The same objects and conditions which cause English speakers stimuli also cause the stimuli of speakers of foreign languages. Not only can we know of each other's stimuli, and of the conditions that cause them, but we can come to knowledge of the same truths, the same external conditions, each from our own perspective and limited to the stimuli which our own idiosyncratic physical frames give us access to. Fixing the stimuli that one's neighbour is exposed to goes a great way towards fixing the stimuli that one will be exposed to oneself. Worlds can be imagined in which this isn't so, in which our stimuli give each of us access to a distinct and private realm, worlds in which we all share the phenomenologist's plight with regard to everything.

In such worlds translation would be impossible, but that's alright since in such worlds communication (a fortiori linguistic communication) is impossible too.

So sound is the underlying methodology, and so obvious is the fact that translation does not proceed by matching stimuli[31], that it is Quine's addiction to matching stimuli that requires explanation rather than the success of the modified account of translation just presented.

Fortunately we need not look far to find the explanation. It is easy to see why strong empiricists required stimuli (or rather sense-data) to be identifiable across subjects, for the distinguishing claim of strong empiricism was that all factual talk could be reduced to talk of sense-data, in particular that each person's assertions could be reduced to assertions about their sense-data. (It would be no good for epistemological purposes if my assertions were reducible to your sense-data. The point of the strong empiricist program was to explain how we knew our assertions to be true, and I have no special access to your sensations.) If each of us were referring to different types of sense-data, then it would turn out that each of us were speaking a different language. If each of us only understands assertions about the sorts of sense-data that she experiences, and no-one shares access to the same kinds of sense-data, then our assertions are all incomprehensible to each other. It is, of course, open to strong empiricists to develop an alternative account of language according to which it is no bar to two people speaking the same language that the assertions of one cannot be reduced to assertions about the sorts of sense-data to which the other is liable to be exposed, but to give away this much leaves one wondering why reducibility should be insisted upon in one's own case. If a more tenuous relationship than reducibility is sufficient to establish that another person is speaking the same language as myself, it must be taken as being sufficient in my own case as well. It may still be of course that my sensations are the only evidence for the truth of such assertions, but requirement of reducibility is done away with. This is another path from strong to weak empiricism.

Davidson's claims about the end of empiricism thus merely recapitulate the end of strong empiricism. The kind of semantics which took the matching sense-data (or stimuli) between subjects as essential for translation was a product of both an empiricistic focus on evidence *and* a limited understanding of logical analysis — (we only know the world through our sensations therefore any statement about the world must be reducible to a statement about our sensations).

---

[31]Not by virtue of the impossibility of identifying stimuli across individuals in a manner that preserves both their distinguishable local properties (to which the stimulated subject has access) and their evidential import. But by the more prosaic observation that people whose senses are impaired in some way or other (such as by blindness) manage to speak the same language as their peers despite lacking sensitivity in some areas and being somewhat more sensitive to cues which others overlook.

Davidson's claims go beyond this of course. He denies the distinction between scheme and content, empirical content and contentive theory, but the plausibility of this claim relies solely upon the failure of the inter-subjective identification of stimuli.

It is both ironic and surprising that Quine, the most famous critic of reductive empiricism, should be guilty of holding to this third dogma (that sameness of meaning requires same observational capacity). But the surprise is lessened when we reflect that Quine's revolution was only ever intended to be partial. While he rejected reductionism in general, he still supposed to be true of that fragment of language constituted by observation sentences, the sentences whose "stimulus meanings may without fear of contradiction be said to do full justice to their meanings"[32]. It is hence natural for Quine to have inherited the standards of translation (identity of stimulus meanings) for those sentences from his reductionist predecessors. Where reductive empiricism went wrong, according to Quine, was in taking as true of every sentence what was only true of observation sentences. While we may fault Quine for being insufficiently thorough in his rejection of reductive empiricism, we may yet admire its cause — his adherence to the principle of least change. While in this case it led Quine to underestimate the consequences of the rejection of reductive empiricism, still the principle is by and large so worthy that its over-application in this case is understandable.

2.2.4   Each to their own

Rejection of this last toxic remnant of strong empiricism allows us to avoid the insuperable difficulties in the individuation of stimuli just explored. All along these difficulties have been due to the need to individuate stimuli in such a way that distinctly formed individuals could be exposed to the same stimuli. Without this constraint the individuation of stimuli is a relatively easy matter.

Quine, when he gave up on identifying stimuli across individuals, suggested taking classes of sensory nerves as stimuli. An individual being exposed to a given stimulus when the stimulus' members are just his firing sensory nerves. In many ways this is an admirable suggestion, but it relies upon a prior identification of sensory nerves. If, for instance, stimuli are taken as classes of the superficial nerves whose firing rate is directly affected by the external environment (e.g. by the impact of photons, or molecules upon them), and exclude the nerves implementing our senses of proprioception and balance, then distinguishable stimuli with different evidential import may be conflated. But to include among sensory nerves those whose firing rates are sensitive to such internal conditions as tension or orientation, blurs this distinction between sensory and

---

[32](Quine, 1960, p. 42)

other nerves, for the firing rate of any nerve is sensitive to various (more or less sophisticated) forms of prodding about. The more we discover about the interactions between nerve cells the less easy it is to draw a sharp distinction between sensory and non-sensory nerves. It has, for instance, recently been discovered that some retinal ganglion cells, intermediaries in the chains of nerve cells connecting photoreceptors to the brain, are not only excited by the firing of connected photoreceptors, but are sensitive to light in their own right[33]. (They are thought thought to play an important role in synchronising circadian rhythms with the rising and setting of the sun.) Such ganglia are obviously sensory cells despite not being on the absolute periphery of the nervous system. Being absolutely peripheral not a necessary condition of being a sensory nerve, and being affected by factors other than the excitation of connected nerve cells is not sufficient.

Defining "sensory nerve" by brute enumeration of the various types of sensory nerve that we currently recognise is unsatisfactory because it implies that "There are sensory nerves unknown to modern science" is false, whereas it is an obvious truth. Sometimes physiologists and other concerned scientists appear to distinguish sensory nerves as those whose *proper function* is to respond to environmental conditions. But the coherence of the notion of the proper or normal (or healthy) function of a nerve (or any other organ) is the subject of serious debate in philosophy, and we are best avoiding dependence upon it if we can.

Better to borrow a different leaf from the physiologists' dictionary and take sensory nerves as those directly affected [34] by the stimuli of their containing organism. This inverts the order of definition, defining "sensory nerve" in terms of stimuli rather than the other way around. This is, I think, the proper way to go about defining our stimuli, instead of trying to distinguish some aspects of our reactions to external impingements (i.e. the firing of sensory nerves) as (giving rise to) sensations by their local properties, we must focus instead upon the role such mechanisms play in implementing our dispositions to react to external conditions. Once this approach is accepted, the reasons for rejecting significantly subcutaneous loci of stimulation in section 2.2.1 again come into play. The difference now is that we are free to take stimuli as occurring at our bodily surfaces despite the variation in our physical forms.

Instead of looking to our bodies' immediate reactions to external impingements for stimuli, we are better off focusing upon the impingements themselves. As a first approximation we might take stimuli just as patterns of physical impingement upon our bodies. Fix the position, velocity (relative, to the subject's

---

[33](Berson, 2003)
[34]i.e. without other nervous intermediaries

centre of mass) and type of every particle striking a subjects body, the strength, orientation and type of every field passing through his surfaces, and the stimuli to which the subject is exposed are fixed. But such detailed stimuli outpace subjects' sensory sensitivity, breaching our second constraint on the individuation of stimuli — that they must be distinguishable by the subjects exposed to them. Breach of this principle risks circumscribing the range of theories compatible with a subject's stimuli too tightly, excluding some theories he might hold compatible with exposure to such stimuli, and undermining the principles upon which the more or less Quinean method of translation of the previous section relies.

Stimuli must therefore be taken to be classes of patterns of physical impingement whose members are indistinguishable to the subjects liable to be exposed to them (i.e. which no training regime will induce them to differentiate). There is here a further restriction on the inter-subjective identification of stimuli, since even subjects whose outwards forms are identical may differ in their sensitivity to their environments, and hence the subjection to some one pattern of impingement carry different evidential import for each. That these are classes of physical patterns guarantees that two stimuli instances cannot be counted as instances of different stimuli by virtue of some variation in some property which does not affect their causal powers (e.g. their etiology), that the equivalence relation which divides them into classes is indiscernibility guarantees the preservation of evidential import. For subjects like ourselves, these stimuli already serve to fix the arrangement of their bodies — fix the location of each impact between a body and the molecules of an atmosphere that surround it and you pretty much fix its surface. Fix the class of such patterns of collision which a subject can distinguish and you fix, among other things the position of his body as finely as he can discern it. Even for subjects who are not enclosed in a atmosphere but exposed, say, to an insensible form of radiation, stimuli still fix their bodily arrangement. Though subjects cannot detect the particles passing through their surfaces, such impingements still fix the location of their surfaces (provided they are sufficiently dense), and subjects' ability to distinguish between their own bodily arrangements still serves to differentiate stimuli and allocate patterns of impingement to them. If we wish to extend this definition of stimuli to cover those bizarrely rare subjects whose surfaces are not subject to a dense array of impingement, stimuli will have to be supplemented with explicit specifications of the locus of each subject's bodily surface. For the most part taking stimuli as classes of indiscernible patterns of physical impingements upon organisms bodies is adequate.

The empirical content of a theory is the sum total of the stimuli sequences compatible with it. That plausible theories will include among their empirical

contents stimuli sequences to which only some observers will access is no novelty of the suggested identification stimuli. All empiricists, excepting only the most solipsistic, accept that each of us (except perhaps god) has access to only a tiny fragment of the sum total of evidence presented to all observers. Despite this, it was supposed that these fragments were, by and large, sufficient to warrant the consensus of opinion embodied in our public scientific theories and much of common-sense. The novelty of the current approach, entailed by the rejection of trans-individual stimuli, lies in holding that the evidence available to each of us is (in practice) distinct. It is staggeringly unlikely that any two people will ever be exposed to the same stimuli (*a fortiori* the same stimuli sequence). Given my stature I will never be exposed to any stimuli to which a subject having a body proportioned like Arnold Schwarzenegger's will be exposed, but that is no reason to fail to count his stimuli as evidence; they are simply evidence to which I do not have access.

Though stimuli are thus private, in the sense that no more than one individual will be exposed to the same stimulus, translation is still possible, because each of us can know of the stimuli to which our peers are exposed. This is not sufficient, of course, overly fickle reactions to stimuli would defeat the suggested method of translation. Language use relies upon uniformity. We, following Quine, have relied upon cross-community uniformities to guide us as to which of the particular subjects dispositions to verbal behaviour will be stable through time. The uniformity underlying our speaking the same language (and the capacity to translate the speech of foreigners) does not consist of any uniformity in our sensory apparatti, nor in the stimuli or experiences to which they give rise, but is based in the alignment of our disparate sensory and nervous systems to the common publicly accessible world.

2.3 The scope of human knowledge according to weak empiricism

> When we run over libraries, persuaded of these principles, what
> havoc must we make?
> David Hume — *An enquiry concerning human understanding*

Strong empiricism stipulated sense-data as the only objects we know of. Weak empiricism is more generous, allowing that each of us knows of objects beyond our own experiences, but only through our stimuli. On this view one may know of apples as the objects which cause one to be exposed to certain stimuli when in front of one and well-lit, other stimuli when bitten into and so on. As we saw in section 1.1 however, this "and so on" covers a multitude of

sins. We need to posit apples because no recursive list can do justification to the diverse and indirect paths by which apples can register in the stimuli to which each of us is exposed. Not only is there no recursive list of stimuli sequences indicating the presence of apples (i.e. incompatible according to our theory with the absence of apples) — hence the lack of a reduction of talk of apples to talk of sense-data — but there is no isolating that fragment of our theories solely relevant to apple-spotting. The link between apples and the stimuli realised at my surface may be arbitrarily long *and* it may be mediated by any kinds of objects and terminate in any stimulus.

Ordinary languages manage to meet the demands of representing our interconnected, open-ended world, precisely because there is no reduction of talk of external objects to talk of sense-data, and because individual sentences have, in general, no particular proprietary range of confirming/infirming stimuli. Fortunately weak empiricism commits us to neither restriction, but this may leave us wondering just what weak empiricism does commit us to — what kinds of theories contravene the empiricist doctrine of justification.

Historical orientation

Empiricism has always been a critical epistemic programme, intended to to delineate the limits of human knowledge (and often accompanied with the criticism of specific claims, particularly those of philosophers, for transgressing this boundary). Hence empiricists write books with titles such as "An Essay Concerning Human Understanding", "Principles of Human Knowledge" and "Human Knowledge, its scope and limits". Most of the historical answers to this question are, however, built upon the limited understanding of logic underpinning strong empiricism. The need to quantify over unsensed objects, to refer to them by description, and to use contextual definition in reductive "analyses" were all largely unknown to empiricists up until Bertrand Russell (hence Hume and Berkeley's desperate course of treating objects as bundles of sensations).

In Russell's empiricism, the machinery of quantification calculus and contextual definition was recognised, giving rise to a distinction between objects known by acquaintance and objects known by description (in terms, eventually, of objects known by acquaintance). In Russell's hands this distinction was not used primarily as a constraint upon the acceptability of theories. Though Hegelian and Kantian idealists (and other philosophers broadly decried as practising metaphysics) claimed to know truths concerning objects that could not be known through either acquaintance or by description, and were thus open to the criticism, based upon Russell's distinction, that they were claiming knowledge that was in principle unobtainable, historically this line of attack was not much

pressed. The development and promulgation of the distinction was a feature of the linguistic turn of early twentieth century philosophy, but it was not one of the reasons for that turn. (In Russell's case the driving reasons for rejecting idealism were problems arising from idealist doctrines concerning internal relations.)

Though he did not much rely upon it in his criticism of other philosophies, the critical principle of Russell's empiricism was that "Every proposition which we can understand must be composed wholly of components with which we are acquainted"[35]. By the time he was publicly advocating this principle however[36], his battle against the metaphysicians was already as resolved as it was ever going to be, and instead of using this principle to attack metaphysical theories Russell relied upon it to defend his own empiricism — emphasising the much wider scope allowed for human knowledge under this "modern analytical empiricism" as against the strong empiricism of Locke, Berkeley and Hume. The purpose was to show that this new empiricism was unlike its forebears in being compatible with our knowing many propositions of common-sense and science to be true; though Russell put the matter in opposite fashion, describing the project as the logical construction of physical objects from sense-data. While the emphasis upon construction is in many ways laudable, being unlike analysis in not imputing uniqueness, we can only accept Russell's account of the matter by taking "construction" very liberally — in the same sense in which numbers are logical constructions, which is to say logical constructions of other objects. Logical construction, on this view, is the cobbling together of pre-existing non-logical components by logical means, better to view the object of such liberal constructions as the general term true of just the target object(s) rather than the objects themselves, for these objects are no logical fictions. This liberal notion of construction delivers the goods, giving full ontological weight to the objects the constructed general terms are taken as true of; whereas strict logical construction would show how to incorporate talk of these logical fictions without adding them to one's ontology. Strict construction would show how to eat one's cake and not have it, liberal construction shows how to describe the cake in terms of its ingredients.

The construction of mathematical, physical, and other unsensed objects was, in Russell's liberal practice, a matter of showing how to produce general terms (descriptions) whose extension included these unsensed objects by means of the logical combination of terms referring to or true of the objects with which we were directly acquainted (and classes). It was natural for Russell to emphasise the logical construction involved in this project, he was confident in both em-

---

[35](Bertrand, 1959, p. 32)
[36]The first article in which Russell proposes the distinction was published in 1911.

piricism and the capacity of the empirical sciences to produce knowledge, and thus the emphasis rested upon the new logical tools which showed how to reconcile these apparently incompatible views. Failure of the construction would, by itself, have been seen as a sign of the inadequacy of the logical tools, much as the failure of previous attempts to produce foundations for mathematics cast doubt over the tools of analysis rather than the mathematical truths they were set to analyse. Our confidence in the adequacy of Russell's logic has increased since he began his constructions. The stock of quantificational calculus has been increased by a confluence of evidence. Church's lambda calculus, Turing's machines, the theory of recursive functions, all conspire together to assure us that the boundaries of quantificational calculus mark a significant point in the complexity of intellectual and computational systems which no physical implementation has been seen to surpass. If the logic tools it supplies are inadequate to rendering weak empiricism compatible with some of our knowledge claims then there is little hope of a superior alternative to save the situation.

Mathematics was the target of Russell's constructions because it is the clearest example of a body of apparently necessary truths and knowledge of necessary truths seems incompatible with the empiricist doctrine of justification and Russell's principle. Showing these truths to be purely logical would have immediately reconciled the existence of this body of knowledge with Russell's principle, for no logical terms referring to objects or properties beyond one's acquaintance would have been required for the enunciation of mathematical truths; and provided logical combination does not introduce new sources of evidence (so long as logical terminology can be learnt from empirical evidence) it would have also reconciled mathematical knowledge with the empiricist doctrine of justification. But such reconciliation was not restricted to mathematics, that was to be only the beginning. Other fields which claim to produce knowledge of necessary truths were to be reconciled in like manner — by showing the truths they proclaimed to be logical truths, such was the programme for analytic philosophy.

Alas such reconciliation was not to be. In settling for the liberal kind of logical construction, admitting objects with full weight and not living up to the strict standards he proclaimed for logical construction, Russell failed to reduce mathematics to logic. One way of coping with this was to widen the scope of analytic truths, to include not only logical truths but also statements true by definition. On this view the axioms of set theory were to be construed as definitions, and the truths of mathematics thus true by virtue of definition and logic. But the world still has to supply objects of which these definitions are true. Pure mathematics perhaps could be construed as a vast hypothetical, "if there are countably many sets then ..."[37] but such an approach is both

---

[37]Indeed one does occasionally come across mathematicians who proclaim this view, claim-

misplaced and unhelpful when it comes to dealing with the more practical uses of mathematics as exemplified in the more rigorous sciences. These concerns were known at the time, centring on anxiety over the axiom of infinity, and leading some to propose the existence of logical objects.

In the event, these shortcomings were overshadowed by the discovery of Gödel's theorem which states that no recursively axiomatisable formal system can capture all the truths of mathematics as theorems. Thus no matter how liberal the assessment of logical truths, mathematics could not be reduced to logic without raising the same problem of empirical access, but this time with regard to the logical truths themselves.

Russell's reduction of mathematics thus failed, at least directly, to provide an acceptable explanation of our knowledge of necessary truths. But though his attempted reduction of numbers failed to eliminate them (to show them as logical fictions), it succeeded, in a back-handed manner, in showing that numbers were indispensable The very indispensability of references to numbers in scientific theories capturing sophisticated empirical contents showed the program to have been misconceived. The need to posit numbers in our scientific theories establishes their empirical bona fides.

This approach to mathematical objects requires giving up on the necessity of mathematical truths. The postulation of objects is part and parcel of our general postulation of objects in the theories required to support our empirical predictions. If the world were so simple that adequate empirical theories could be stated without use of mathematical objects there would be no such objects. Strong empiricists were mislead by the absence of any reduction for mathematical statements, the lack of any infirming stimuli for individual statements, into supposing that these sentences must be known in some other way entirely divorced from empirical observation. But weak empiricism recognises that this requirement is too strong, that in a complex world the relations of objects to observations is more indirect and diffuse.

While such an approach may serve as the basis for an explanation of our knowledge of mathematical truths, it leaves us wondering about other, non-mathematical, putatively necessary truths. In philosophy particularly there are disputes over statements between parties sharing the same empirical expectations (insofar as they are relevant to the dispute at any rate). The indispensability of numbers, which assured them of their place in our empirical theories, seems not to extend to these philosophical claims. In particular it does not seem to extend to the empiricist doctrine of justification itself. This has given rise to the charge that empiricism, or positivism as it is has more disparagingly

---

ing, in their philosophical moments, that mathematics is not true, but what would be true if.

been called, is pragmatically self-refuting — that "positivism produced a conception of rationality so narrow as to exclude the very activity of producing that conception"[38].

It was never as obvious as the critics of strong empiricism made out, but this charge was true of strong empiricism in its various forms. The proponents of the verificationist principle might insist that "The meaning of a sentence is its method of verification" should be taken as true by definition or analytic, and hence not meaningless by its own lights, but at some point it had to be connected with the vocable blasts and inscriptions by which humans make assertions. Not only do we need empirical evidence that the world provides objects of which these definitions are true, but the notion of a definition, and of a definitional truth need to be spelt out in verifiable terms at some stage; the appeal to definitions only served to defer the problem. The positivist/verificationist programme suffered both from difficulties in isolating a sense-data or protocol language[39] and from a lack of successful reductions to it, but with the obvious appeal of empiricism and difficulties in appreciating the logical restrictions imposed by strong empiricism, the flame was kept alive until it was snuffed out by Quine in *Two Dogmas* and *Word and Object* where he pointed out that there was no empirical distinction between definitional truths and statements expressing true beliefs held universally and tenaciously (by the members of the relevant speech community) [40] such as the sentence "There have been black dogs" (held true by competent English speakers). Quine's criticism can thus be viewed as a vindication of the traditional charge that positivism is pragmatically self-refuting.

Quine didn't put the matter that way. His charge was that reductive empiricism was incompatible with (weak) empiricism, not only that sentences of the form "...is true by definition" are not verifiable by themselves, but that the general term "true by definition" cannot be paraphrased by any logical construction of empirically accessible ingredients; putting the matter in Russellian terms, the property of being true by description cannot be known by acquaintance or description. This naturally leaves us with the question of whether weak empiricism is pragmatically self-refuting in the same way that strong empiricism was claimed (correctly) to be self-refuting by its critics; and indeed the critics of empiricism have often taken the success of their criticism of strong empiricism to show that empiricism in general is self-refuting without properly considering the differences between its strong and weak forms.

One way to attempt to exonerate weak empiricism from this charge would

---

[38](Putnam, 1983d)

[39]Largely for the reasons discussed in §1.0.3

[40]Except of course for the intuition of analyticity, and it is of course up for grabs just what that is an intuition of.

be to to specify a class of terms referring to the directly sensible objects and the predicates true of them (as Russell somewhat misleadingly puts it, these terms would be names for objects known directly), and then specify the logical apparatus by which general terms might be composed from these fundamental elements and then to produce a translation of the empiricist doctrine of justification in this austere language. Even if completed, such a task would not show weak empiricism to be true; in a sense it would not even show it not to be self-refuting. Not only might there be other sentences known to be true which could not be rendered into this austere idiom, but it might be alleged that the standard of translation was too liberal, by virtue of leaving some extra-empirical component of meaning out of account, and that the austere version of the empiricist doctrine of justification and its English counterpart did not have the same meanings. Provided that the standard of translation involved met the empirical constraints upon good translation however, the empiricist principle of justification would be shown not to be self-refuting by its own lights and that is all the vindication required.

Such a construction of the empiricist doctrine on content would be a massive and massively complicated undertaking. If the terms "believes", and "experience" can be defined in the austere idiom of logic and terms referring to objects known directly, it is as the theoretical terms of a complex psychological theory. The obstacles to producing an austere version of this theory are overwhelming. In the first place it is not even clear what the truths of psychology that we would seek to capture as the theorems of such a theory are. Not only do we lack an austere version of a psychological theory adequate to the prediction of behaviour, we lack any explicit version at all. We have nowhere near the same clarity in understanding which sentences we would wish to capture as the theorems of our psychological theory as those who sought to reduce mathematics had of the truths which they wished to capture in their reductions.

An explicit formal psychological theory would only be the beginning. Such a theory would, presumably, serve to generate definitions of the theoretical terms of psychology in the terms of non-psychological terms — the terms which psychologists come to understand as part of their specially psychological training in terms of those they must understand prior to that training if they are to get on with it. These prior terms would not be, in Russell's terminology, names of objects known by acquaintance. The connections between these terms and the objects known by direct acquaintance are so multifarious and inter-twined that a theory capable of producing definitions of these in terms of objects known by acquaintance would have to be a theory of pretty much everything. For consider, psychological subjects on distant planets might communicate with us by manipulating their planet's geography, or its weather, or the spectrum of radi-

ation emitted by their star, nearer ones can make inscriptions, vocable blasts, statuary, smoke signals — the list of the manifestations by which psychological states might be revealed is, like the states and processes of nearly any objects, open-ended and insusceptible of separate enumeration. And these are just the difficulties surrounding the psychological aspects of the empiricist principle of justification, let alone those arising from the notion of justification.

The direct vindication of the empiricist principle of justification is thus a mammoth task. The production of a translation of the empirical principle of justification in the austere idiom of directly observable objects and properties is not a task so much as a research project, not to be carried out in one massive burst of genius but in a more piecemeal fashion. One step would be to translate attributions of propositional attitudes into non-intentional language, at the other end works like Carnap's Aufbau der Welt shows (in combination with Lewis' account of the definition of theoretical terms) how to translate attributions of colours to points in three-dimensional space into a language in which the only objects named are points in two-dimensional sensory space. This translation is not an elimination of the references to points in physical space in favour of those of the sensory space, it is not a logical construction in the strictest sense and does not show how to dispense with references to these extra-sensory objects altogether, but it shows how to construct descriptions of them from sensory components; how to eliminate names for these objects without eliminating reference to them. The points of three dimensional space and their colours are the objects and properties which come nearest and near enough to fulfilling Carnap's desiderata for attributing colours to them.

While fragments of this project may be successfully completed by philosophers in relatively short periods of time, there is no expectation that the whole will be completed in the near future, and our current inability to produce the required translation of the empirical principle of justification is thus no knockdown objection to the overall project. On the other hand one is justified in being suspicious of this begging off of direct vindication. Much the same excuse of extreme complexity was made by the strong empiricists to explain their inability to produce their translations, but it turned out that their principles — the empiricist principle of content and the verification principle of meaning — really were self-refuting. Not only were these principles unstatable in any language of which they were true but, as we have seen, so too were a host of other statements of ordinary English, and the appeal to complexity merely served to conceal strong empiricism's fundamental problems.

Of course, one way of vindicating the program to which weak empiricism gives rise is to carry it out, and equally of course this method of vindication is not available to us at the early and middle stages of this program. The

contention of the critics of weak empiricism (or rather of empiricism generally, for often no distinction is made) is that we can already see that weak empiricism is false and that there is little or no point in devoting further resources to its execution. Importantly, the critics' charge is not that merely that the truth of the principle of justification is incompatible with our knowing it to be true[41], but that this principle is just one of a whole class of claims, the philosophical claims, which the empiricist principle of justification implies cannot be known.

Like statements of mathematics, philosophical claims were supposed to have their truth values necessarily and necessity is, if our empiricism is correct, inscrutable; but this is not grounds for holding the empiricist principle of justification to be self-refuting. The current charge is that the empiricist principle of justification implies that philosophical claims, including itself, cannot be known at all, and not merely that they cannot be known to be necessary. It is quite compatible with our empiricism for the empiricist principle of justification to be only a contingent truth (or indeed for the necessary/contingent distinction to be simply ill-conceived) so long as it is a truth that can be known. The problem is that the philosophical truths, like those of mathematics, were supposed to be discerned by pure reason and not by empirical investigation, and unlike the truths of mathematics, this claim appears to be true. Mathematics was redeemable because it is indispensable in generating the empirical predictions of our scientific theories, but philosophical claims, however, do not appear to share this virtue. Disputes over philosophical claims generally occur between parties who do not differ in their empirical expectations in any way that either thinks is relevant to the dispute; philosophical claims are not settled by experiment.

Since philosophical claims appear to be unlike mathematical statements in not being indispensable to our empirical theories, one might naturally hope that they are sufficiently unlike to be redeemable by strict logical construction — that philosophical truths might be defended as logical truths. But it is obvious that most philosophical claims, the ones over which there is dispute at any rate and

---

[41]I have, in this section, played loose with the distinction between the epistemic limits and the semantic limits implied by empiricism The empiricist principle of content and the verification principle are both explicitly semantic, they demarcate the limits of what can be meaningfully said. The empiricist principle of justification is an epistemic doctrine, demarcating what we can know. It is very plausible that this principle can be shown to imply, via arguments of much the same kind as we employed against the possibility of a shared language of phenomenology, or by the standard of translation it gives rise to, that we can only speak of the types of objects of which we can have knowledge; but that argument has not been advanced here. Though blurring the difference between the two limits has served both to exhibit the similarities between criticism of strong and weak empiricism, and between Russell's and our view of the motive for logical construction (for Russell's principle is semantic, and explains his view of logical construction as a form of semantic analysis) the only principle being defended is the empiricist principle of justification. We are concerned here not with the objection that the principle implies its own meaninglessness, but only that it implies that it cannot be known. This is still enough to show that the empiricist program we have envisaged is misconceived.

the empiricist principle of justification in particular, are not logically valid —
not true under every re-interpretation of their non-logical terms. In order then
to defend these statements' status as logical truths, philosophers who have taken
this line, including nearly all of the analytic philosophers in the early twentieth
century, postulated a 'deep grammar' or hidden logical structure, to be revealed
by logical analysis. Thus the title "*analytic philosophy*". The claim is thus
that the analysis of philosophical truths, including the empiricist principle of
justification, produces logically valid statements.

This move, the postulation of 'deep grammatical structure' and a technique
of logical analysis for its revelation, may suffer from much the same flaw that
affected the verificationists' appeal to definition in defending their verificationist
principle. Just as the verificationists were unable to spell out their notion of
definition in terms that were empirically accessible by their own standards, so
too weak empiricists may be unable to spell out the notion of logical analysis in
terms that are epistemically accessible by their standards. If the only constraint
upon logical analysis is that the analysandum and analysans be translations of
each other, and the only constraint upon translation is preservation of evidence
(i.e. dispositions to verbal behaviour prompted by stimuli), then any sentence
which competent speakers maintain to be true in the face of any experience will
be a logical truth under some analysis or other[42]. Analytic philosophy is thus
reduced to a near pointless collection of arguments over which of a variety of
good translations is preferable. The question of whether this is so, and whether
in general our knowledge of logical truths is compatible with weak empiricism,
shall occupy us in the next chapter.

For the moment let us get clear on the charge we propose to answer; it is
that (weak) empiricism leaves scope only for a limited kind of scientific thinking.
Popperian falsification remains intact: if a theory implies that all swans are
white and one sees a black swan one may conclude that that theory is false.
It may also leave scope for the computation involved in using a theory (if our
knowledge of logical truths and entailment is compatible with empiricism). It
may take time to perform the computations involved in deriving the predictions
implied by the combination of one's theory and some novel observation. But
there is, according to the critic, no scope for *reasoning*, no scope for sitting
down and *thinking* about a matter and coming up with a good reason to change
one's mind in the absence of any unexpected observations. Thus no reason for
preferring one or another side in a dispute between parties which differ none
in their empirical predictions, no reason to assent to "All justified beliefs are
justified by experience" rather than its negation and thus no justification for

---

[42]Though not any class of statements. Hence the immunity of mathematics as a whole to
such treatment

believing the empiricist principle of justification. Our goal is to show that this charge is false.

Showing that weak empiricism leaves scope for the justified resolution of philosophical disputes would not, of itself, show that the empiricist principle of justification is compatible with our knowledge of it. But it would leave no reason to suppose it incompatible, it would be a vindication of our empiricism of the same strength as Moore's refutation of idealism.

Our goal is thus to show that knowledge of philosophical truths (or rather of the correct resolution of philosophical disputes) is compatible with weak empiricism. In fact, I propose to aim slightly lower than attempting to provide a compatible positive justification of such knowledge. Instead I propose to give an account of philosophical defects — what can be wrong with a theory despite its fitting all the empirical evidence. This is sufficient to provide an explanation of how philosophical disputes can be justifiably resolved that is compatible with weak empiricism. It gives rise to a plausible account of positive justification — if we can have good reason to believe that every theory which makes no false empirical predictions, and has at least the predictive power of our current science (i.e. is compatible with a narrower range of stimuli sequences) and which includes the English sentence "Some justified beliefs are not justified by experience" is defective then we have good reason to believe that sentence to be false — while allowing us to avoid the general issue of how belief in scientific theories is justified. If one holds a quasi-Popperian view according to which belief in our current scientific theories is justified because of the repeated failure of attempts to falsify them, then if the addition of the English sentence just quoted produces a defective theory one has reason to believe the empiricist principle of justification. If one holds instead a view according to which it is not beliefs, per se, but changes in belief that require justification, then such defects will give a good reason for changing from a defective view (in which the empiricist principle of content is denied) to one which is not defective (in that respect at least). While I will touch upon these issues concerning justification in defending weak empiricism against other charges the main focus remains on the question of whether weak empiricism implies such a limited scope of human reasoning that it precludes the justified resolution of philosophical disputes, and hence precludes the justification of a belief in weak empiricism itself.

While investigation of this question might proceed without the clear delineation of the limits upon theories that can be known according to weak empiricism, there are still good reasons for clarifying those limits. Such clarification serves to make the commitments of weak empiricism more concrete, making clear the target of our defence and reassuring us that the defence of weak empiricism does not turn upon an unacceptably liberal interpretation of the position such as

sets unempirically open limits to human knowledge. In fact, the limits implied by the form of weak empiricism I espouse (the techniques by which descriptions may be logically constructed from fundamental terms) are often criticised as being too restrictive. Defence against those criticisms is beyond the scope of this work; our focus remains on the charge that weak empiricism precludes the practice of philosophy, and in particular the justification of belief in weak empiricism itself. If we omit the exploration of these criticisms and the rejoinders to them then we find that in delineating our stimuli the task is more than half completed already, for the techniques of logical construction which we may practice upon these foundations are well known and easily explained. So let us return to the point at which we began this section and delineate the restrictions empiricism places on justifiably believable theories.

2.3.1   Foundations and Constructions

Though our techniques of, and motives for, logical construction overlap to a great extent with Russell's, our attitudes towards the fundamental materials upon which these constructions are based and the features distinguishing them are vastly different. These differences permeate the undertaking and are reflected in different attitudes towards the final constructions. Russell distinguished objects known by direct acquaintance on modal grounds; they were the objects whose presence was registered infallibly. Russell thought these objects included universals and that the relationship between them and the terms reporting acquaintance was naming. Weak empiricism differs on all these points.

Our fundamental elements are stimuli, chosen not because of the certainty of our acquaintance with them, but because our current best science tells us that all the belief justifying influences of the world upon us are mediated by events at our sensory surfaces. The difference between this view and Russell's is the difference between seeking to explain our knowledge of the world in scientific terms and seeking a first philosophy specifying the foundations upon which science might be built. It is striking that Russell's epistemology exemplified a form of weak empiricism (for he does posit irreducible unsensed sensibilia) which postulated certain foundations for scientific knowledge. Quine invariably treated the two matters as indissolubly linked. On Quine's view, as soon one accepted that references to non-sensory objects are ineliminable (the attempt to find "in some subtle construction of sets upon sets of sense impressions a category of objects enjoying just the formula properties of objects" given up as fruitless), one no longer had reason to seek to ground empirical science upon indubitable foundations, no reason not to turn to psychology for an account of how our observations are related to our beliefs.

Despite differences in attitude towards these foundations, they do still serve as foundations. The two key components of weak empiricism are shared: there is a certain class of beliefs concerning one's own experiences for which experience (exposure to stimuli) provides a direct and immediate justification, and all we know of the more distant world is the sequence of these directly known objects which it presents to us and (granting that something like induction is knowledge supporting) that it is disposed to present certain sequences of these objects to us in future. The differences concern the justification for these principles and the confidence in the foundational beliefs they hold warranted by the corresponding experiences. For Russell direct acquaintance with an object gave an absolute warrant for believing that one was so acquainted, for us too exposure to a stimulus justifies the belief that one is so exposed, but the confidence warranted falls short of certainty. On Russell's view this certainty was the distinguishing feature of acquaintance, our focus upon stimuli is justified by our scientific theories; Russell's account would give the epistemic foundations of science, ours the epistemic foundations according to science.

Though we take our foundations to have less than the absolute solidity Russell accorded to them, still the techniques for building on them are the same, as are the epistemic limits they entail. Let us approach these limits, and the techniques for achieving them, a step at a time.

Obviously some theories composed in a language composed solely of reports of concurrent exposure to stimuli, are knowable according to our empiricism. The sentences of this language might be translated as "Lo stimuli S". Such renderings owe much to the language into which they are translated — they impute names for stimuli where objectively there are merely distinct responses to them. Such humble beginnings carry no ontological commitment to stimuli, indeed no ontological commitments at all. At this primitive stage competence in the use of the 'language' and epistemic access to the truth value of its statements are not to be distinguished; both consist simply in the ability to make distinct responses to distinct stimuli. In taking these responses as statements referring to stimuli and subjects, we are implicitly assuming that this austere vocabulary is a mere fragment of a larger whole, in which the apparatus of quantification is clearly at work. The very question of whether stimuli are named within such puerile "theories", whether the statements of such a theory refer to either stimuli, stimuli instances, and/or the stimulated subject is simply a matter of which structure we wish to impose on the its translation into a more fully rounded language

Though we thus take such responses as reports of passing stimuli only with the benefit of an external framework supplied by some other language (or imposed in hindsight once this rudimentary language has been incorporated into a

richer whole) it is clear that the truth of the assertions made in such a language are within their proponents' epistemic reach (so long as each sticks to reporting his own stimuli). Competence in the making of assertions in this language might be inculcated by stimulus-response training, and it was by the possibility of such training that we distinguished stimuli in the first place. We might even go further and inculcate the appropriate dispositions to assent and dissent to these sentences as they are or are not accompanied by the requisite stimuli, but such promptings would disrupt the stimuli to which the subjects were supposed to be responding. This provides one powerful reason why language learning begins with references to more removed objects.

Even within the realm of sensory experience we quickly come to matters for which such amorphous sentential structure is inadequate. Not only is knowledge of passing stimuli obviously compatible with empiricism (both strong and weak) but so too is knowledge of the stimuli sequences one has been exposed to. There are a countably infinite number of stimuli sequences of finite duration, too many for a sentence reporting each to be learnt, and the machinery of quantification begins to play a useful role. There are many ways in which the requisite reports of stimuli sequences might be constructed. One language adequate to making just such reports would have a two place primitive predicate expressing temporal precedence, and for each stimulus (to which the speaker is liable to be exposed) a predicate true of just its instances. If we followed the practice suggested in section 2.2 we would also require a two-place predicate interpreted as "lasted for longer than" though, if our experiences were suitably discrete, we could equally take each stimuli instance as lasting for the minimum distinguishable period of time. The difference is between taking one unchanging pattern of surface impingement as one enduring stimulus instance or a multitude of succeeding instances of the same stimulus, both approaches are acceptable. The important thing is that the language be rich enough to express the temporal order of stimuli instances and at least the relative duration of each. This might give rise to concern that our sensory language is already too rich, for it will include a translation of "every stimulus instance is not before itself" a statement which would be false if time were circular rather than linear. Thus far we our language has been restricted to talk of objects and properties with which we are directly acquainted, and it is unreasonable to include the linearity of time among such directly known matters. But the difficulty may be avoided, for no matter what geometry time may have on the large scale each of us experiences within his own life a definite order of his experiences, better then to take our temporal ordering relationship as "temporally precedes in my lifetime", true of any pair of stimulus instances of the same subject of which the first is temporally closer to his birth than the second.

This almost completes our account of the objects and properties that may be known by acquaintance and the primitive unstructured terms referring to them. From this basis references to horde upon horde of other objects are to be introduced as descriptions. Though Russell was the first to clearly elucidate contextual definitions for definite descriptions the descriptions we are imagining are not tiddly examples like Russell's analysis of "the author of Waverley". Rather we look to whole theories to provide the complex descriptions of the objects they introduce terms for, and do so on a wholesale basis. Each theory giving rise to a horde of definitions, one for each of the terms it introduces. David Lewis has given a very clear exposition of how such definitions may be constructed[43]. He gives as an example a detective proposing a theory of a crime with the aid of three new theoretical terms "X" "Y" and "Z". The detectives theory is that "X, Y and Z conspired to murder Mr. Body. Seventeen years ago, in the gold fields of Uganda, X was Body's partner ... Last week, Y and Z conferred in a bar in Reading" and so on, more and more details concerning X, Y and Z are given as part of the theory. Lewis' claim is that as a result of being told the detectives theory, and understanding the other terms used in it, the audience of the detectives exposition can now use the terms "X", "Y" and "Z" to refer to the members of the triplet coming nearest and near enough to realising the detectives theory. These terms get their meaning from the theory used to introduce them and may be defined as referring to the corresponding members of the tuple uniquely realising the theory. The claim of weak empiricism is that all the terms of any sentence which we can be justified in believing can be introduced in this way in a theory in which the only non-logical terms are those reporting (potential) direct acquaintances — the terms specified in the previous paragraph.

There are however several immediate problems with this claim. Firstly, Lewis' account of the definition of theoretical terms only serves to define *names*. We would thus be unable to introduce predicates such as "is blue", and a complete lack of such predicates precludes even the introduction of descriptions of objects other than those already known by acquaintance. If we were restricted to the austere idiom outlined we would never be able to refer to any objects but instances of our stimuli. In addition to the terms we have already recognised as giving rise to justifiable claims, we must admit just one more — "is a member of". This will allow us to introduce names for the extensions of theoretical predicates, and these names combined with "is a member of" can serve to state theories empirically equivalent to those in which the theoretical predicates feature. For example, instead of the theory in which the predicates "bacterium", "cell", "culture", are used we can turn instead to the equivalent which features

---

[43](Lewis, 1970), and (Lewis, 1972)

names of the corresponding extensions "class of bacteria", "class of cells" etc., or to make it clear that these introduced terms need have no explicit internal structure "bacterium-kind", "cell-kind" etc. A fragment of the theory which gives these terms their meaning is "Members of bacterium-kind are very small. Each member of bacterium-kind is a member of cell-kind. A multitude of members of bacterium-kind of the same species growing together comprise a member of culture-kind ...". Clearly this formulation is empirically equivalent to the less verbose original which featured predicates rather than the corresponding phrases of the form "member of X-kind".

Admittance of the term "is a member of" to our austere idiom seems to run contrary to our empiricist principles. For in one fell swoop we acknowledge all the truths of mathematics to be within our epistemic scope. This seems at odds with our justification of our knowledge of mathematical objects on the grounds that they are indispensable in our scientific theories. We would have expected numbers to be introduced as the values of physical quantities, for the number 3×10^9 for example, to be introduced as the number related to photons as their speed in meters. Our concerns on this point can be somewhat assuaged however. We are seeking to demarcate the sentences which are within our epistemic reach, not those we actually know, and the denotations we can use not those which are guaranteed to have referents. In a world so simple that it gave rise to observational patterns so simple that they did not require numbers for their prediction we would have no reason to suppose that there were numbers, even though we could in principle produce descriptions that would be true of them. The standard of empirical accessibility being proposed does not pre-empt our justification for believing in numbers, but leaves room for it. In the other direction, the admittance of the term "is a member of" to our austere idioms reflects subject's capacities to generalise over the objects they postulate. Without this capacity the postulation of objects is rendered moot, for postulation only has its point when an infinite number of objects are postulated, talk of any finite number of objects being capturable within purely truth-functional logic which is ontologically neutral. The connections between such an infinite range of objects and our observations can only be captured in a recursively axiomatisable theory by means of predicates which these objects satisfy. A horde of objects over which no true generalisations could be made would give rise to patterns of observations which no such theory could predict. Generalisation over objects is a non-negotiable component of their postulation. Possessed in the most general degree this capacity for generalisation gives rise to the capacity to comprehend the statements of mathematics. Of course individuals may not possess such capacity, they may be able to recognise only a finite number of kinds of objects, perhaps natural kinds, but we cannot

rely upon such limitations in demarcating the range of theories which subjects may be justified in believing. It is incumbent upon us not to place arbitrary or ad hoc restrictions upon subjects' capacities to make generalisations, for we wish to trace epistemic limits implied by our empiricism not by psychology.

A second problem is that on Lewis' account theoretical terms refer to *the* objects coming nearest and near enough to realising the theory that introduces them. If the theory has multiple realisations (or no exact realisation and many equally near realisations) then the terms fail to denote any of these objects. Whereas, as both proponents and critics of weak empiricism have recognised, if our scientific theories are realised by some objects at all then they will also be realised by an infinite number of suitable permutations of them. On this point we must follow Carnap's account of the introduction of theoretical terms rather than Lewis', taking such terms as naming the components of some arbitrarily chosen realisation of the relevant theory, though the choice as to which is arbitrary. We may make some concession to the ordinary intuition that such arbitrary decisions are misplaced in determining the reference of the names introduced in our scientific theories by insisting that such indeterminacy only be tolerated at the most basic level, from which, if weak empiricism is true, the indeterminacy is ineliminable[44]. We can insist that any term that can be introduced in a theory stated entirely in terms of prior, or old theoretical terms, only denotes an object if that theory is uniquely realised relative to a fixed interpretation of those terms. This secures the noted indeterminacy at a level beneath that of ordinary language, whose terms are exclusively theoretical terms (relative to the austere idiom of observational terms we have delineated), reconciling our ordinary intuition that multiple realisation is a defect in a theory with indeterminacy of translation and reference implied by our empiricism.

The third objection is that translation into the austere idioms we have envisaged mis-represents ontological commitments. For the result of such translation will include the extensions of the various predicates occurring solely in the original among its ontological commitments. These sets must be reckoned as among the values over which the quantified variables of the translation range if the resulting theory is to be true, whereas the source of the translation may carry no such ontological commitment. The fragment of the semi-translation given above, for example, imposes the postulation of an object named by "bacterium-kind" where the original fragment had only the predicate "bacterium" which does not name anything.

One way out of this problem might be to produce a method for defining theoretical predicates analogous to Lewis' account of the definition of theoretical names. Rather than taking as a definition sentence for the new term a sentence

---

[44]See (Winnie, 1967) John Winnie the implicit definition of theoretical terms for a proof

identifying the object referred to by a name with that referred to by a description couched in the old terms understood prior to the introduction of the novel theory, we might take a sentence asserting that the novel predicate is true of just those objects satisfying an open sentence couched in the previously understood terms. But in fact such a defence would be misplaced. While it is interesting to consider whether or not such a device as just suggested could be carried through, it is completely irrelevant to the defence of the empiricist demarcation of knowable theories under consideration.

Our earlier statement of the limits of human knowledge implied by weak empiricism was that all the non-logical terms of any sentence which we might be justified in believing could be introduced in the manner of theoretical terms into a theory in which the only non-logical terms were predicates expressing exposure to the instances of a stimulus, terms expressing the immediately discernible temporal relations of stimuli, and membership. This principle applies to each individual, each sentence an individual can be justified in believing must be translatable into a language in which the only predicates expressing exposure to stimuli are just those expressing exposure to stimuli to which *he* is liable to be exposed. In putting things this way we have followed Lewis' device of explaining how such terms may be introduced without pause for an explicit definition. But of course the method Lewis proposed was a full-fledged method for giving the definition of novel terms in old terms, for eliminating the novel in favour of the old while preserving meaning. Lewis's account of the preservation of meaning was (like almost everything else everything else in Lewis' philosophy) unashamedly modal. He justified the claim that the new definienda and the descriptions he offered as their definiens were alike in meaning on the grounds that they denoted the same object as each other in every possible world. By contrast, the only standard for good translation that we have proposed is preservation of empirical content, but any mapping that is a translation in Lewis' sense is also a translation in ours, the Ramsey sentence of a theory (formed by replacing all novel named with existentially quantified variables) implies all the same O-sentences, and is hence compatible with just the same range of stimuli sequences as the theory itself, and the meaning postulates or the definitions they give rise to imply no O-sentences, no truths about stimuli, except logical truths.

The claim is thus merely that all the everyday statements any particular person can be justified in believing can be translated into the austere idiom in which the predicates correspond to the stimuli to which that individual may be exposed. Translation in this sense in not guaranteed to preserve ontological commitments, indeed it is quite patent that ontological commitments can be endlessly and pointlessly enlarged upon in such translations[45]. The quantifi-

---

[45]That the positing of universals is just such a pointless imputation is the heart of Wittgen-

cational standard of ontological commitment I have advocated applies in the
first place to the untranslated statement of subjects' theories, provided they
are already in quantificational form, and in that form if the subject has no
call to quantify across the extensions of his predicates those sets are not to be
rated among his ontological commitments. If they are translated, either be-
cause they were not in quantificational form already, or for some other reason
(as at present), then the ontological commitments imputed are relative to that
translation. Indeed if ontological commitments were preserved by translation
there would be no call for such relativisation. The objection that translation
of a subject's tongue into an austere idiom as sketched may mis-represent that
subject's ontological commitments is thus irrelevant since it is no part of our
view that translation preserves such commitments.

This attitude towards the relation between the original version and its aus-
tere translation is a further difference between the current view and Russell's.
For Russell, the objects known by direct acquaintance were meanings; and trans-
lation into a tongue in which all terms referred to such acquaintances a unique
analysis revealed the true meaning (and hence the true ontological commit-
ments) of the subject's assertions. This allowed him to parley the absolute
certainty of such acquaintances into a more general notion of analyticity or
truth by virtue of meaning alone.

This difference of attitude is also reflected in divergent responses to the
objection that the limits on human knowledge implied by our empiricism also
imply that no two people can mean the same thing by their assertions. Russell
embraced this conclusion "It would be absolutely fatal if people meant the same
things by their words. It would make all intercourse impossible and language the
most hopeless and useless thing imaginable, because the meaning you attach to
your words must depend upon the nature of the objects you are acquainted with,
and since different people are acquainted with different objects, they would not
be able to talk to each other unless they attached different meanings to their
words"[46]. This is a result of his view that the translation into terms referring
to objects known by acquaintance is analysis — revealing the one true meaning
of the statements analysed. No such claims are attached to the view currently
being urged; these austere translations are not put forward as (canonical for-
mulations of) sentential meanings. All that is guaranteed of translation on our
view is the preservation of fit with dispositions to verbal behaviour. Translation
into an austere idiom has no special privileges beyond this, serving neither to
identify the meanings of a speakers tongue nor the ontological commitments he
undertakes within it, but as the hallmark of statements whose truth values are

_____

stein's objection to the coherence of the philosophical arguments over Universals and tropes
    [46](Russell, 1972)

not epistemically inaccessible. Sameness of meaning between speakers, in the only sense remaining to us, is simply a matter of the success of homophonic translation.

2.3.2   The status of logic

We have now given a statement of the limits of human knowledge according to weak empiricism; viz., that every statement an individual may know must be translatable into a language in which the only non-logical terms are n predicates corresponding to the stimulus types to which that individual is liable to be exposed, each true of just the instances of the corresponding stimulus. This limit has not been so much argued for as simply asserted. We have, in particular, not elucidated the relation of justification. The only standard for justification easily read off from our account of the connection between scientific theories and observational evidence is the purely negative standard of falsification — exposure to a stimulus sequence to which a theory implies one will not be exposed provides, somehow or other, a justification for believing that that theory is false. That there is more to be said on this point, let alone the nature of our warrant for positive confidence in some one among the empirical theories that are compatible with our observations is painfully obvious, though to what extent we should turn to professional epistemologists or to the scientists who know their theories to say it is less clear. My intent is to stick to the narrow question of whether weak-empiricism excludes philosophical reasoning, and so precludes the knowledge that weak empiricism is true.

In seeking the nature of the justifications for philosophical assertions it is natural to turn to logic; for according to their proponents the truths of philosophy are like the truths of logic in being necessary truths discoverable by the exercise of reason unalloyed with experience. Indeed it was these similarities that lead many philosophers into claiming that philosophical claims were logical truths in their "deep grammar". This is one way in which the justification for our belief in logical truths might be extended to cover philosophical truths, but there are others. It might be thought that a denial of a logical truth breaches the norms of rationality, and hoped that the denial of philosophical truths might be similarly improper. Or perhaps the denial of logical truths might breach linguistic norms, the rules for proper use of the language in question. All these approaches rely upon the positing of objects realising properties or relations to which our epistemic access is not obviously compatible with our weak empiricism. The first approach, according to which true philosophical claims are disguised logical truths, hinges upon the relation of synonymy, which is supposed to hold between philosophical truths and their analyses. This is

a more exclusive relationship than that expressed by "may be translated as" or (for open sentences) "is co-extensional with", one for which we are owed an account of our epistemic access. The second approach has something of the opposite problem; if the relationship translation of is, as we have envisaged it, too weak to support philosophical analysis it is too strong to allow for any breach of rational laws. For translation on our account preserves empirical content, and inconsistent empirical theories are incompatible with every stimulus sequence, hence no-one holding a theory compatible with any stimulus sequence can be translated as asserting a logical contradiction (if homophonic translation renders them as asserting a contradiction it must be abandoned). If the laws of reason are to be transgressed by the assertion of contradictions then something must over-ride the preservation of empirical content in translation. We are owed an account of both our epistemic access to this something and to the rules of reasoning which it allows us to transgress The third approach faces the same problem accompanied with the hope that the requisite norms, and standards of translation may be cashed out in terms of linguistic propriety.

By contrast, viewing logic as a mere inference engine — a tool for formalising the inferential connections between sentences underpinning the construction of scientific theories, with logical truths the flying buttresses of the resulting edifice — appears much more readily reconcilable with our empiricist scruples. Quantificational calculus provides us with an epistemically accessible account of logical truths — they are the theorems of quantification theory, and our epistemic access to the process by which such theorems may be derived is comparatively unproblematic. But capturing these statements as the theorems of a formal system does not justify our belief in them. For justification we must turn elsewhere, something can be made of their indispensability for scientific theorising, but what is the nature of our warrant for believing this inference engine adequate?

These problems, and the question of whether logical truths are necessary, and if so whether our warrant for believing in that necessity might also extend to some philosophical claims, shall occupy us in the next section of this work.

Part II

Inference

Chapter 3

In search of necessity

A naturalistic account of the sensory evidence for our theories does not amount to a naturalised epistemology. The old yearning of strong empiricists for certainty was directed to two components of scientific theories: observational foundations and inferential connections between sentences. Like rising damp, certainty was supposed to seep upwards from the foundations and permeate the scientific edifices built upon them. In the felt need for certainty of inference the yearning lives on, its satisfaction requiring a non-natural element which can be either located in us, or in the world around us.

Once again Husserl is an eloquent opponent of the view endorsed here, regarding the naturalisation of reasoning as absurd. "[If reasoning is a natural process] is it not the case that the logical forms and laws express the accidental peculiarity of the human species, which could have been different and which will be different in the course of future evolution? . . . But at once another piece of absurdity arises. Can the cognitions by which such a view operates and the possibilities which it ponders make any sense themselves if the laws of logic are given over to such relativism? Does not the truth that there is this and that possibility implicitly presuppose the absolute validity of the principle of non-contradiction, according to which any given truth excludes its contrary?"[1].

At least part of what Husserl requires, in less exciting talk, is that truths of logic be necessary (though not that the truths of logic exhaust the necessary truths). There have, broadly speaking, been two approaches to necessity, one sought to explain necessity as a feature of the world independent of us, and the other which explained necessity as a product of our minds. Aristotle's approach to modality is an exemplar of the former, and Kant's of the latter.

The main difficulty empiricists have traditionally had with the former sort of approach, is accounting for our epistemic access to modal properties. The tra-

---

[1](Husserl, 1964)

ditional reason for rejecting Aristotelian essences being our inability to provide
any method to distinguish essential from accidental features. In recent times
modal realism has become more popular, largely as the result of developments
made in modal logic by Kripke, but it has never been acceptable to empiricists.
As a result of Kripke's development of models for quantificational modal logic
many modern modal realists base their analyses of possibility and necessity in
terms of possible worlds rather than taking the distinction between essential
and accidental features as fundamental. It should be noted however that while
Kripke's developments have shown (some) modal logic(s) to be consistent, they
do not show that there is any way of reducing de re necessity (essential prop-
erties) to de dicto necessity (necessary propositions), which is not either trivial
(for instance treating all properties of everything as essential) or arbitrarily se-
lective. This complaint only applies to modal logics while they are interpreted
as being modal (i.e. while "□" is interpreted as "necessarily"). Perhaps 'modal'
logic has striking applications in other areas (Hintikka hints at cognitive sci-
ence, another possibility is to interpret "□" as a Prior-style temporal operator)
but these have no connection with de re *necessity*. While Kripke's work has
shown a variety of modal logics to be formally consistent, he has not shown
which of this variety captures the notion of necessity — which system's "□" is
to be interpreted as "necessity". Despite his formal results there remains no
motivated choice of which quantified logic to treat as the correct account of
modality and the traditional complaint of empiricists remains in force, essential
features remain as inscrutable as ever, and hence so too does quantified modal
logic except as a variety of uninterpreted formalisms.

Even moderately realistic approaches to modality, such as David Armstrong's
combinatorial account of modality, according to which possible worlds are re-
combinations of the fundamental properties instantiated in the sole actual world,
is incompatible with our standard of evidence. There is no guarantee that there
is only one scientific theory on whose predicates we may base this re-combination
and even if there were, we would still require a plausible standard of individ-
uating properties, for properties used in such a fashion are themselves modal
notions and little headway has been made. Armstrong's account requires there
to be objectively existing universals of which we, through proper application of
scientific method, gain knowledge.

Nothing stops us from simply nominating the combinations of predicates we
are prepared to consider capable of being satisfied by the same objects, but
this does not establish the kind of 'absolute validity' Husserl felt need of. This
pale shadow amounts to relying upon logical validity (defined formally, without
recourse to modal notions) in place of validity (i.e. necessary connection) at
least if we resort to a standard logic for the machinery of our inferences. Even

if validity is distinguished from logical validity by extending to cover 'meaning postulates' or other statements compelling assent, reliance upon the notion of logical truth renders the claim that logical truths are absolutely valid trivial. Much the same attitude can be taken towards essential properties, by simply nominating the properties which we wish to hold essential. But mere nomination of such things by fiat does not establish the absolute validity of anything.

3.1 Conceptual Necessity

> Alice laughed. "There's no use trying," she said: "one can't believe impossible things."
>
> "I daresay you haven't had much practice," said the Queen. "When I was your age I always did it for half-an-hour a day. Why sometimes I've believed as many as six impossible things before breakfast."
>
> — Lewis Carroll, *Alice through the looking glass*

If problems of epistemic access preclude taking modality as an objective feature of the external world, the alternative is to take it as something we impose upon the world — a product of our language/concepts or their application to the world. This shift has its technical repercussions, we may drop the distinction between accidental and essential properties, so long as we are prepared to go without quantifying over modal contexts, treating as the bearers of modal properties not the objects to which we refer but the sentences or thoughts in which we refer to them. This course might equally be taken in a realist manner — by taking modality to be a feature not of objects but facts. But there is no reason to suppose the resolution of the world into such facts is itself scrutable; again we are struck by the lack of a guarantee of a unique best scientific theory, the lack of a unique set of properties and relations we may discern. The advantage of a conceptual/linguistic approach, over a 'factual' is supposed to be that that we have better access to our own concepts than to the facts. A cost of the conceptual course, if we wish to count the truths of mathematics as necessary, is that we may only include denumerably many numbers in our ontology. Given more numbers than this there would be necessary truths that could not be accounted for by the operation of recursive procedures on a finite number of primitive concepts, hence necessary truths we could never conceive of. Some conceptualist accounts of mathematical truth and our knowledge of it have staggering ramifications: perhaps the least dramatic is the rejection of higher infinities than the denumerable, but rejection of the law of the excluded middle and (hence) of argument by reductio ad absurdam have also been ad-

vocated in the pursuit of a conceptualistic foundation for mathematics. I shall not attempt to assess conceptual accounts of necessity on these grounds, instead sticking to questions of epistemic access and whether the explanation of necessity as conceptual/linguistic is any better on this score that an account given in terms of an independent and mystically perceived 'platonic realm'. On this count there is at least one immediate advantage to the conceptual approach — it allows us to thankfully turn our backs on the problems of epistemic access to the distinction between essence and accident; necessity attaches to judgements, statements or propositions in the first place and not to the more commonplace objects the judgements are about. Many, however, do not currently share this attitude and continue to attempt to reduce de re to de dicto necessity. We may marvel at their labour, but we need not investigate it. De dicto necessity alone would serve to clarify the absolute range of possibilities, and to distinguish valid from invalid inferences. It would fulfil the lingering desire for certainty we have inherited from reductive empiricists.

Adrian Heathcote has suggested that a dilemma reminiscent of Socrates argument in the *Euthyphro* arises in the explanation of necessity on conceptual/linguistic grounds. Are conceptual (/linguistic/analytic) truths necessary because they are conceptual or conceptual because they are necessary? "For example, does the law of non-contradiction seem true because of the arbitrary meaning assigned to 'and' and 'not', or is the meaning constrained to reflect some antecedent necessary logical truths?" [2]. Obviously if we seek to avoid modal realism we must attempt the first alternative; but then, Heathcote goes on to complain, "If we try to explain the necessity of necessary truths as arising out of analyticities then we will be left with nothing to explain how these analyticities arise. Language will be entirely contingent and so, consequently, will the necessary truths!". At this point however the analogy with the *Euthyphro* breaks down, if the good is nothing but the whim of the gods then it carries no moral weight; one may reasonably respond "So what". But one is not freed from the constraints of one's own language by recognising that those constraints are contingent.

This point has been clearly recognised by most empiricists, certainly by Ayer and Carnap. Carnap was quite clear that "In the language of the inhabitants of Britain 'Either it is raining or not raining' is analytic" was itself contingent. Indeed the whole point of the project was to explain how contingent beings such as ourselves whose epistemic access is mediated solely by experience could know some truths to be necessary, and the point of analyticity was to explain this by showing how contingent beings such as ourselves could give rise to such necessities. The supposed difficulty of the contingent foundations of necessity

---

[2](Heathcote, 2002, pp. 90-91)

is in fact the whole point of the exercise[3]. Only someone heavily addicted to modality could see this as a defect. The price to pay was that the necessity produced was only relative to language (as per Ayer) or conceptual framework (as per Kant). In the latter case a form of absolutism appealing to both Husserl and Heathcote can be redeemed provided our concepts are immutable and shape the phenomenal world, for then they are part of the necessary bedrock of this constructed world; though only by their own lights there is no other light to choose from. (There is no difference between claiming that position is absolute and claiming it is relative to the only true frame of reference. Relativism requires other equally good frames of reference.) If however one recognises other alternative conceptual schemes, and does not hold that one's current scheme shapes one's observations to fit then the adequacy of one's conceptual scheme is open to question. Some recalcitrant experiences are best adapted to by a change of scheme, others by a the re-evaluation of the sentences within that scheme. Still, one must always use some scheme, rely on some language to frame one's judgements and relative to that framework necessity is unshaken.

Such relativism still does not imply that merely thinking will make modal facts so. Even if necessity is language or 'framework' relative it is still a matter about which we can be mistaken even if that judgement of necessity or analyticity can only be made from within its own framework in turn. It is compatible with this approach that our reasoning processes might habitually or even inevitably (albeit contingently) go astray, misjudging or presupposing an improperly limited (or liberal) range of possibilities. Logic, on these views, is yet the study of how we ought to reason if we wish to preserve truth, and remains distinct from the psychological study of how we do reason.

The appeal of a linguistic/conceptual approach to modality is easy to understand. Prima facie it provides a way for empiricists to account for our knowledge of necessary truths, for surely we have access to our own minds and concepts, or equally to the conventions of our languages. If necessary truths are true in virtue of such internal considerations rather than some modal facts out in the world then the problem of our epistemic access is solved. It was this epistemological point, and not, for instance, a love of naturalism, or a prior hostility to metaphysics that underwrote analytic philosophy (though such hostility was certainly a motivating factor, the reasons for finding metaphysics unacceptable were epistemic).

---

[3]Heathcote goes on to say that only Quine was radical enough to "embrace this horn of the dilemma". There is here a faint echo of the terrible argument sometimes imputed to Quine that since any sentence can be revised there are no analyticities (cf,. §4.2). But surely it is obvious even on the most superficial reading that Quine's complaint wasn't that the only evidence for analyticity was behavioural (a fortiori contingent) but that there was no such evidence.

It should be noted that the value of this epistemic constraint, and hence the value of explaining necessity in terms of truth by virtue of meaning/linguistic convention alone (i.e. analyticity), relies upon a diagnosis of what it is that our "intuition of necessity" is an intuition of. There is no doubt an intuitive sense of difference between 'Spinsters are unmarried' and 'Some humans have died'. Taking this intuition at face value is an example of the pernicious effects of allowing introspection as evidence which we bemoaned in connection with phenomenalism. To repeat the maxim, the states we are aware of through introspection *do not come with labels attached,* and this applies as much to intuitions as to any other sensation. What this intuitive sense of difference is a sense of, is not settled by the quality of the intuition itself. Here as elsewhere the result of thinking otherwise is to stultify theory at the point where a terminology to describe our intuitions is first constructed. Even where progress is made, as for instance recasting this intuition as the apprehension of analyticity rather than the apprehension of necessity by the faculty of reason, progress is terribly slowed. To be sure, an alternative account of the 'intuition of analyticity' would be of interest, and almost equally surely Quine's diagnoses of it as either a rough sense of the disruption in communication attending dissent to the sentence in question (pace *Word and Object*) or metaphorical distance from the experiential boundary of our theories (pace "Two Dogmas") are both wanting, but though such an account of this intuition is desirable it is not directly required either in semantics or epistemology.

There was of course a reason philosophers took this intuition to be one of necessity, namely that when questioned people would assert that these statements could not be false. But if our sense of what could be the case hinged not on our sense of what is the case in other possible worlds or awareness of meanings/concepts (or more generally of linguistic/conceptual constraints) but on the constraints we impose on extrapolating from what we have observed to be the case to what we believe may (actually) occur elsewhere (or something equally acknowledged as contingent), then our sense that these statements could not be false would not indicate anything other than their importance in such extrapolation, and may prove irrelevant to necessity.

While it is thus open to explain away necessity rather than explain it, strong empiricists appeared to have a a good starting point for explaining necessity at face value. The quasi-cartesian notion of indubitable sense-data that we found wanting in §1.0.3 appears at first glance to give rise to just the desired distinction. If such sense-data were either mutually independent, capable of occurring in any combination, (like the sense-data that the empiricists tended to envision) or mutually exclusive (like the stimuli we settled for) and strong empiricism were true (i.e. every sentence could be reduced to one about sense-data) then the

necessary or analytic sentences would be those true of any possible combination of sense-data (or stimuli sequence).

Strong empiricists supposed sense-data to be known completely and indubitably. They were a species of ideas and, as Berkeley puts it, "I cannot be deceived in thinking I have an idea which I have not. It is not possible for me to imagine that any of my own ideas are alike or unlike that are not truly so. To discern the agreements or disagreements there are between my ideas, to see what ideas are included in any compound idea and what not, there is nothing more requisite than an attentive perception of what passes in my own understanding."[4]. This makes sense-data (or at least ideas) appear to be especially well suited to being foundations of modality. Potential misjudgement of which sense-data one is presented with is not a critical flaw in itself, but if it is possible to misjudge sense-data at all, then it seems possible to attribute to some sense-data a property that no sense-data possess, or to fail to attribute to some sense-data a property that all sense-data possess. This possibility is however precluded by analyses of necessity built upon sense-data. The problem arises because sense-data are required to do double duty: serving as conceptual atoms, whose recombinations in sense or imagination mark the limits of possibility, and also as part of the causal chain between the external world and our minds. The former role requires indubitable knowledge by acquaintance, but the latter exposes them to revision in the development of our ongoing scientific theories concerning perception.

Uncertainty about the nature of sense-data does not put paid to attempts to found an account of modality upon them, but it gives rise to a divergence between metaphysical modality, the range of combinations that sense-data *are* actually susceptible to, and epistemic modality, the range of combinations *we take them* to be susceptible to. This distinction is not particularly connected with attempts to found modality on sensory foundations, but arises in any attempt to found modality on uncertain foundations. It arises in Putnam's philosophy, for example, because he attempts to base modality on essential properties discerned by fallible natural science. Putnam claims that water is essentially H_2O; the same desire to move the modal qualifier to initial position in the sentence that is the technical mark of conceptual analyses of modality leads him to take this as entailing "It is metaphysically impossible for water not to be H_2O". But he makes much of the possibility of us being exposed to an *epistemic counterpart* of water that is not H_2O, that is the predicament the inhabitants twin earth are supposed to be in. Putnam's proposal is to live with the dualism of metaphysical and epistemic modality where strong empiricists (and, until recently, analytic philosophers) had hoped to explain the appearance

---

[4]Section 22 from *The Principles of Human Knowledge*

of the former in terms of the latter.

Certainty and transparency in acquaintance was only one of three virtues strong empiricists attributed to sense-data which equipped them to serve as modal foundations. The second was that all our beliefs and imaginings were attitudes towards combinations of sense-data or *ideas* which derived from sense-data (either by direct though faint imitation or, though this was a point of controversy between Locke and Berkeley, through abstraction). This is a dim doctrine, and one we would do well to avoid if we may, but it is an indispensable part of conceptual/linguistic accounts of modality. We have taken empiricism as an epistemic doctrine, and sought to demarcate the limits of human knowledge; advocates of this second doctrine take the same boundary as marking the limits of human belief and assertion, and see in empiricism a semantic or psychological doctrine. It is easy to see the how this applies to the strong empiricists: their epistemic doctrine was that sense-data were the only objects we could know of, the parallel psychological/semantic doctrine is that sense-data are the only objects about which we can have beliefs or make assertions. Our statement of the limits of empirical knowledge[5] at the end of part one also admits of just such a parallel: take translation into a language containing only logical terms and predicates expressing exposure to stimuli and the temporal relations between such exposures as a criterion of meaningfulness rather than of epistemic accessibility. It is tempting to take this boundary as marking the stronger limit, and more tempting yet to at least give the matter some consideration. However, as I intend to reject the postulation of concepts/meanings and the distinction between conceptual/analytic truths and synthetic truths on altogether separate grounds I propose to avoid any close consideration of this semantic/psychological doctrine. In the co-operative spirit of a reductio ad absurdam I propose to simply conflate the limits of what can be known with the limits of what we can believe for the duration of our consideration of conceptual/linguistic accounts of necessity.

The third virtue strong empiricists attributed to sense-data was that the ideas which echoed them could be re-combined in the imagination at will. The only limits upon imaginable combinations of ideas were the limits of possibility. Failure to *see* a black swan might be due to an unco-operative world, but failure to *imagine* a black swan could only be due (according to the most naive forms of strong empiricism) to black swans being impossible. Thus it was that Berkeley took the failure to imagine a line that was neither great nor small, black, white nor red, nor of any other determinate colour as demonstrating that that there was no abstract or general line named by the term "line".

---

[5]We are currently investigating whether our admission of logical terms required a justification which might, in turn apply to other truths than the merely logical.

The early British empiricists tended to talk of examining their own ideas as kind of looking within themselves, and this is not mere happenstance. Visualisation dominates the early British empiricists discussions of the imagination. One good reason for this, rarely recognised at the time, is that vision is on of the most nearly standardised of the senses. As we saw in §2.2 trying to provide a inter-subjective standard of identity for stimuli taking sensations of touch into account appears a hopeless task. While the same is true to a lesser extent of patterns of retinal irradiation or visual sensations the range of variation in the sensitivities of the retina of clear-sighted individuals is so minimal as to be easily ignored. Still, while both the uniformity of our retina and the comparatively developed knowledge of gross physiological structure of the eye may have contributed to British empiricist's focus on vision, perhaps the main reason for this focus was simply that the empiricist's accounts of necessity and conceivability works best when applied to the visualisation of positive possibilities. Can there be red squares, or isosceles triangles? Yes, one can imagine them. When applied to other sensory modalities the results are less clear: Is there a taste of strawberry marmalade on orange slices? Well, yes, I can imagine eating strawberry coated orange slices and I know it will taste of something, but for my own part at least I cannot imagine what it will taste like. This is perfectly in accord with Locke's psychological empiricism, according to which minds begin as tabula rasa and are equipped only with ideas echoing the sense-data to which they have been exposed. Psychological empiricism implies that there is a distinction between the range of one's imaginations and the range of possibilities — the scope of one's imagination may be narrower for want of experience. Objects may instantiate properties one has never observed and hence cannot imagine. But this much might be conceded without too much adjustment by empiricists wishing to maintain an internal sensory account of our epistemic access to modal truths. Locke, in taking this line, supposed each thinking being to be capable of judging the identity and difference of its own ideas infallibly. Thus one can tell with certainty that *Red is not Blue* because one can distinguish with certainty the idea of Red from the idea of Blue; and one can tell that *A Circle is a Circle* because one can identify occurrences of the idea of a circle. But this device works only because the identities revealed by the contemplation of one's ideas are already supposed to be necessary identities and the logical principles governing identity (in particular the transitivity of identity, under which Locke tries to subsume all reasoning) are assumed to be necessary truths. In fact Locke relies upon the impossibility of imagining a circle that is not a circle or a colour that is both red and blue. This is all the more apparent in those truths which Locke proclaims to be self-evident which do not easily fall under identity principles such as that *Two bodies cannot be in the same place* or that *A man*

is not a horse. In all these cases the underlying justification is that one fails to imagine objects of which the proposition is false: one cannot imagine a circle that is not a circle, or a man that is a horse, or two bodies in the same place. The empiricist account of modality (for psychological empiricists) is thus that a proposition is necessarily false if a person acquainted with the ideas expressed in it cannot imagine any objects of which it is true. The empiricist account of our knowledge of necessity works best when applied to visualisation, because it is fairly straightforward to take the simple ideas we are acquainted with through vision as coloured points or regions, but it is far from clear what the simple ideas of the other sensory modalities such as taste are, and thus more difficult to apply the empiricist account of our knowledge of modal truths. (Am I unable to imagine the taste of strawberry marmalade on orange slices because there is no such taste or because I have not been acquainted with the requisite simple tastes. Perhaps strawberry on orange is a simple taste in its own right).

Even on their home ground of visual imagination British empiricist's explanations of our knowledge of modal truths run into trouble. Thomas Hobbes famously took himself to have squared the circle, and one may suppose that he visualised (each step of) the geometrical construction by which he thought himself to have achieved this task; but not only have Hobbes constructions been shown not to have squared the circle, it has been proven that there are no such constructions. It would seem that Hobbes visualised an impossibility, and hence that his visual imagination proved not to be a reliable guide to modality. However empiricists have at least two plausible lines of defence. Berkeley, always warning of the confusions due to misleading nature of language, might have diagnosed Hobbes' conceit as turning upon a mislabelling of his ideas. Perhaps Hobbes has misused the terms "same area as" to express the idea of "an arbitrarily close approximation to the same area" or perhaps the object from which he began his constructions was not a proper circle. In this way an empiricist might reconcile the claim that Hobbes' visualisation had revealed a necessary truth while denying that that truth specified a method of squaring the circle. Descartes is the best known proponent of the other line of defence, which relies upon distinguishing between combinations of ideas that are clearly and distinctly conceived from those that are not; only clearly conceived combinations of ideas reflect genuine possibilities. On this view Hobbes may not have conceived of anything clearly when he took himself to have squared the circle. This has the disadvantage of introducing an epistemic gap between us and our imaginations, we may erroneously take vague and imprecise imaginations to be clear and distinct. At its most unbridled such scope for error would again lead to an externalist account of modality with the attendant problems of epistemic access.

Historically the final nail in the coffin for attempts to provide sensory foundations for modality was not struck by any philosophical objection but by discoveries in empirical science. Since Einsteinian relativity was introduced in 1905 physicists have delighted in telling us that physical space-time is a non-Euclidean space, to which our visual imagination could not do justice. Einstein himself was well aware that "philosophers have had a harmful effect upon the progress of scientific thinking in removing certain fundamental concepts from the domain of empiricism, where they are under our control, to the intangible heights of the a priori"[6]. Under his influence the physicists succeeded in bringing the concepts of space and time "down from the Olympus of the a priori" and the tremendously wide dissemination of at least the most superficial aspects of his theories in popular western culture has nearly eradicated the belief that the visual imagination is a reliable guide to physical possibility.

Though we cannot visualise some of the objects postulated in modern physics, it might yet be thought that we can visualise their observable effects upon us. We might for instance imagine stars which are almost blocked from view by the Sun being visible at a slightly different position than they would be if physical space were Euclidean and three-dimensional. We can even visualise beams of light being deflected in their passage past the sun or other massive objects. This suggestion marks the shift from the traditional empiricist semantics in which the meanings of (non-logical) terms are sensory ideas, to the verificationist semantics of their positivistic heirs. The class of observations confirming or infirming a statement[7] is not itself a sense-datum nor a sensory idea. If these classes are not recursive but only recursively enumerable then (on the assumption that our own mental powers are restricted to the computation of recursive functions) their members cannot all be envisaged at once, rather we must settle for a disposition to to generate the corresponding ideas or expectations. Something of this can be found to be implicit in the British empiricists' account of ideas, for their ideas were connected with each other through association (in modern jargon, stimulus-response conditioning). Thus the idea of a ball approaching a solid object was associated with the idea of it bouncing off in another direction. Such associative connections implied that visualisations were connected together to form a network of expectations. The postulation of unvisualisable objects in physics shows that the nodes of this network need not themselves be sensory ideas.

So it was that sense-data and the ideas echoing them were found to provide inadequate foundations for modality. They are neither known infallibly, nor do

---

[6](Einstein, 1956, p. 2)

[7]Supposing, as both the British empiricists and the positivists did, that individual sentences can be confirmed in isolation.

they exhaust our epistemic/semantic reach, nor does the failure to visualise an object of a given type show that it is impossible for objects of that type to exist. The shortcomings of sense-data did not however put paid to the attempt to find internal foundations for modality. Instead concepts were posited in their stead. These concepts were the old sensory ideas shorn of their sensory component. Just how these concepts came by their representational content (i.e. were true of or extended over other objects) if not by resemblance was and remains a matter of dispute. But these concepts were to have all the virtues required for an internal account of modality that sense-data so sadly lacked. Since to form a judgement or belief is, on the conceptualist account, to stand in a relationship to a composition of concepts alternative concepts, unlike alternative sense-data, are inconceivable as a matter of course. This absence of alternatives, together an understanding of philosophy as conceptual analysis, are reflected in (early) Wittgenstein's claim that " If one tried to advance theses in philosophy, it would never be possible to debate them, because everyone would agree to them"[8]. Theses and anti-theses reflect genuine possibilities whereas the negation of a necessary truth reflects not a genuine possibility but only conceptual or linguistic confusion. Putative assertions of such statements are to be dismissed as either a mis-expression of some other belief, or incoherent babble. These are the same two explanations of conceptual error we noted above — Berkeley's account in terms of linguistic confusion, and Descartes requirement of clear and distinct conception — stripped of their sensory aspect. Philosophical dispute is not precluded, but such disputes are to be resolved by showing them to be merely notational, or that one or other party had confused their concepts (perhaps using the same word to express different concepts on different occasions), or had so totally failed to grasp the concepts normally expressed by the terms of their attempted assertions that they were not really asserting anything at all and there was never really a dispute between two views in the first place — never by showing that a disputant's assertions were false.

Concepts also inherited, mutatis mutandis, the second requisite virtue. Just as the British empiricists took belief to be an attitude adopted towards composites of sensory ideas, their conceptualist inheritors took beliefs to be attitudes towards (or relationships with) composites of concepts. The limits upon such compositions thus marked the boundaries of our epistemic/semantic reach.

The third virtue went over into the new form too, though in a more modified (and hence more contentious) fashion. For Berkeley it was the failure to visualise an object that marked the impossibility of that object's existence. The failure of his imagination was the justification for concluding that there cannot be round squares of unsensed objects. Conceptualists tend to allow that

---

[8](Wittgenstein, 1953), comment 128

necessary falsehoods can be thought and asserted, and hence that the constitutive concepts allowed of the requisite composition. One must have a concept of a round square if one is to deny that any exist. But, just as introspection of attempts at visualisation was thought to reveal modal truths, the private contemplation of concepts is thought to reveal necessary truths and falsehoods. Still the old Berkeleyan sense that necessary falsehoods were distinguishable by the lack of a corresponding composite of ideas did not disappear altogether, manifesting intermittently in analytic philosophers' occasional denunciations of contradictions and other necessary falsehoods as nonsense.

Concepts were thus custom made to possess all the virtues required of modal foundations. At last philosophers had a plausible sketch of how belief in modal truths might be justified which at least appeared to be compatible with empiricist epistemology. A certain self-discipline and the abandonment of some over-blown ambitions was called for, but some proclamations of discoveries of necessary a priori truths, or necessary and sufficient conditions, so characteristic of philosophical work from the Ancient Greeks to the early twentieth century, might at last be vindicated. In particular, conceptual truths might be invoked to quell skepticism. With regard to skepticism about physical objects for instance, conceptual analysis might reveal necessary connections between observations and the objects postulated in reports of them. But such feats can only be accomplished if we can form justified beliefs about which concepts both believers and skeptics are expressing in their claims. If concepts are supposed completely private, not only do we face the same problems as phenomenology[9] and there can be no public collaboration in the discovery of necessary truths, but even worse, even our capacity to understand one another's everyday assertions is rendered inexplicable. For conceptualist semantics explains the necessary truth of assertions and claims in terms of the concepts those claims express. But this relies upon the principle that assertions generally have their content in virtue of the concepts they express. If the expression of concepts is inscrutable then one cannot tell what, if anything, other people are saying.

The very promise of success in refuting skepticism made the need for public evidence of concepts all the clearer. Just as conceptual analysis was expected to provide a refutation about physical objects, it was also expected to provide a refutation of skepticism about other minds. Traditional British empiricists had taken perception as a completely private affair, and tended to look to God's benevolence to ensure that our private perceptions reflected a larger reality. So long as the logical tools required to elucidate the connections between our observations and our talk of objects were unrecognised, no account of how we gained knowledge of anything beyond the screen of our own mental

---

[9]cf. §2.1

cinema could be provided, God's help appeared to be required and skepticism about other minds presented no special difficulty except as a problem for our knowledge of God's mind (on which point Berkeleyan/Cartesian theology was at least self-consistent). But conceptual analysis, together with a more developed understanding of the logical operations by which complex concepts might be composed, promised an alternative justification for belief in external objects. To remain dependent upon God's grace for our knowledge of other minds was thus increasingly intolerable.

The most natural account of our knowledge of other people's concepts and the public evidence for ascriptions of relations to concepts and conceptual contents arises from taking concepts as objects posited in psychological theories, theories for which the relevant public evidence is public behaviour. Undertaken naively this approach leads to a form of psychologism, according to which conceptual truths reflect mere psychological regularities. At their most extreme, advocates of psychologism took logic to be an attempt to discern general laws about psychological posits (such as ideas, beliefs and judgements) predicting observable regularities (e.g. that everyone who assents to each of two sentences will also assent to their conjunction). Boole is remembered as one such advocate, chiefly for the title of his book *An Investigation of the Laws of Thought*. In the early twentieth century psychologism was a hot topic, especially in German speaking countries, with Husserl and Frege the leading critics. A great deal of effort was spent by them and their supporters and opponents in determining just what counted as psychologism and just why it was unacceptable but, in brief, there were two central and independent lines of attack.

The first was based upon the principle that psychological laws could not provide a justification for absolutely certain belief in logical and conceptual truths. Psychological theories are empirical and fallible, their laws inductive generalisations which cannot be known with certainty. Furthermore, unless believing truly was a distinguishable type of *psychological* state from believing falsely[10], such generalisations could only serve to record the opinions of the common herd. Visual illusions and schoolboy fallacies provided examples of false beliefs which could be reliably induced and suggested, naturally enough, that some false beliefs might be so invariably induced as to be held universally. Universality not only fails to guarantee necessity, it fails even to guarantee truth! So, at least, the criticism goes. It is particularly associated with Husserl, though it lives on, as we have seen, in such variants as Heathcote's objection to explanations of necessity in terms of analyticity in which it is directed to

---

[10]To see the sorts of problems this raises consider combining the view with a functionalist account of psychological states. It would have to turn out that true and false beliefs played a systematically distinguishable causal role mediating stimuli inputs and behavioural outputs. See (Fodor and Society, 1987) for an example of someone jumping up an down on the point

linguistic rather than psychologistic explanations of necessity.

The critic's premise must be conceded, the inductive generalisations of psychological theory are not known with certainty nor do they justify certain belief in conceptual truths. But it is not obvious that this implies that the concepts directly encountered in reasoning and other mental activities cannot be identified with the objects postulated in some psychological theories. At least at first glance it appears perfectly compatible for belief in conceptual truths to be predicted by laws of thought and justified by the direct examination of concepts. Mismatch between the confidence with which we hold conceptual truths and the qualified warrant for such confidence provided by empirical psychological investigation (if any) can be explained by the existence of a separate justification provided by conceptual analysis.

If this objection to identifying concepts with psychological postulates is to carry any weight at all it must be because a psychological explanation for belief in conceptual truths undercuts the justification for those beliefs. Suppose psychology and associated sciences developed enough to allow people's beliefs to be determined by examination of their physiology and that a psychological/physiological mechanism was found in all or nearly all humans which reliably caused them to believe that it is raining when asked "Is it raining?". Clearly such a discovery would undermine our confidence in the instances of the belief that it is raining instantiated by human beings. Not only would it undermine our confidence in future but it would turn out that the justifiable confidence with which had people believed it was raining all along was generally much lower than thought before this mechanism was discovered. The belief that it is raining could be caused by events that that did not warrant any confidence in the belief that it was raining, and only to the extent that we were sure that our belief was not due to such aberrant causes (i.e. that no-one nearby had just asked "Is it raining?") could our confidence in such judgements justifiably re-approach the levels we had previously had in them.

The mechanism of belief causation just imagined does undermine the justifiable confidence with which the type of belief it causes can be held. But it is easy enough to imagine other mechanisms of belief causation which do not do so, and easy to describe some which do so only to a lesser degree: Say psychologists discovered that being struck by small drops of water while believing oneself to be outside, reliably caused the belief that it was raining in most human adults. Such a discovery would show that we had over-estimated the justifiable confidence with which "It is raining" beliefs can be held to some degree. It would turn out, among other things, that being sprayed with a garden hose was more likely to deceive than had been thought, and the confidence of our assessments of the truth values of assertions of "It is raining" by humans would be unjus-

tified if, knowing this, we did not take it into account. Though the discovery of such a mechanism would still undermine our justifiable confidence in "It is raining" beliefs, it would do so to a lesser degree than the discovery of the mechanism described in the previous paragraph. Bayesians tell us this is because, in part, the probability of being rained near given that small drops of water are striking one's body and one believes oneself to be outside is much greater than the probability of it raining near someone given that they have been asked "Is it raining"[11]. If we were prepared to describe more complex and discriminating mechanisms which caused "It is raining" beliefs, there is no reason to suppose we could not describe one which justified the same degree of confidence that we currently hold in such beliefs or even higher.

Since the mere existence of a psychological belief causing mechanism need not undermine the justifiable confidence with which the resulting belief can be held, psychologism's critics owe an account of why identifying concepts with psychological posits is problematic. It is granted by both sides that such a theory would have laws covering the causes of beliefs. (Functionalism is marked by the stronger claim that psychological states are identified by their causal role and is the subject of some debate, but surely it is reasonable to assume that psychological states have causes and that at least some psychological belief-desire theories had better tell us about those causes if belief-desire psychology is to be worth pursuing at all.) But why, a conceptualist might ask, should the discovery of the causal limits of our psychological machinery and the limits on the patterns of belief it can give rise to, in particular that there are some beliefs it will always cause to be held with certainty, undermine our confidence in those certain beliefs? Obviously, since these are supposed to be necessary truths it isn't that, like the first crude "It is raining" belief causing mechanism imagined above, it causes the belief even when it is false. The critic's answer is that our realisation of some contingent empirical psychological theory is irrelevant to the necessity of some truths. Since those truths are supposed to be true *no matter what psychological theory we realise* then our realising a certain psychological theory provides no reason to believe those truths. In Bayesian terms the probability of it being true that "It is raining or not raining" given that one realises a particular psychological theory is the same as its probability of being true given that one doesn't, hence conditionalising on the discovery that one realises a psychological theory shouldn't change one's confidence in the belief that "It is raining or not raining" but, according to psychologistic accounts of necessary truths that is just what causes this belief; since the belief in necessary truths

---

[11]The full Bayesian story involves the prior probability of it raining and the probability of it not raining given that one is being struck by water droplets and believes oneself to be outside/has just been asked "Is it raining".

is caused (solely) by something that does not justify it the resulting belief is unjustified.

The problem can be put without the Bayesian framework. All parties to the dispute agree that whether humans realise this or that particular empirical psychological theory is a merely contingent matter. Psychology is not to be deduced from conceptual truths but inferred from observations of behaviour. If our concepts are the objects posited in these theories then it is a contingent matter which concepts we have and which of our beliefs are conceptual truths; but, the critics have it, if necessary truths are just conceptual truths then it is a contingent matter which truths are necessary. Husserl's rhetorical question as to whether "logical forms and laws express the accidental peculiarity of the human species, which could have been different and which will be different in the course of future evolution?" is thus the conclusion of a reductio against psychologism. The critic contends, in a nutshell, that psychologism implies that it is only contingent that certain truths are necessary whereas, according to standard modal intuitions, necessary truths are necessarily necessary, hence psychologism is false. The defence against this charge turns on the rejection of the first of this arguments premises.

The critic's mistake, according to the defence, is to think of creatures realising different psychologies as grasping the same concepts (i.e. capable of asserting or believing the same truths) at all. On that image the same truth could be a conceptual truth for psyches realising one type of psychology, and a non-conceptual or factual truth for psyches realising another. If that were so then psychologism together with a conceptualist account of necessity would imply that necessity was relative and so, since necessity is not relative, one or the other must be false. The mistake, the defence goes, is the result of failing to recognise that the ultimate truth bearers are beliefs composed of concepts. Psyches for which different truths are conceptual truths, must have different concepts, and hence the same truth cannot be conceptual truth for one type of psyche and not for another. Any creature for which a logical truth is not a conceptual truth is not capable of asserting or denying or even believing that truth since, ipso facto, they lack the requisite logical concepts.[12]

Thus far the defence has shown that psychologism does not imply relativism nor that some truths are necessary truths only contingently. But the critic may still wonder what reason there is to suppose that putative conceptual truths are true. Conceptualists always had a hard time explaining just why concep-

---

[12]Whether distinct psyches could share just some of their concepts is still a matter of dispute. Critics of conceptualism tend to argue for an implausible conceptual holism in the hope of a reductio ad absurdam, while proponents of conceptual explanations of necessity have usually attempted to delineate conditions under which subjects may share just some of their concepts.

tual truths had to be true, regardless of their position on psychologism. Kant
explained some conceptual necessities in terms of conceptual containment: ac-
cording to Kant the sentence "John's red rose is red" is a conceptual truth be-
cause the concept expressed by "John's red rose" already contains the concept
expressed by "is red". Locke explained conceptual truth in terms of conceptual
identities and differences, on his view "Red is not Blue" is a self-evident con-
ceptual truth because the concepts expressed by the terms "Red" and "Blue"
are distinct. These two views are quite distinct, Kant's explanation cannot be
made to apply to "Red is not Blue" unless we suppose that colour concepts each
contain within them the denial of the other colour concepts and this gives rise
to an infinite regress (for each denial will similarly have to contain the relevant
colour concept which will itself have to contain further denials) while Locke's
account appears to tacitly rely upon an unexplained distinction between merely
distinct concepts and mutually exclusive concepts; but the identity and compo-
sition of concepts are such murky matters that no absolutely clear account of
the differences between the two can be given. In any case neither explanation
is easily brought to bear on some logical truths. "Either it is raining or it is
not raining" is not of subject-predicate form, and while we may grant that the
concept expressed by "it is not raining" is distinct from that expressed by "it
is not raining", so (presumably) is the concept expressed by "Today is Friday"
and yet it obvious that "Either it is raining or today is Friday" does not express
a conceptual truth since that sentence sometimes expresses a falsehood.

The most famous attempt to apply the conceptualist account of neces-
sary truths to logically valid truths is given in Wittgenstein's *Tractatus* logico-
philosophicus. Wittgenstein explicitly states in the preface that his aim is to
draw a limit to thought or rather to the expression of thought, or in other words
to give a conceptualist account of necessity. His account of why the world must
satisfy our conceptual truths begins with then assertion that "There must be
something identical in a picture and what it depicts, to enable the one to be a
picture of the other at all."[13]. He goes on to assert that this something in com-
mon is the "form" and the form of an object is "the possibility of its occurring
in states of affairs."[14] Thus, to recast of talk of forms in terms of possibility,
for a belief to represent an arrangement of objects, there must be a one-to-one
correspondence between the ways it is possible to arrange the concepts com-
posing that the belief (to form other beliefs) and the ways it is possible to
arrange the objects represented. It is striking just how often "Wittgenstein's
correspondence theory of meaning" is criticised in recent philosophical litera-
ture (like psychologism it is a doctrine better known for being false than for

---

[13](Wittgenstein, 1922, 2.161) but see 2.16 through 2.171
[14]*ibid., 2.0141*

any of its features) and even more striking how infrequently it is noted that the correspondence involved is modal. If one hears Wittgenstein's correspondence theory of meaning mentioned at all, one is apt to hear it criticised on the ground that "we cannot get outside our own heads to check if our concepts correspond with objects" whereas the point of Wittgenstein's doctrine is precisely that we do not need to "get outside our own heads" in order to be assured that our concepts correspond to the elements of states of affairs, it is guaranteed by the very nature of representation. Conceptual truths must be true of the objects we talk of if we are to talk of them at all "we cannot say of an illogical world what it would look like" and so on.

Wittgenstein's account of our knowledge of modal truths relies upon two principles. Firstly we know the possible combinations of our concepts directly. Secondly, whatever our concepts represent must be something with an isomorphic structure of possible combinations of its component parts — the same form or "mathematical multiplicity". If this second principle can be known and logical implication preserves epistemic access, then it follows that we can know which arrangements of the objects represented by our concepts are possible. Famously, Wittgenstein runs into difficulty stating this second principle. According to Wittgenstein a belief represents its subject matter by means of the arrangement of the concepts which compose it, and represents that subject matter as being arranged in a corresponding manner (i.e. the manner to which the relevant isomorphism between possible arrangements of concepts and objects maps the arrangement of concepts composing the belief). A belief thus marks out one possible arrangement or a class of possible arrangements [15] of the objects it represents and says that one of these possibilities is realised. But this account breaks down for this second principle. There is no range of possible forms among which the form of our concepts (and the world) is one; and the claim that the structure of possible beliefs is isomorphic to that of possible states of affairs does not mark out those possible worlds in which this isomorphism obtains or rather it marks out all of them for it excludes no possibility. It is difficult to see how this principle can feature in a justification of our confidence in conceptual truths; at very least it is circular, since it is a conceptual truth itself, and as such, in any case could never be an essential. The problems of explaining how such a claim could at be both utterly vacuous and an essential part of our understanding of language and necessity motivated Wittgenstein's

---

[15]There is here a problem that Wittgenstein fails to address. Pictures are never general in the way that some beliefs are. One can believe that the policeman's hair was red or green (and neither of the disjuncts) but one cannot have a picture in which the policeman's hair is red or green. Though Wittgenstein clearly acknowledges that some beliefs are general in this fashion (a proposition is like "a space bounded by a solid substance in which there is room to move"), and also equally clearly asserts that "A proposition is a picture of reality" he never seems to recognise the conflict between the two. Let us be generous.

dark distinction between saying and showing. On Wittgenstein's view "Either it is raining or it is not raining" expresses a conceptual truth and says nothing about the state of the world or of our language; but *that* it is a conceptual truth shows something (unstatable) about the concepts of which it is composed.

Just how dark Wittgenstein's distinction between saying and showing was is the subject of many philosophical papers and articles (what can't be said can't be shown and can't be whistled either sums up many critic's responses) but it is not our problem here. There is nothing in this principle of isomorphic possibilities which presents difficulties for psychologism. If according to empirically discovered laws certain contingent arrangements of neurons are disposed to produce firing patterns possessing the requisite mathematical multiplicity then they realise instances of the relevant concepts. Of course most arrangements of neurons will not possess such dispositions, and will not realise the corresponding concepts, and of those that do the properties they possess that are irrelevant to the dispositions realising the requisite pattern are irrelevant to cognitive functioning. Our beliefs in necessary truths cannot be justified by empirical psychological investigations, psychologists cannot hope to discover new concepts or new conceptual truths (though their investigation might prompt conceptual analysis which can[16]). Any such putative discovery of a novel conceptual truth must be backed up by first person conceptual analysis if the generalisations which give rise to it are not to be dismissed as non-cognitive interference. Conceptualism is thus incompatible with a strong form of psychologism which takes concepts and conceptual truths to be discoverable by empirical investigation. For Wittgenstein that would be to confuse psychology, or more generally science, with philosophy, the resulting "discoveries" of conceptual truths due to the imposition, by scientists unpractised in techniques of conceptual analysis, of their own conceptual confusions. While that strong psychologism which claimed to have found in psychology an alternative path to the discovery of necessary truth is thus incompatible with conceptualism, that does not preclude the identification of concepts with the objects postulated in empirical psychology.

Even if this defence does refute the critics' first line of attack on psychologism that does not show that concepts can be identified with psychological postulates. That depends in part, of course, on the relevant psychology. The psychology of molluscs is presumably not sophisticated enough to warrant the identification of logical concepts with any of the objects playing a role in the etiology of mollusc behaviour. All parties to the dispute over psychologism agreed that psychologist's observations were contingent and the theories they justified only

---

[16]This is in a way an inversion of the role introspection can play in psychological investigations (cf. §2.1). Where for the psychologist introspective observations can only suggest avenues for public investigation, for the conceptual analyst public investigations can only suggest avenues for conceptual analysis.

contingently true (this was at the core of the first line of attack). The defence showed, at very best, that it was possible for concepts to be identical with psychological posits, it was left as a contingent matter, to be discovered by empirical psychology, whether such identities actually obtained. The second line of attack upon psychologism is that such identities do not obtain as a matter of contingent fact. But this is only to put the objection superficially; if as a matter of fact, the objects which our empirical observations justify positing do not include concepts, then it seems we have no reason to suppose that other people have concepts or are acquainted with them at all. Indeed the modest psychologism we have been defending is only a little more than a conceptualistic denial of skepticism about other minds, a defence of the principle that our empirical observations can justify beliefs about the concepts and beliefs of other minds; still, it is little more than this since in focusing upon *psychologism* we have shifted to the narrower claim that there can be *behavioural* evidence of other people's beliefs and concepts. In any case the arguments that concepts may be identifiable with psychological posits we have considered readily go over into allowing identifications with other empirical posits as well.

Perhaps there really have been some philosophers who have been skeptical about the existence of other people's beliefs, but most have sought to reconcile our knowledge of one another's beliefs and concepts with our failure to live up to the empirical predictions of a theory whose objects can be identified with concepts. Conceptual errors of all kinds, the assertion of inconsistent statements, the denial of conceptual truths and the assertion of conceptual falsehoods, are commonplace according to such philosophers, but this does not show that we do not have concepts and beliefs, but rather the very capacity to make such errors is constitutive of our possession of concepts. On this view the fundamental mistake of psychologism, and indeed of the sort of conceptualism we have described so far, is the presumption that concepts are posited in a positive theory based upon the quasi-introspective observation of them, instead these philosophers have suggested that concepts are posited within a normative theory of our behaviour and beliefs. On this view logic, and perhaps more broadly philosophy, is the study of how we ought to reason and hence not to be confused with psychology which is the study of (among other things) how we do in fact reason.

Prospects for a knock-down argument against a positive psychological account of concepts based upon the assertion/belief of conceptual falsehoods are poor. Even if it is granted that, according to psychology, people do believe and assert logical contradictions, that does not show that the laws of logic, for instance, cannot be psychological laws. They cannot be exceptionless laws of course but it is not clear that psychological laws need to be. Perhaps the req-

uisite psychological laws are the laws of an idealised theory (on such a view
rational beings are like frictionless planes, or mass points.); or perhaps speakers
do not assert contradictions when the contradictory statements are all being
attended to or are somehow "in focus" (on that view the laws of logic would
be the laws of the cognitive psychology of subjects who could keep all their
beliefs "in focus" at the same time); or perhaps subjects fail to draw all the
logical consequences of their beliefs because the drawing of such conclusions
requires computational resources of which subjects have only a limited supply.
Even if psychology says *nothing* about such caveats or idealisations and gives
no explanation of why peoples' beliefs fail to respect logical laws and relations
that doesn't show that those laws aren't laws of psychology (nor that the psy-
chological theory in question is false). Psychology is a special science and the
laws of special sciences do sometimes admit of exceptions which do not falsify
them. The best-known proposed explanation for this is that the laws of special
sciences are qualified by ceteris paribus clauses, which cannot be spelt out in
the terms of the science in question.

On the other hand, no one has the remotest idea of what the relevant ceteris
paribus conditions might be (or what the point is of a science in which they
are so widely falsified). In geology and other special sciences failure of ceteris
paribus conditions can be explained in the terms some other science — the
corresponding explanations for the failure of laws of rationality is conspicuous
by its absence. Also conspicuous by its absence is the lack of any account of
why psychologists should idealise to rationality, or how closely ordinary believers
approximate this idealisation. In fact psychologists seem to have a great deal
of interesting claims to make about peoples conceptual failures and propose all
sorts of laws to explain people's irrational behaviour.

Between the morass of qualifying clauses and other exceptions that positive
conceptualists can call upon and the generally woeful state of cognitive science
(as the relevant part of psychology has come to be called) there is precious little
hope of coming to a decisive resolution of the dispute between positive and
normative conceptualists. Still, the striking virtue of the normative account is
that it is happily compatible with the common-sense view that poor old Hobbes
really did think he had squared the circle, and that in doing so he made a grave
error. Norms, unlike laws, are built to be breached.

On the normative view there need not be anything in a person's mind or their
dispositions to behaviour that determines the conceptual content of their beliefs,
or the concepts expressed in their assertions. To have beliefs and concepts need
not be a matter of realising any positive theory but of being held to a certain
sort of account, or being properly held to such an account.

There are many different accounts of the norms upon which concepts and

conceptual content depend and they have been taken as applying to a wide range of different objects, states and actions. Most analytic philosophers followed Ayer in taking the relevant norms to be rules governing the use of language, but other approaches have been suggested. Lotze conceived of logic as an ideal of thought, approximated by our actual enquiries and some of Dennett's comments suggest that he takes a similar view[17]. McDermott has proposed a view which contains both positive and normative elements; he treats beliefs as the elements of a positive psychological theory, but relies upon norms to fix which assertions it is proper to make in which belief-states. thus for Ayer it is simply wrong to *assent* to "It is raining and not raining", for Dennett and Lotze (and perhaps Husserl) it is wrong to *believe* "It is raining and not raining", and for McDermott there is no mental state in which it is proper to assert "It is raining and not raining".

No matter which norms are relied upon in a normative analysis of concepts and conceptual truth, the same problems of epistemic access which drove conceptualism in the first place still apply. Though many modern philosophers hold that all norms are human creations, it has sometimes been suggested that conceptual norms are objective features of the external world: orthodox catholic doctrine has it that rationality is part of the image of god in which man was created, and there are less religiously inspired views according to which rationality lies in mirroring the properties of external objects in our concepts. Regardless of whether there are external norms and values or all values are created by sentient beings, such external norms are clearly unsuited to be linguistic/conceptual norms which give rise to conceptual truths. The same reasons that precluded externalist accounts of modality and motivated attempts to explain modality in terms of our concepts, also preclude the explanation of modal truths in terms of external conceptual norms. Just as concepts are, on the positive account of them, something about us which gives rise to modality so too the relevant conceptual norms must be our creations; the result of our adoption of rules or laws or systems of rights and obligations created through social (and perhaps private) practices.

Normative analyses of concepts have their own special problems in addition to those of epistemic access. They invert Kant's account of rationality and moral duty, for instead of "deducing" norms from the requirements of rationality, the objective is to derive rationality (which we may take as the absence of conceptual error) from normative practices or attitudes. Analyses of conceptual/linguistic necessity in terms of intentional normative attitudes are circular. If normative status depends upon intentional attitudes (such as believing to be improper, or intending to punish) then nothing is accomplished by such analyses except

---

[17]Though he is generally cagey about the norms of rationality, he is explicit in claiming that intentional agents ought to *believe* the logical consequences of their beliefs.

finding a new entry point into the circle of interdefinable intentional terms. It might please draconian headmasters and petty officers to believe the unpunished life is meaningless, but it will do little to allay the concerns of those who find intentionality and modality mysterious. If circularity is to be avoided without rendering our epistemic access to these norms inexplicable, the relevant norms must be a product of, or implicit in, our social practices. In recent times a version of this approach to conceptual content has been developed by Robert Brandom in his book *Making It Explicit* but it is perhaps best known from the later works of Wittgenstein. Indeed the shift from a positive to a normative account of conceptual content and necessity is the key difference between early and late Wittgenstein. As we would expect in such attempts, both Wittgenstein and Brandom emphasise actions, practices, sanctions, rewards and punishments, and take them as the fundamental states and events in terms of which our concepts are to be understood.

One must be something of a Calvinist to analyse intentions, beliefs and such like in terms of normative practices such as punishment and reward. It must, at the bottom level, be deeds and not intentions which render people liable to punishment and reward; and these responses must themselves be taken as punishments or rewards independently of the intentions of the participants. Such an approach appears tremendously unpromising. Ordinarily we think of punishments and rewards as being intimately connected to the beliefs of the participants of the social practices involved. One supposes, intuitively, that we can make sense of the suggestion that being killed as a sacrifice was a *reward* for victory in the ritual ball games of Aztec society, only against the background of the Aztec's religious beliefs. It is far from clear that there is any theory in which the term "punishment" plays a role except those incorporating belief-desire psychology. Any such theory will need to posit states and objects internal to the people realising a social system (for much the same reason that we need to posit external objects cf. §1.1.2), but the only candidate theories are belief-desire psychology (in which the internal objects posited are beliefs, or to make the matter clearly circular, concepts) or various forms of non-intentional science. It is plausible that theories of the first sort can give an account of punishment, but non-intentional theories of more basic sciences seem hard put to make sense of even the notion of behaviour. A thorough-going attempt to give a normative non-intentional account of concepts and conceptual content will have to re-invent an entire theory of behaviour and social practices which recognises actors, punishments and rewards, victims and perpetrators, and explains their actions without relying upon the attributions of beliefs and desires. It is far from clear that there is any such theory to be had. While the special problems facing normative analyses of concepts are substantial, we shall not dwell upon them

here. Instead we shall stick to our focus on psychologism and epistemic access.

Clearly, if the relevant norms are the products of human practices, we must be able to distinguish actions and/or states complying with them from those which breach them. Indistinguishable objects/actions cannot play a different role in human practices and it is those practices which give rise, supposedly, to the relevant norms. It is less clear that we must also have epistemic access to these social norms themselves. Less clear because it might be thought that the relevant norms do not cash out into any positive account at any stage, that it is norms "all the way down". But this is to put norms (and hence concepts) beyond our epistemic reach altogether. The attractiveness of the Calvinist explanation of (conceptual) norms in terms of social practices was that it explained our ability to understand one another; to explain it in terms of norms based upon yet further norms renders such understanding incomprehensible.

The matter is not quite as clear cut as that however. There is scope for a final somewhat desperate defence of a normative analysis of concepts, for like the concepts they are intended to explain, norms may be taken as something we impose upon the external world rather than something to be found within it. (A conceptualist taking this stand might, in a Wittgensteinian mode, proclaim that the limits of my norms are the limits of your language[18]) But this simply replaces the problem of explaining epistemic access to to other people's concepts with that of explaining our epistemic access to other peoples norms. The same ploys we have explored in the investigation of concepts can be replayed in the investigation of the relevant norms. One might attempt to identify norms and values with the objects or types of objects posited in some positive theory, much as we thought to identify concepts with the elements of a positive psychological theory, or alternatively one may give a normative account of the relevant norms. In this way the norms may be heaped upon norms, and the shift to positive theory and the connection to the evidence justifying our knowledge of other people's concepts and their knowledge of ours indefinitely deferred.

Still our norms and hence our concepts must cash out as the entities of some positive theory (presumably psychology) at some point, for it is only positive theories which make positive empirical predictions, and only through this connection with the empirical observations which are the sole evidence for them that we are justified in believing theories at all. Norms that depend solely upon other norms "all the way down" leave our capacity to know of other people's norms (a fortiori our capacity to know ourselves to be judged by the same linguistic standards that we judge ourselves, and hence understood) inexplicable.

We began our investigation of psychologism intending to examine and eval-

---

[18]This would be completely contrary to Wittgenstein's somewhat mystical attitude towards ethics, but that is another matter.

uate the two main lines of argument against it, one based upon the premise that our positive knowledge of concepts was unlike that arrived at as a result of empirical theory, and the other on the premise that concepts were posits of a normative theory. We have found these criticisms not to be completely without merit. It is true that the results of our empirical investigations cannot justify knowledge of modal truths, and in particular that psychological generalisations cannot justify believing a truth to be necessary. It is a contingent matter that creatures on earth have the psychological structures they do and contingencies cannot justify belief in necessities. But this does not show that concepts cannot be identified with psychological posits, for while it may be granted that that each of us knows her own concepts with a certainty that cannot be (justifiably) matched by our knowledge of empirical psychology, that does not show that our concepts cannot also be known through empirical means. We have perhaps been uncharitable in calling this last weak thesis "psychologism", for surely some of psychologism's critics have only argued against the psychological justification of conceptual truths; call it what one may, the first line of criticism we have considered not give any reason to suppose that the identification of concepts with psychological posits is incompatible with a conceptualist explanation of our knowledge of modal truths.

Equally, while it may well be true that concepts are fundamentally normative we have seen that this does not imply that concepts are not psychological entities. Certainly common-sense has it that people occasionally do believe impossible things and fail to believe some necessary ones, but the critic's argument from this — that since the laws of this normative system are sometimes breached while the laws of true empirical theories are never breached, concepts cannot be identical with the objects posited in any true empirical theory — is invalid. The counter-argument against this criticism went beyond merely showing that psychologism is compatible with a conceptualist account of modality, for the principle that human-made norms must be humanly evaluable (i.e. that compliance with the relevant norms be empirically determinable) upon which this counter-argument relied can be met by the supposition that normative status might be human projections that did not resolve into positive elements at any level — that it was norms "all the way down". In arguing that this supposition was false we did not propose any substantial ethical theory, nor attempt to delineate the empirical evidence for something being an ethical agent nor the evidence of its forming ethical judgements of other objects. Instead of any such constructive argument we relied upon the common-sense principle that each of us does sometimes know our assertions to have been understood (to have been taken as meaning what we took them as meaning) together with our empiricist principles about the evidence for this knowledge to imply that there must be

some cashing out of ethical theories and normative statuses in positive terms at some level or other without giving any positive characterisation of how norms were cashed out. This argument thus shows that if conceptualism is true then psychologism, or at least the identification of concepts with the objects of some positive theory or other, must be true. The argument for the compatibility of psychologism with a conceptualist account of modality was thus that this conceptualism (together with some empiricist principles and the common-sense principle that we sometimes know what we are talking about) implied psychologism, and hence must be compatible with it if conceptualism is true at all.

The epistemic argument for psychologism is indeed obvious, and if it were not for the very bad reputation that psychologism sustains as a result of the criticism of philosophers such as Husserl our extended defence of it would have been misplaced. Hopefully, our defence of psychologism has made clear that it is not merely a plausible doctrine (at least on our very modest interpretation of it) but is implied by any plausibly empirical conceptual/linguistic account of modality. Indeed this provides one of the few common points between the varieties of conceptualism which we have mentioned. Concepts have variously been taken to be the elements of beliefs or assertions, positive or normative entities, the product of public convention or private intention, but through all these variations the key requirement of evidence persists. If there is no empirical distinction to be drawn between conceptual truths and other more happenstance truths then conceptualism does not merely lack the power to quash skepticism, it lacks all plausibility. Epistemic considerations drove us, and many analytic philosophers before us, to look for an internal explanation of modality in terms of conceivability or conceptual/linguistic truth. Necessary truths were supposed to be true by virtue of meaning alone, and meanings, being either a product of our social conventions or ideas encountered directly in the imagination were unlike the essential properties of Aristotelian ontology, were something about ourselves to which we had epistemic access.

Thus far we have seen that conceptualism (together with empiricism) implies psychologism, or at least that there is empirical evidence for the conceptual structure of our thoughts and assertions, in the next chapter we investigate whether conceptualism is true.

Chapter 4

Quine's arguments against analyticity

> "The only evidence Quine produced to support this remarkable claim [that synonymy is hopelessly vague] was that he, Quine, could not clarify the notion in a few pages"
>
> — Putnam, "Two Dogmas revisited"
>
> "Quine's 'jump' from the claim that past attempts to establish the [analytic-synthetic] distinction faces serious difficulties to the claim that the distinction is (in principle) an untenable dogma is obviously not a step in deductive reasoning"
>
> — Gila Sher, "Is there a place for philosophy in Quine's theory"

The common view of Quine's "Two Dogmas" is that it does not contain any solid argument against the coherence of a distinction between analytic or conceptual truths and synthetic or non-conceptual truths distinction. Instead, Quine is thought to present a range of difficulties for traditional semantics reliant on the notions of synonymy and analyticity, none of which are shown to be insuperable, and then suggest an alternative which does not face the same problems. One of the purposes of this section is to show that this criticism is mistaken — that while "Two Dogmas" does contain some transparently worthless arguments (and others have been mistakenly attributed to it) it also contains the suggestion of one very good argument for the incoherence of analyticity which is both good and widely unacknowledged.

The other key text in which Quine urges rejection of the analytic-synthetic distinction is *Word and Object*. Unlike "Two Dogmas" it is generally agreed that *Word and Object* contains a positive argument for rejecting the analytic-synthetic distinction, which proceeds via the indeterminacy of translation and is

quite distinct from that (if any) of "Two Dogmas". Criticism of this argument tends to focus on the charge that this indeterminacy is an artefact of Quine's underestimation of the evidence upon which translation may be based. Another purpose of this section is to show that this criticism is also misplaced; that while Quine's behaviourism is mistaken, the main argument of *Word and Object* does not depend upon it nor upon any undue restriction of the range of evidence upon which translation is based, and indeed that it is much the same argument as the good argument of "Two Dogmas".

The main purpose of this section however, is to present a good argument for the epistemic inaccessibility of analyticity. It would appear that fulfilment of this purpose precludes fulfilling the two just mentioned, for if this argument's conclusion is true then there is no clear distinction between invention and exegesis; for that distinction is simply another aspect of the analytic-synthetic distinction. But the appearance is partly illusory, and the inaccessibility of the analytic-synthetic distinction does not preclude credit being given where it is due. For even those most skeptical of an analytic-synthetic distinction admit a distinction between good and bad translation. Though purpose of this section is not exegesis of Quine's arguments, I do hold that the arguments presented here are reasonable translations of those in "Two Dogmas" and *Word and Object*, whereas many of the arguments that have been attributed to those works are not.

Before attempting to explicate Quine's good argument for the epistemic inaccessibility of analyticity, I propose to consider some of the rather silly arguments for rejecting the analytic-synthetic distinction that have been attributed to him and in particular to "Two Dogmas". The purpose of this is not criticism for its own sake; belief in these misleading interpretations is so widespread in the philosophical community and so deeply entrenched that it occludes Quine's good arguments and the conclusions to which they tend. We will be in a better position to perceive the good argument those works contain once our minds are clear of these erroneous interpretations.

4.1   Circularity

The argument most commonly attributed to "Two Dogmas" scarcely deserves to be called an argument at all. This argument is that since the definitions of "analyticity" that Quine considers in "Two Dogmas" are circular then analyticity is incoherent. This is the argument Gila Sher and Hilary Putnam criticise in the quotations opening this chapter. Unlike some interpretations we shall consider it does have the right conclusion. But apart from this its chief virtue is that it is so transparently bad that no-one believes it, unfortunately this has

not prevented many commentators from attributing it to Quine.

It is obvious that the defectiveness of one (or several) definitions does not show its definiendum to be incoherent. Both "Large machine for crushing cars" and "cat" are woefully defective definitions of "cat" , but that throws no doubt whatsoever on either our epistemic access to cats or the coherence of "cat". Even the absence of *any* acceptable known definitions is not fatal to significance. One of the most famous figures in western philosophy spent his lifetime unsuccessfully looking for people who could define all sorts of perfectly significant terms such as "good" and "knows". For the argument from circularity to carry any substantial weight it would need to demonstrate that *every* definition of "analytic" was circular, and that cannot be established by considering any handful of examples.

Though the argument from examples of circularity is clearly worthless, it is also clear Quine does spend the larger part of "Two Dogmas" arguing that particular explications of "analytic" are circular. (Of *Two Dogmas'* six sections, fully three and two halves are devoted to arguing that explications of "analytic" are circular, from halfway through §1 till halfway through §5). Naturally one wonders if the critics are correct and the main argument (by word count at least) of "Two Dogmas" is worthless. The saving point is that the circularity of the definitions considered is not offered as sufficient reason to believe analyticity is incoherent; that conclusion is to be justified by the accompanying positive argument from holism. The failure of these explications sets a puzzle — its rhetorical purpose is to rouse dogmatic sleepers — the answer to the puzzle is provided by Quine's positive countersuggestion that "our statements about the external world face the tribunal of sense experience not individually but only as a corporate body."[1].

Though the arguments for the circularity of particular definitions of "analytic" serve a largely rhetorical purpose, they do not lack cognitive weight altogether. If on first reading these arguments serve primarily to open up for discussion the possibility that "analytic" is incoherent, in hindsight they also serve to reconcile the wide-spread use of "analytic" in philosophical circles with its incoherence. At first glance our capacity to distinguish competent from incompetent use of "analytic", the widespread consensus as to which statements to project it as true of, its role in explicating necessity and even our capacity to appeal to it in such explications, all suggest that "analytic" is perfectly meaningful. The arguments for circularity show that these trappings do not assure significance. A circle of definitions sets up a class of discriminable definitional truths and allows incompetent use of terms defined in it (which contravenes the relevant definitions) to be distinguished from competent use (which does not), but unless this circle touches upon statements which play a significant role in

---

[1](Quine, 1960, p. 41)

the generation of empirical predictions, the truth-values of the statements containing these terms are epistemically inaccessible and, according to Quine, those statements are therefore nonsense.

While circularity arguments cannot establish the epistemic inaccessibility of analyticity, they are the main form of attack upon proposed definitions of "analytic" and serve to show what is wrong with apparently plausible accounts of analyticity and the semantic theories in which it features. They thus have an importance proportional to the plausibility and popularity of the accounts they refute.

One *extremely* plausible and popular approach to explicating "analytic" is to postulate a canonical notation in which statements' fundamental semantic structure is explicit, and then to define analyticity in terms of the formal properties of statements in this notation. The variety of linguistic/philosophical theories which have taken such an approach to analyticity is breathtaking and the form of statements' canonical renditions have been variously called their "logical form", grammatical form", "deep structure", "conceptual structure" or "semantic representation". In philosophy this approach is most closely associated with logical atomism.

Logical atomism's canonical notation is that of truth-functional logic (or, more plausibly, quantificational logic[2]), and it is just the statements rendered as logically valid in this canonical notation that are taken to be analytic; but any of the host of philosophies which take concepts with full ontological seriousness and rely upon them to resolve questions of necessity, analyticity and synonymy are equal exemplars of the approach. Such philosophies are still common, though nowadays they are highly disputed, and though common-sense is at least superficially committed to a trinity of facts, thoughts, and statements (in which synonymous statements express identical thoughts) it is far from clear how much of these common-sense theories may be redeemed by sympathetic paraphrasis and how much must be rejected as false, but in linguistics and cognitive science the postulation of a "semantic level" (or semantic markers or semantic features) in which synonyms have identical representations is so ubiquitous and so little questioned that it is part of the orthodoxy.

Logical atomism's programme was at least clear. Belief in the necessity of non-logical truths was to be justified solely by producing their logical forms (by means of conceptual analysis), and then showing the resulting form to be logically valid. The production of such justifications and hence of the knowledge of necessary truths was to be the special task of philosophy, whereas science's task was to produce knowledge of contingent truths. The flaw in this scheme, given that it is to be one of circularity, may already be obvious to the reader.

---

[2]The difference is between Wittgenstein's and Russell's versions of logical atomism.

Both the original sentence and its canonical translations[3] must be synonymous if the validity of the latter is to guarantee the analyticity of the former. But the two are synonyms just in case the bi-conditional composed of the original statement and its canonical version flanking "if and only if" (e.g. "Bachelors are unmarried if and only if unmarried men are unmarried") is itself analytic. If analysis was called for at all, and the original statement not a logical truth as it stood, then neither will this bi-conditional be a logical truth, and its analyticity can only be revealed by further conceptual analysis in turn. But the synonymy of this bi-conditional and its analysand amounts to the analyticity of yet another bi-conditional which, though more complex, yet contains the same unrefined clause (the original sentence converted to a sentential clause) as the first. No progress is made from this point and an infinite regress ensues. At least some truths must be discernibly analytic without being resolved into their component atoms if any but logical truths are to be known to be analytic. But if some truths can be known without such decomposition why not all? The mysteries of conceptual analysis preclude a clear answer to this question.

This problem was noted at the time and even titled "the paradox of analysis" but, as its title suggests, it was not taken as a serious criticism of logical atomism. It was easily cast (and dismissed) as akin to Meno's so called paradox of knowledge and like it taken to be an insubstantial piece of sophistry.

In linguistics the situation is less clear cut, empirical concerns about how our brains are related to assertions and the nature of mental representations intrude, but the very identification of the "semantic level" relies upon the ability to identify synonyms prior to comparing their representations at this level. At the very least each statements must be determinably synonymous with their representations at the "semantic level" independently of the other. But if statements' meanings can be determined prior to their semantic level representations, it undermines much of the reason for supposing that there is such a level in the first place.

The objection applies to more modern semantical approaches too, of which prototype semantics is just one. No advance is made by declaring general terms synonymous just in case they have the same prototype, for we must first determine that each is synonymous with the composite general term formed by application of "is similar to" to a term true of the prototype; i.e. whether 'bird' is synonymous with 'is similar to a robin', or "opinionated grandmother" is synonymous with 'is similar to Fodor's granny'. The difficulty is better recognised by some prototype theorist than it was by most early twentieth century analytic philosophers: "Prototypes do not constitute a theory of representation

---

[3]Or some intermediate semi-analysand. Both Russell and Wittgenstein allowed that analysis might be a partial matter.

of categories. Although we have suggested that it would be reasonable ... if categories were represented by prototypes ... such a statement remains an unspecified formula until it is made concrete by inclusion in some specific theory of representation."[4]. And of course such concretion cannot simply take the form of more prototype theory. Prototypicality, intersubjective consistency in judgements of how well individuals exemplify the categories they fall under, is not the basis of a semantic theory, nor can it found judgements of synonymy, but is instead a further piece of evidence for scientific linguistics to explain.

Still, though this argument against defining analyticity in terms of the formal properties of statements in canonical notation cuts a swathe through semantic theories, like all of the circularity arguments it does not show that there is no way of defining "analytic" nor that the linguist's "semantic level" does not exist.

4.2   The first argument from Universal Revisability

Even though the argument from circularity is a non-sequitur, it is not the worst argument frequently attributed to "Two Dogmas". That title must go to an argument from universal revisability. Putnam has dubbed this argument "the 'historical' argument" and it begins appositely enough on a historical note: Again and again in the history of science scientists reject sentences their predecessors had been taken to express unrevisable truths. No longer do we believe the certainties of yesteryear — that space is Euclidean, the passage of time absolute, and causation deterministic — each taken in their time not as mere empirical generalisations but as necessary truths, the world's joints at which Plato enjoined us to carve. From this premise the argument proceeds by a kind of induction to the conclusion that any of the sentences we currently hold to express necessary truths may be found to express falsehoods in future. It might even turn out best to revise quantificational logic itself, perhaps in the face of discoveries in quantum mechanics, and thus deny logical truths. The susceptibility to revision is sufficient to show that these truths are neither necessary nor analytic. Necessity, according to proponents of this argument would appear to be an illusion engendered by over-confidence.

The simplicity of the argument doubtless provides a great deal of its appeal, but the inductive argument is not merely bad, it is irrelevant. That a sentence expresses an analytic or conceptual truth was never supposed to imply that sentence's immunity to re-evaluation, but only that such re-evaluation would required changing the belief expressed by that sentence. Universal revisability

---

[4](Rosch, 1978)

alone by itself does not establish that there are no analyticities, it is universal revisability *without change of meaning* that is required.

4.3 The second argument from Universal Revisability

The second argument from universal revisability has one virtue the first lacks — it was actually advanced by Quine in "Two Dogmas". Even better, its premises justify its conclusion. Its flaw is that it takes a far stronger form of holism as a premise than the mere failure of reductionism. No reason is given in "Two Dogmas" to suppose that the falsehood of the first of the eponymous dogmas implies the falsehood of the second.

The holism Quine relies upon in "Two Dogmas" is stronger than the mere failure of reductionism in two respects. Here we are concerned with the first, which the argument from universal revisability depends upon, and shall defer consideration of the second until we consider the argument from universal maintainability.

Quine's holism is presented in "Two Dogmas" as the doctrine that no statement has empirical content in isolation. This strong holism implies universal revisability; if no statement has empirical content in isolation, then a fortiori no statement has vacuous empirical content in isolation, and no statement is verified come what may. For any statement there must at least be some conditions (and some background assumptions) which are compatible with its negation. This universal revisability implied is universal revisability without change of meaning, and hence implies that there are no analytic truths, for if any sentence can be denied without change of meaning then no sentence is true by virtue of meaning alone.

This strong holism not logically implied by the negation of reductionism. In point of logic alone that statement (viz., "It is not the case that each meaningful statement is equivalent to some logical construct upon terms which refer to immediate experience"[5]) only implies that *some* statements have no sensory equivalents, not that *none* do. But yet the inference to the stronger claim might be justified upon some more subtle explication of reductionism, or with the support of additional premises. The question is to be settled either by further argument or a counter-example.

No such argument is to be found in "Two Dogmas" and by the time of *Word and Object* Quine divides statements [6] (somewhat nebulously) into observa-

---

[5]See the opening paragraph of "Two Dogmas" for the description of reductionism on which this is based.

[6]More strictly *occasion* statements

tional and non-observational, of which the former have their empirical content
in complete isolation from other statements and the latter (Quine appears to
tacitly assume) have no empirical content in isolation at all. Still that does not
address the question of whether holism somehow implies that there are no state-
ments which cannot be meaningfully denied. While Quine does not present any
argument for that implication however, he does unwittingly provide a counter
example.

According to Quine his holism issues "essentially from Carnap's doctrine
of the physical world in the Aufbau". He explains that the language of the
Aufbau is holistic because it includes statements of the form "Quality q is at
$x;y;z;t$" which lack empirical content in isolation. The truth values given to these
statements cannot be determined in isolation but only apportioned "in such a
way as to maximize and minimize certain over-all features"[7]. The canons of the
Aufbau "counsel us in the use of 'is at' but not in its elimination" — they do
not issue in translation of those sentences into terms referring solely to stimuli
or experience.

Clearly the language described in the Aufbau is holistic in the weak sense
that reductionism is not true of it. Despite this some statements are included
in every theory constructed according to Carnap's precepts. The evidence upon
which the Aufbau's constructions are based are coloured points arranged in a
two dimensional pseudo-visual field, but no matter how the points of this field
are coloured, any theory constructed in accordance with the Aufbau's canons
will always include these statements and not their negations. The theories
produced attribute colours to points of three dimensional space — hence the
"z" in "is at x;y;z;t" and fix the point from which this three dimensional space is
being viewed. The guiding idea is to postulate the most plausible arrangement
of objects and observer that would produce the specified arrangement of colours
across the observer's two-dimensional visual field. Compliance with Carnap's
canons never produces a theory attributing two distinct positions simultaneously
to the observers point of view, nor one containing a logical contradiction (unless
the specification of the given visual-field is itself contradictory). Apportioning
falsehood to the corresponding statements is never compatible with Carnap's
canons for all that the theories they produce are holistic ones. If the Aufbau's
canons are semantical rules, fixing the meaning of the terms whose use they
counsel, then the statement of those theories which may be rendered into English
as "The observer is at just one point in space at a time" is unrevisable without
change of meaning, or to put it another way, hence true by virtue of meaning
alone.

It is slightly over-stating the case to claim this as a counter-example to the

---

[7](Quine, 1953, p. 40)

claim that every statement in any holistic language can be revised, for that claim assumes we can make sense of the claim that the Aufbau's precepts are (or can be taken as) semantical rules which fix the meaning of the terms of the language they describe. If that claim is nonsense then it would equally make no sense to insist that the breach of these precepts, in particular the denial of one of the statements they required to be taken as true, involved a change of meaning. But the incoherence of the notion of a semantic rule, and of an analytic truth was supposed to be a consequence of Quine's argument, not an additional premise. But the system of the Aufbau does show us that there are systems of rules which are undoubtedly holistic (even if doubtfully semantic) and which yet require that some sentences are not to be revised. The failure of reductionism does not imply that there are no unrevisable statements, and Quine's argument via universal revisability is worthless.

The second argument from universal revisability is no critic's invention. Quine advances it when he argues that it is "folly to seek a boundary between synthetic statements, which hold contingently on experience, and analytic statements, which hold true come what may" since "Any statement may be held true come what may, if we make drastic enough adjustments elsewhere in the system. Even a statement very close to the periphery can be held true in the face of recalcitrant experience by pleading hallucination, or by amending certain statements of the kind called logical laws. Conversely, by the same token, no statement is immune to revision."[8]. Though commentators are thus not mistaken in attributing this argument such attributions have had the effect of spreading two very serious errors.

The first is that holism implies universal revisability. We have seen that this is not so, but this implication is become so widely assumed that some commentators take the two claims to be equivalent — "Quine invokes the Duhemian under-determination argument. If any relatively high-level theoretical claim can be held true come what may, then Carnap's philosophical moves here do not work."[9], and even where they do not the inference goes unquestioned.

The most pernicious historical effect of this argument from universal revisability however, has been to spread the illusion that the intended conclusion of "Two Dogmas" is that there happen not to be any analytic truths, rather than that analyticity is epistemically inaccessible and hence "analytic" incoherent (by empiricist standards). Even though Quine's proposing the argument from universal revisability is genuinely confusing, it is surprising that this confusion is so widespread despite Quine's unequivocal statements of his intention in "Two

---

[8](1953, p. 43) It is pedantic to point out Quine's confusion of use and mention, but it is perhaps justified by his reputation as a stickler on the point.

[9]Alan Richardson, Two Dogmas about Logical Empiricism, p.153, see also p.157

Dogmas" itself. For example — "That there is such a distinction [between analytic and synthetic] to be drawn at all is an unempirical dogma of empiricists, a metaphysical article of faith."[10]. One would have thought such declarations would make the matter clear. That they have not is due in no small part to Fodor, who is at least clear in his mis-interpretation. "When Quine says 'No a/s,' he presumably means 'No analytic sentences.' "[11]. The argument from universal revisability is a key part of this mis-interpretation "To the extent that [Quine's] arguments are relevant to the more general notion of analyticity, that's because the potential revisability of a statement shows that it isn't analytic; and many philosophers hold that this potential spans the entire language. Whether they are right, however, is an empirical question. So the issue of what analyticities there are turns on a variety of unresolved empirical matters"[12].

By contrast the stated point of "Two Dogmas", and the conclusion for which Quine suggests the best argument, is not that empirical investigation reveals that no statements are analytic, but that empiricism implies that no empirical investigation can reveal analytic truths at all. The arguments from universal revisability not only fail to provide any reason for believing that there are no analytic sentences, but their supposition that there is empirical evidence incompatible with the existence of analytic truths is directly contrary to the conclusion of Quine's central (and best) argument which is that the analytic-synthetic distinction is epistemically inaccessible. From this principle, together with a general sense of empiricism, he infers that the analytic-synthetic distinction is incoherent.

4.4   The argument via Universal maintainability

In addition to the good argument for the epistemic inaccessibility of analytic truth, Quine gives a very bad argument. And though it has the right conclusion — that analytic truth is epistemically inaccessible — it infers this conclusion from a principle that is not itself implied by the failure of reductionism — namely that every statement can be held true come what may.

The argument from this principle of universal maintainability to the epistemic inaccessibility is neither so immediate or so obvious as the argument from universal revisability to the absence of analytic truths. Superficially, the existence of non-analytic truths held true come what may only appears to show that "statement competent speakers hold true come what may" is not an adequate definition of "analytic". As we noted in consideration of the arguments

---

[10](Quine, 1953, p. 37)
[11](Fodor, 1992a, p. 25)
[12](Fodor, 1998, p. 21)

from circularity, the disqualification of *a* definition of "analytic" does not show that there is no good definition to be had. We shall consider this step in the argument more closely in connection with Quine's best argument, for now let us seek the more the glaring fault in this one.

Like the first step of the argument via universal revisability, the argument for universal maintainability, insofar as one is given at all, proceeds from a premise that is essentially stronger than the mere negation of reductionism. Again, logically the negation of reductionism implies only that *some* statements cannot be translated into purely sensory terms, and even if a statement's being so untranslatable implied it must also be maintainable under any circumstances (given radical enough re-evaluation of other statements), the two together would imply only that *some* statements could be maintained come what may. In this case however, the broader argument (for epistemic inaccessibility) does not depend upon the over-generalisation. That there are just some non-analytic truths held true come what may is sufficient to discredit "held true come what may" as a definition for "analytic". In this case the fault is minor and we shall not dwell on it further.

The worst problem is that statements that have infirmation conditions in isolation, and hence cannot be held true come what may, can yet be irreducible to sensory terms. The failure of reductionism thus does not imply that any statements can be held true come what may. The tremendous significance of the defect in Quine's philosophy of language on this point contrary, and its serious and enduring effects on modern philosophy merit emphasis. This error is not washed away in Quine's shift of opinion between "Two Dogmas" and *Word and Object* with the admission of observation sentences. Though these sentences are an admitted exception to both universal maintainability and universal revisability , Quine continues to underestimate the subtlety and complexity with which individual statements may be connected with observation and yet admit of their own individual empirical content.

The key error is a failure to recognise that there are statements of an intermediate sort which, while not living up to the strict standards set for observational statements nor admitting of being ordered by their "observationality" yet have some empirical content in isolation. Imagine for a moment that there are observation statements and that "Here is a rabbit" is one of them, and "Here is an electron" is not; now consider the empirical content of the conjunction of the two "Here is a rabbit and here is an electron". Clearly the negative stimulus meaning of this statement (the class of stimuli prompting dissent from it in competent speakers) must be a superset of the negative stimulus meaning of "Here is a rabbit", but of course if "Here is a electron" doesn't have any stable positive stimulus meaning (regardless of idiosyncratic background assumptions)

then neither does the conjunction of which it is a part. The conjunction obviously has some empirical content in isolation, all the negative content of "Here is a rabbit" and more, but equally obviously it is irreducible to sensory terms.

Once it is realised that statements can have some empirical content in isolation, can be outright confirmed or infirmed by some stimuli sequences regardless of background assumptions, and yet *not* be translatable into sensory terms, *not* have completely fixed empirical content in isolation, the whole idea that theories must be connected to their empirical content via statements having fixed empirical content completely in isolation (or even to a degree of isolation that can be graded along a linear scale) can be dismissed. We can recognise that a statement can have a stable core of its positive stimulus meaning whose members always evoke assent and a penumbra of stimuli sequences which may or may not evoke assent depending upon background assumptions, so too, mutatis mutandis, for negative stimulus meaning. Having recognised that there are such statements the whole need for observation statements (in Quine's strict sense) evaporates. So long as a statement has a reliable and discernible core of stimuli sequences which always evoke dissent to it, it can serve as both an entering wedge for translation and a connection between theories and their empirical contents. It should have been obvious that there were no observational statements in Quine's sense, that even the most observational statement must be connected with theory and assent to it evocable through those connections.

The distinction between observational and non-observational statements has something of both the vices attributed to the analytic/synthetic distinction. Taken as an extreme dichotomy between statements whose stimulus meanings vary not one whit with varying background assumptions, the distinction is perfectly clear but vacuous — it is clear what an observational statement would be like, its just that there are no observational statements in this sense. But taken as a graded notion observationality is incoherent for we have no standard by which to compare the observationality of different sentences, indeed once one begins to question the degree of observationality of various sentences the ridiculousness of the notion shines through. Which is more observational "Here is a rabbit and here is a bachelor" or "Here is a rabbit and something that is not a bachelor"? Are we moved by overwhelming preponderance of obvious non-bachelors over obvious bachelors to claim that former is the less observational? Let us vary the example — is "Red" less observational than "Green" because of its theoretical connections with litmus paper — it can be evoked from a properly trained chemistry student by just some examination text about litmus paper and chemical solutions — or is the latter less observational because of the convention of colouring ground wires green in Australian domestic wiring. Are we to be moved by the fact that chemistry is more "theoretical" than the study

of domestic electrical circuitry or by the fact that there are more electricians than chemists? Surely the question is silly for there is an open-ended variety of ways of detecting green clad wires and red litmus paper. The closest we can come to making sense of a degree of observationality is to take it as the ratio of the stimuli sequences in the stable core (whether positive or negative) to the fickle penumbra; but this device will not do the job, for in all except trivial cases there will be a countably infinite number of stimuli sequences in each.

In *Word and Object* Quine's oversight of irreducible but individually con/infirmable statements is fuelled by his bizarre enthusiasm for stimulus-response conditioning as the psycho-semantic mechanism of choice; but its origins lie in his conflation of two very different senses in which statements are said to "have empirical content in isolation" in Two Dogmas. In the weak sense "has empirical content in isolation" is roughly equivalent to "is translatable into sensory terminology", and in this sense of course holism does imply that some statements do not have empirical content in isolation. But the stronger sense in which statements face the tribunal of sense experience "only as a corporate body" is not implied by holism. Indeed far from implying that it is nonsense to talk of the confirmation and infirmation of individual statements at all, the central positive doctrine of holism — that whole theories do have (empirical) content (/confirmation/infirmation/compatibility conditions depending upon the particular brand of holism) — implies that it is perfectly coherent to talk of the content of individual statements, for the content of individual statements can easily be defined in terms of the content of whole theories. An individual sentence may naturally be called compatible with a given stimuli sequence just in case it is a member of some theory compatible with that sequence and incompatible otherwise.

This definition is couched in terms of of compatibility and stimuli sequences (the notions explicated in part 1), but it may as easily be cast in terms of confirmation — an individual statement is naturally said to be confirmed by a given stimuli sequence if it is a member of every theory confirmed by that sequence. Whatever one takes the fundamental semantic translation guiding relationship to be, so long as it is a relationship between evidence and whole theories, a corresponding relationship between that evidence and individual statements can be defined. Holism is distinguished not by the lack of such relationships but by the combinations in which it may obtain. Reductionism, on this view, is the doctrine that the content of any whole theory is simply the intersection or super-position of the contents of each of the individual statements composing it. Holism, by contrast is the doctrine that statements contribute more than their compatibility conditions to the compatibility conditions of the theories they compose — that sometimes the addition of a statement compatible with

a given stimulus sequence to a theory compatible with that same sequence (together with all their combined logical implicants) produces a theory which is not compatible with that sequence. Quine's strong holism — the doctrine that no individual statement admits of con/infirmation in isolation — implies that that every statement and every statement's negation is compatible with any stimulus sequence, and though this is consistent with holism it is not implied by it.

Granted a holistic language which includes at least one projectibly true statement which nonetheless admits of infirmation in isolation it is easy to construct a holistic language of which no statement is without its own individual empirical content. Consider a language such as English which is sufficiently sophisticated to express currently plausible scientific theories and which includes the statement "There are no flying pigs", and let us choose as our contentive statement "There is not a flying pig present". Clearly our current best theories predict that this statement is true on every occasion, and equally clearly, despite this there are stimuli sequences in which flying pigs feature so conspicuously as to prompt any competent English speaker exposed to them to deny this statement. It is, of course, part of the empirical content of our current theories that no observer will or has ever been exposed to such stimuli, but such stimuli would infirm this statement in isolation.

Now consider a language having the same lexicon as English and the same standards of grammatical correctness — each English sentence is a sentence of this novel language and vice-versa — but in which sentences express (are translatable as) the conjunction of the homophonic English statement and the English statement "There is not a flying pig present". Thus in this new language the sentence "Electrons have negative charge" expresses what "Electrons have negative charge and there is not a flying pig present" expresses in English Now even by Quinean standards the homophonic mapping between this language and English is not a good translation, for it does not preserve stimulus meanings, it does however preserve the empirical content of any large slab of English theorising sufficient to imply that there are no flying pigs. This language is thus capable of expressing theories with empirical content that cannot be possessed by any theory composed solely of statements that can be translated into sensory terms, and is hence a holistic language. But every statement of this language is infirmable in isolation, therefore holism does not imply that any statements maintainability a fortiori that all are. It might be argued that this imagined language is defective on some other grounds, that for some reason or other languages cannot be composed of statements which can all be infirmed in isolation, but holism does not carry any such implication and there might be some other reason for supposing that some statements of any holistic language

must be maintainable under any circumstances; but the argument we have been considering was to be from holism to universal maintainability and thence to the epistemic inaccessibility, and this argument is no good as it stands since we have no reason to suppose that the premise is not true of some languages of which the intermediate conclusion is false.

4.5 The good argument

The common flaw in Quine's bad arguments for the epistemic inaccessibility of analytic truth is the assumption that we can identify which (holistic) language a speaker is using, but that there are no analytic truths in such languages Hence Quine's claim that in a holistic language "Any statement can be held true come what may ...[and] Conversely by the same token no statement is immune to revision."[13]. As we have seen, this claim is false, in some holistic languages there are statements that must be held true come what may and statements that must be rejected in the face of some evidence. So long as a distinction can be drawn between the dual components of language and extralinguistic experience, even if only on a global scale, a distinction may be drawn between those sentences true by virtue of language alone (those which are members of every theory expressible in that language which are not incompatible with every stimuli sequence) and the others.

The good argument is that we cannot identify just one language a speaker is using — that for any (logically adequate) language in which a sentence expresses a non-logical analytic truth there is another language whose speakers would be (or would ought to be) indistinguishable in point of their semantically relevant dispositions, but in which that sentence expresses a universally held belief rather than an analytic truth; and (less importantly) that for any language containing a sentence which expresses a universally held non-analytic belief there is another language whose speakers would have the same relevant dispositions in which that sentence expresses an analytic truth. What needs to be shown is that the speakers of such alternative languages cannot be distinguished on any semantically relevant grounds and that such languages exist.

Historically, attempts to demonstrate the limits of the evidence relevant to linguistic affiliation/competence have centred on radical translation. A radical translator knows only of subject's public verbal responses to publicly determinable stimuli, and the key claim is that on this meagre evidence radical translators can produce translations upon which no further evidential supplement can improve. The only justification Quine offers for this claim is that this

---

[13](Quine, 1960, p. 43)

is all the evidence anyone has to go on in learning their first language. *Word and Object* begins with the observation that "Language is a social art. In acquiring it we have to depend entirely on intersubjectively available cues as to what to say and when. Hence there is no justification for collating linguistic meanings unless in terms of men's dispositions to respond overtly to socially observable stimulations."[14]. The point is even more explicit in *Epistemology Naturalised* where Quine declares that "What I have said of infant learning applies equally to the linguist's learning a new language in the field. If the linguist does not lean upon related languages for which there are previously accepted translation practices, then obviously he has no data but the concomitances of native utterance and observable stimulus situation."[15] The claim then, is that public utterances in response to publicly determinable stimuli are all the evidence that children or linguists have to go on in learning how to speak/translate novel languages, and since children and linguists succeed in these projects on this meagre diet, it must be sufficient to supply their wants, and hence sufficient to determine which language the child learns to speak and which translations of his utterances into other tongues are good.

This defence of radical translation is too hasty and nearly entirely misplaced. It is not clear, after all, just what the evidence children learn their language from is, nor even that children do *know* which language they speak or that their peers speak it also. It may be that children's diet of evidence is richer than there observation of the conditions under which their competent linguistic compatriots assent to statements. Even if it is not, the matter does not rest there. Some psychologists tell us[16] that children do not take linguistic regularities to be distinct from more fundamental nomological regularities, and take their peers speaking of their native tongue to be as natural and inevitable as the falling of massy objects to the ground. Of course the children are wrong, for there is a world of alternative languages of which they are unaware. One may well suppose that just as children simply assume that their language is the only one, so too they simply assume (and do not know) that it is spoken by their peers. Indeed many psychologists insist that the principles upon which language learning depends are to some extent innate and hence (one supposes as a good empiricist) not justified by anything at all. Perhaps only those who can justify these innate beliefs, really know what language they speak whereas the rest of us just know how to speak it.

The strategic mistake in this style of defence is to attempt to justify the adequacy of radical translation's narrow evidential base directly on epistemic

---

[14](Quine, 1960, p.1)
[15](Quine, 1969, p. 81)
[16]I inherit the point from Fodor (see (Fodor, 1992a, p. 77)).

(i.e. empiricist) grounds. This is natural, we are used to using empiricism to circumscribe the available evidence and the claims that may be justified by it, indeed the current argument is, simply an application of this technique. But in this case we wish to show that the evidence available to radical translators is sufficient to good translations, and critical empiricism is simply misplaced. The best that can be done along these lines would be to show that some people whose evidence is thus circumscribed nonetheless succeed in learning languages or producing justified translation. But it almost as unclear if anyone succeeds in radical translation, as it is that radical translators' evidence is sufficient in the first place. The claim that linguistic competence is a matter of a dispositions to assent exercised by the presentation of evidence (and perhaps a prompting query) must be justified (in the first instance at least) by a positive account of what languages are rather than how we come to know them. It is not directly to empiricism, but to holism that we must turn if we are to show that radical translation admits of no improvement.

Thus far we have treated holism as little more than the negation of reductionism (and criticised Quine whenever he relied upon a stronger claim to support his arguments), but it is of course a positive doctrine viz., that meaning or significance is purely a matter of empirical content (confirming/infirming/compatible evidence) and that its primary bearers are whole theories[17]. Holism is the doctrine that it is whole theories that are compatible with/confirmed/infirmed by evidence (in the first instance at least, we must allow for the derivative sense in which individual statements may have such contents). A more sophisticated account would treat of degrees of confidence in disparate theories and degrees of con/infirmation, but such complications are a distraction from our current purpose, and we will do well to imagine that each person holds just one theory, and disregard the degree of confidence with which they hold it. Holding a theory, on this view at least, is simply a matter of holding the statements that comprise it true. Thus the only requirements of linguistic competence concern which statements one can permissibly hold true in the face of the passing show, and which passing show is compatible with the assertions of one's compatriots, and hence this is all that is relevant to translation and meaning.

This secures the argument *from holism*, on this point at least, but it may leave one wondering what reason there is to believe holism in the first place — why hold that theories have empirical content or that this empirical content is all the content there is. At this point Quine gestures towards an empirical justification — science reveals "there is no clairvoyance". But this is to trivialise conceptualism which was never committed to any form of extra sensory perception, but hopeful of a form of justification which did not depend upon

---

[17]The previous section should serve to clarify this last part of the claim.

any external influence at all. For our own part a more modest justification will do. Recall that we are investigating the charge that weak empiricism is self defeating, and it is already clear from our refutation of strong empiricism that if any form of empiricism is true then it must be one compatible with the failure of reductionism for we have already seen that the statements of our scientific theories (and the austere idioms described at the end of part one) cannot be reduced to sensory terms. Of the two stronger components of holism, the claims that empirical content is all the content there is and that whole theories have this content, the first is pretty much just a less clear statement of a semantic version of weak empiricism — so long as empirical contents are simply empirical evidence under another name we may cast our weak empiricism in parallel terms as the doctrine that empirical content is the only content that we can be justified in believing; and as for the second, provided that to believe something is to believe that it is true then statements (or the propositions they express) are the only candidates on offer. The austere idioms, in terms of which we defined weak empiricism as a doctrine concerning translation at the end of part one, are holistic languages

Having demarcated the evidence relevant to linguistic competence/compliance, we now need to show that this evidence is insufficient to determine linguistic affiliation uniquely. The entering wedge for this component of the argument is that holism implies theories are under-determined by the evidence — i.e. that there are incompatible assignments of truth values to the statements of any holistic language equally well fitting some ranges of total evidence. This much is readily shown, for if each course of experience had only a single compatible assignment of truth-values to the statements of some language, then each of those statements must have its own content strictly in isolation.

Though the theories in a holistic language are under-determined by the evidence, subjects' dispositions to hold statements true are not. Just as a subject cannot be disposed to both stop and go upon driving up to a red traffic light[18], so too a subject cannot be disposed to both hold and not hold true a statement that is a member of just some of the theories compatible with(/justified by) the exercising evidence; and to hold true all of the statements that are members of any of several compatible(/justified) theories is not to hold any of those theories (and hence not to be a competent speaker of the language in question), indeed it is not to hold any theory at all unless the resulting hodge-podge of statements happens to be closed under logical implication.

Though one cannot be disposed to both stop and go at red traffic lights one can be disposed to have a ninety percent chance of stopping at red lights and a 5 percent chance of driving through. In like fashion it might be thought

---

[18]Flesh out the exercising condition as one will.

that speakers might have propensities to hold each of the theories which have as their empirical content the (total) evidence presented to them, and in this fashion all of the proprieties of language use (all of the linguistic equation) might be manifest in competent speakers dispositions to hold sentences true. But the suggestion, though it may be true, is completely misplaced. The under-determination of evidence by theory does not require competent subjects to remain forever undecided between empirically equivalent theories, but shows instead that the choice is arbitrary and may be made with full confidence for there is no risk of going wrong — whatever theory a subject alights upon will be true and the others false; in part by virtue of the match of empirical fact and empirical content and in part by virtue of the subjects very choice which itself fixes the relationships between word and object.

Since competent speakers may be absolutely confident (or as confident as they are that their experiences match the empirical content of their theories) in just one of several empirically equivalent theories, so too they may be disposed to become confident in just one of several empirically equivalent theories when confronted by empirical evidence with which just they are compatible. The key point is that while theories expressed in holistic languages must be under-determined by the total evidence, the dispositions of competent speakers to hold theories in response to that evidence can be completely determinate. There can be an excess of valid ways of expressing empirical contents over and above those competent speakers are disposed to use. According to holism a language can be viewed as a mapping between whole theories and empirical contents. But while this mapping must be *many-to-one* in any holistic language, the mapping realised in a competent speakers dispositions to hold theories in response to their experiences may be *one-to-one*. Thus there can be linguistically proper ways of expressing some empirical contents which are not manifest in any competent speakers dispositions. To be a competent speaker it is enough to only hold theories which express the empirical contents with which one has been/expects to be exposed to *it does not require that one to have a disposition to hold every theory which expresses the empirical contents one expects to be/has been exposed to.*

If any many-to-one mapping from theories to empirical contents were a holis-tic language the indeterminacy of which language a speaker was using and the epistemic inaccessibility of analyticity would follow almost immediately For con-sider a (holistic) language L in which a sentence S expresses an analytic truth (i.e. is a member of all theories that have non-empty empirical contents). Com-petent speakers of L must assent to S come what may. If any mapping from classes of statements closed under logical implication to empirical contents cor-responded to a language, then there is another language L′ with the same logi-

cal terminology which in addition to mapping all the same classes of sentences (which are theories in both languages) to all the same empirical contents also maps some other classes of sentences (theories of L′) which do not include S to some of the empirical contents which can be expressed in the original language. The theories of L′ are even more under-determined by evidence than those of L for in L′ there are more ways of expressing the same empirical contents. More importantly the competent speakers of L of whom linguistic propriety requires them to reject any theory which does not contain S cannot be distinguished (on semantically relevant grounds) from speakers of L′ who merely choose to eschew theories which do not contain S in favour of those that do for they have exactly the same dispositions to assent to sentences in the face of experience. Nor is the point merely epistemic, for every competent speaker of L is also disposed to obey all the linguistic proprieties of L′ and hence is a speaker of L′ and every speaker of L′ who eschews theories which do not contain S obeys all the proprieties of L and is hence a competent speaker of L. Empiricism is required to restrict the range of relevant evidence, to preclude the postulation of some inscrutable (presumably mental) difference between the speakers of the two languages, but once that point and holism are granted then we must give up the idea that competent speakers are speakers of just one language.

The complication is that not every mapping from theories to contents is a plausible language. Some mappings are so hopelessly higgeldy-piggeldy as to defy human mastery. While it is arguable that there are such languages, and indeed that humans are competent speakers of them, our argument for the epistemic inaccessibility of analyticity would be easily dismissed if it relied upon them. The most common, and indeed the only plausible way of explaining the human capacity to understand and properly assert novel sentences is to posit some compositional structure within sentences — atoms which can be recombined. Holism is not incompatible with the existence of composable atoms, it may be true of languages comprised of sentences constructed from a finite number of fundamental terms according to recursive rules, and a mapping from theories to empirical contents which is a recursively enumerable function of the arrangement of the terms in those theories sentences. The atoms in these welcome accounts are words; holism precludes only the attributions of meanings to these atoms. Together holism and compositionality can be reconciled by a compositional account of statements' inferential connections (or rather the patterns of statements that competent speakers may hold true). Such approaches are easily confused with inferential role semantics, that ugly halfway house for those who have been swayed by holistic metaphors of webs, or arched columns or nets, or in some other way induced to think that meanings are fixed in part by inferential connections between statements, while yet resisting holism's central

claim enough to insist that the meanings of individual statements are composed in this way. But compositionality does not require atomistic semantics, for though these inferential connections are, together with the empirical contents of some finite number of statements, sufficient to fix the empirical contents of *whole theories* there is no reason to suppose they fix the meanings of individual terms or statements or indeed that there are any such meanings to be fixed.

It is thus not enough that for any language in which there are analytic truths i.e., which only maps theories containing the corresponding sentences to empirical contents, there is a more liberal mapping which as well as mapping exactly the all the contentive sentence classes of the original language to the same empirical contents also maps some other classes of sentences to empirical contents. The more liberal mapping may be dismissable as either beyond human mastery or intolerably ad hoc. The additional requirement of recursive compositionality can however be satisfied by language in which contentive classes are the closures of a recursive class of sentences under a recursive operation producing sentences from sentences, and their contents of each such class is the intersection (i.e. all the stimuli sequence compatible with all) of the empirical contents of truth functions of a finite class of sentences con/infirmable in isolation (but not necessarily reducible to sensory terms) and time specifications which are among its members.[19]

More precisely, the argument can be vindicated by the existence of liberal languages which have the following three features: they possess the standard logical terms of quantificational calculus, in them only recursively axiomatisable sets of sentences are contentive, and each has a finite and invariant class of independently con/infirmable statements such that the empirical content of each contentive class in that language is the intersection of the empirical contents of the members of the closure of the union of that invariant class and explicit temporal specifications (i.e. sentences of the form "It is n seconds since my birth") under the operations of truth-functional composition which it includes. Such a language could be learnt by creatures that were sensitive to their environment and the passage of time and could compute recursive functions (or establish norms satisfied by the computation of such functions), and the pattern of their learning could be reassuringly Quinean (though without the egregious behaviourism). In the very early stages of language learning stimulus-response

---

[19]The admission of time specifications echoes our admission of "temporally precedes in my lifetime" to the austere idiom of empiricism suggested at the end of part one. The lack of precise details of how incorporate predictions of future experience into the more or less Quinean scheme that I have adopted is one of the lingering defects of this work. The device gestured to here is to allow at least relative specifications of time (such as "It is 12,000 seconds since my birth") as independently contentive sentences, and thus to allow the content of theories to be fixed by the membership of such sentences as "Either it is not 12,000 second since my birth or I am seeing red".

conditioning is a near approximation of how the proper use of such indepen-
dently con/infirmable sentences might be taught, an approximation only for too
rigid or brittle conditioning might preclude later expansion of the stimuli elicit-
ing assent, to include conditions which do not universally elicit assent regardless
of background beliefs, (precisely because the subject holds some of the relevant
background beliefs). Unlike Quine's, the account suggested here is compatible
with the pre-dispositions of which direction to relax the observation meaning
of independently contentive sentences and the level at which language learning
can begin belying a sophisticated, ontologically substantive, and pre-linguistic
model of the environment. No position is taken here about how sophisticated
(or non-sensory) a subject's "pre-linguistic quality space"[20] may be.

Likewise, Quine's account of the course in which the apparatus of reference
is learnt (from predication, to relative clauses and abstract terms) is compatible
with, but not essential to, the premises of the current argument. Our assump-
tion is that humans are capable of learning to perform any recursive operations
on letter-sequences, and hence, since syntactic rules of logical implication are
recursive, capable of mastering quantificational logic. This assumption is not
indubitable, and is discussed further in connection with the unconceptualistic
defence of quantificational calculus, for the moment let us mark it down as an
additional premise of the argument, albeit one already relied upon in the main
argument for holism advanced in part one (viz., that only holistic languages can
capture recursively enumerable classes of stimuli/empirically contentive sen-
tences).

Languages whose non-trivially contentive classes (all the contentive classes
that are compatible with just some stimuli sequences) are recursively axiomati-
sable, and whose connection to empirical contents is via a finite number of sen-
tences whose empirical contents can be learnt in isolation are clearly learnable
(assuming humans can perform recursive symbol manipulation). The argument
for the epistemic inaccessibility of analyticity thus appears to depend on showing
that for any language in which a given sentence expresses a non-logical analytic
truth, there is a clearly learnable language whose speakers would be (or would
ought to be) indistinguishable in point of their dispositions to assert to/dissent
from sentences, but in which that sentence expresses a universally held belief
rather than an analytic truth. But as it stands, this claim is demonstrably false,
or so at least if we count the higgeldy-piggeldy collections on contentive sen-
tences as languages, they are not recursively axiomatisable at all. No speaker
of a clearly learnable language could match those dispositions to assent to sen-
tences (nor judge, in general, whether his peers possessed them). Clearly the
scope of the claim needs to be restricted to languages which are themselves

---

[20]The term is Quine's see (Quine, 1960, p.83)

plausible human languages.

Thus far we have only argued that clearly learnable languages are within human comprehension (provided humans can perform recursive operations), we now need to show that all human languages are clearly learnable; for if a language was not clearly learnable the relaxation of conceptual laws and demotion of analytic truths need not produce one, and the central argument against analyticity would yield counter-examples.

For languages which need not be adopted all at once, that may be learnt in piecemeal fashion and in which no particular piece must be learnt first to allow a novice to express any content at all, a reasonably good argument can be made for the content of theories being conferred by a finite number of independently contentive sentences. Any statement that can serve as an entry point into the language (e.g. "mama" or perhaps, in hothousing families, "mama is here") and be used by a novice who has mastered only that single sentence to express the same contents as expressed by (logical closures of attributions of experience to the novice by) fully competent language users must have its own independent empirical content in isolation. And clearly finite creatures such as ourselves are capable of learning only a finite number of such separate lessons. But this does not guarantee that the contents of every theory in every humanly devisable language must be the intersection of the contents of a fixed number of its independently contentive members. No doubt devious Gödelian techniques of encoding content into theories can be contrived which do not solely rely upon a finite number of individually contentive sentences, awaiting only interest and any genuine doubt that they exist to merit a logician's construction of them (though in any theory in which finite communications are useful empirical contents must be mediated by individually contentive statements, there is no guarantee that there must be just a finite number of them). It is tempting to dismiss at least some such languages on the grounds that their adoption requires a conspiracy formulated in a prior language, but the mastery of quantificational calculus (on Quine's description of it) is itself acquired in a series of discrete steps, and it is far from clear that these languages could not be acquired ab initio through equally small and plausible psychological steps without the aid of explicit linguistic instruction.

The second requirement, the recursive axiomatisability of contentive theories and possession of full quantificational apparatus, fares little better. Our generous treatment of novices' capacities to express contents upon mastering less than a handful of observation sentences makes clear that these requirements are not ironclad, for we counted a novice as holding a contentive class of statements true when his disposition to hold just one sentence true under appropriate conditions was exercised. Not only can a novice make contentive assertions

without full logical competence, but there might be complete languages which lack such apparatus. But languages such as these may be put aside on the grounds that they are inadequate to expressing sophisticated empirical theories — incapable of capturing recursively enumerable classes of independently con/infirmable statements. Thus on the one hand we exclude languages which are incapable of capturing recursively enumerable classes of independently contentive sentences as the consequences (requiring assent on pain of incompetence) of recursive classes, and on the other our assumption that humans are incapable of performing non-recursive operations leads us to exclude all languages except those in which the classes of sentences a competent speaker may be simultaneously disposed to give assent are just the closures of recursive classes of sentences under a finite number of recursive syntactic operations.

The question that thus confronts us is whether the notation of quantificational calculus is the only framework capable of rendering the empirical content of sophisticated scientific theories, or if there are irreconcilably different alternatives which rely upon recursive relationships other than quantificational implication to circumscribe the patterns of competent assent. To be sure there are notations based upon functors, or upon geometrical or topological notions, but these alternatives are readily translatable into quantificational notation and may be viewed as mere alternative symbologies for the same logical apparatus. Our question is whether there is an alternative logic, equal in point of the complexity of the theories it can support, but untranslatable into our classical notation. As far as I can tell no proof of the existence or non-existence of such irreconcilable alternatives is known, and I do not have one to offer. At first glance it may seem that a proof that there are no such fundamental alternatives is easily produced, after all the systems fundamental rule of inference is required to be recursive and its closures of recursive classes of statements recursively enumerable, hence we can construct a quantificational theory containing a two-place predicate V such that just those statements formed from $V(x, y)$ by replacing x and y with canonical descriptions of statements in some putatively novel theoretical framework where the statement replacing y may be reached from that replacing x by some chain of statements in which this fundamental rule of inference relates each link to its successor; furthermore we might imagine independently contentive statements being translated into equivalent or at least equally contentive statements in some quantificational notation (perhaps English). In this fashion, and by welding on these observational components by taking the equivalences as part of our quantificational version of the "theories" in such a novel notation we might produce a strictly quantificational equivalent. But this, as it turns out, is of small comfort. Applying this method to Chinese would (imagining that we could specify a primitive rule of inference

for the competent use of Chinese) produce just a rule for putting ideograms in quotation marks appended to "It is true in Chinese that", and specified both the rules of inference among sequences (operations on the quoted ideograms producing nothing but combinations which competent Chinese speakers might assent to) and, for some handful of independently contentive Chinese sentences rules such as "If (an instance of) '癸 子' is true in Chinese then the sentence's asserter is seeing red". Such a method is not a translation at all but rather a (rather ugly) way of specifying the semantics of (a) Chinese (in English) , and even that accomplishment may be doubted when we reflect that our assurance that a theory with the desired theorems "⌜ y ⌝ is a consequence in Chinese of ⌜ x ⌝" (mutatis mutandis for other more novel languages) can be constructed in quantificational notation (i.e. that any recursively enumerable set of consistent sentences can be captured as the theorems of some recursively axiomatisable theory) comes from Craig's theorem.

Such pseudo-translations do not provide a systematic mapping between statements in a given quantificational notation (such as, roughly, English) and those of the novel tongue; no foreign equivalents would be given for "or", or "there exists" of the rest of the distinctively quantificational apparatus. Our question is whether there are incommensurable alternatives to quantificational logic, at best all that such pseudo-translations show is that the existence of a method for converting foreign "theories" into empirically equivalent quantificational theories does not preclude the incommensurability of the two tongues.

So we are left without proof that either of the two requisites of clearly learnable languages are possessed by all humanly learnable languages There are complex techniques for expressing content that do not rely upon a finite number of individually contentive statements, and there may be alternatives to quantificational calculus capable of articulating the content of sophisticated scientific theories. Still the situation is not so bleak, for it is tremendously plausible that no language which does not fulfil these two criteria has ever been used by humans in practice. There is no known incommensurable alternative to quantificational calculus capable of expressing the contents of sophisticated scientific theories, and every such theory that has ever been expressed has been expressed in a notation translatable into quantificational form. Likewise, though it is possible for contents to be communicated via Gödelian schemes, in practice no one has done so, and even the specification of such schemes has always been given in other more straight forward languages. Even if the argument is restricted solely to clearly learnable languages, showing that non-logical analytic truth in such languages is epistemically inaccessible, it is still reasonable to suppose that it yet applies to every language ever used by humans to express sophisticated empirical theories.

The chief controversy over whether all our theories can be recursively axiomatised (upon which our conceit that all non-trivially contentive theories are recursively axiomatisable relies) concerns intentional expressions such as "believes that ..." and "hopes that ..." and the failure of substitutivity they engender. A similar failure occurs in modal contexts too, but surely this is permissible in the context of the overall argument, for surely if being a modal truth is to amount to anything it must amount to more than merely being called a modal truth. We are seeking evidence that such claims are true, but this is to ask how to introduce modal expressions into languages that were previously free of them and justified modal truths into the theories they express, or equivalently how to tell that a given expression of some hitherto foreign language is a modal operator for the evidence for modal truth must pre-exist the theory in which necessity is correctly attributed. We are surely permitted to admit our ignorance of modal truths and modal expressions before we seek to understand them, indeed one of the snippets of philosophy that has made it into a popular soundbite is that knowledge begins with the admission of ignorance.

We cannot afford to be so cavalier about propositional attitudes however, for their predictive power is daily attested; given the choice we must admit the paucity of our quantificational notation rather than spurn them as lacking meaning. The argument touches here upon a deep subject to which we can give only cursory attention, but if the argument is sound and our skepticism about modal cum conceptual truths extends also to facts and properties then it is no wonder that it also extends to propositions, for they are conceptual structures if any are. Still, if propositions are objectionable, there is still need to posit internal or mental objects of some sort, for just as external objects must be posited in theories with complex empirical contents[21], internal objects must be posited in psychological theories attributing complex expectations. Indeed it would seem to be a meritorious re-use of intellectual achievement to base such a psychological theory upon more or less shared theory about the external world (provided that theory was not too far from the truth, and expectation not too divergent from actual experience), the obvious course would be to posit mental correlates of the external objects. This suggestion is perhaps too primitive, too medieval sounding, to have received the attention it deserves from modern philosophers, and, as best it could be put by enlightenment philosophers, it succumbs to a well known fatal objection viz., that it depends upon a spurious relationship of resemblance between mental objects and their external counterparts to link them together, but of course this resemblance can never be determined since the external objects can never be known "as they are in themselves". Something of this objection is often raised against both Locke and Berkeley's empiricisms,

---

[21]See section 1.1.2

and Wittgenstein's *Tractatus* was in large part an attempt to give a relationship of (modal) resemblance that did not require such unobtainable empirical verification; but those philosophers restricted constructions were inadequate to tracing the relationship between posited objects and empirical evidence in any case, so it is no surprise that they could not do justice to mental objects in particular. If such posits are justified it is as part of a psychological theory which attributes expectations to subjects. Obviously there are significant differences between expectations and stimuli, perhaps the most salient is that expectations are general, capable of being fulfilled by a number of distinct stimuli sequences, whereas each instance of n second total physical impingement is a member of just one stimulus sequence. Such generality must affect mental objects too, and scope must be allowed for it; but even if this difficulty can be overcome, it may yet be objected that all that would be shown is that "expectation" is a member of the closed circle of inter-definable intentional terms, and it is true attributions of expectations are not generally independently contentive. One hope is that expectations might be given a neurological account, this would unify biology and folk psychology at this level if no other (for we need not presume that mental objects should admit of any general neurological reduction), but one is struck in this connection by the difficulties of defining energy functions for any neural networks but those that tend unceasingly towards stable states. Still one can take some cheer from psychologist's success in determining (some of) the observational expectations of pre-linguistic infants by means of the duration of their gaze. The point is that some internal objects must be posited in theories that can do justice to the complexities of our dispositions to behaviour, it may be reasonably hoped that neurological cum physiological theories are among them; and the suggestion is that folk-psychology is a great deal less mysterious if interpreted as positing mental objects closely based upon the external objects posited in the sentences embedded in attributions of propositional attitudes.

In any case our justification for believing that quantificational apparatus is adequate even to the attribution of propositional attitudes and the expression of the contents of the theories of which they are members does not depend upon the plausibility of any particular scheme for their regimentation but upon our assessment of human computational capacity. An ability to compute functions beyond the recursive and recursively enumerable cannot be restricted to any one field of study and, though we have noted the possibility of other incommensurable notations which reflect the same computational capacity there is good reason to suppose that folk-psychology is not built upon such an alternative: both because it is built, at least superficially, upon the same logical lexicon, and because the content of the beliefs (and other propositional attitudes) attributed in folk-psychology can be expressed in quantificational notation. While these

consideration do not conclusively demonstrate that all human languages are clearly learnable it remains a highly plausible hypothesis.

With the scope of the argument against analyticity restricted to clearly learnable languages the central claim of the argument, in full, is as follows: For any language possessing the full quantificational apparatus, in which only recursively axiomatisable theories are (non-trivially) contentive, and their contents are determined by those among a finite stock of independently contentive statements which are their members together with such statements as be formed from them by (repeated) conjunction by the phrase "is immediately after"[22] (i.e. any clearly learnable language), in which there are analytic truths which are not (overtly) logically valid there is another clearly learnable language which maps all the theories expressible in the original to the same empirical contents as in the original language, but, which also maps some theories which do not contain the sentences expressing these non-logical analytic truths to empirical contents (i.e. in which those sentences do not express analyticities); any competent speaker of the original language is also a competent speaker of this alternative language, albeit one who eschews, whether by whim or hard-wiring, those theories in which the analyticities of the former do not feature. Likewise speakers of any clearly learnable language in which a given sentence is not analytic who yet always adopt theories which contain that sentence are also competent speakers of a more restrictive clearly learnable language in which that sentence is analytic (i.e. is a member of every contentive theory).

All that remains to be shown is that these alternative languages exist, and that is trivially easy. Consider any clearly learnable language in which there are (overtly) logically invalid analytic truths (i.e. in which any theory which does not contain these analyticities is not compatible with any stimulus sequence). This language may possess the full quantificational apparatus and the contents of any (contentive) theory expressed in it be are mediated by those of its members from a finite stock of independently contentive statements; but obviously the semantic rules of such a language must go further than specifying just these two components since these alone do not fix any logically invalid statements as analytic truths. Consider the language whose semantic rules are exhausted by the specification of these two components, a language whose semantic rules fix the logical terminology and the syntactic realisation of the patterns of logical implication constraining the combinations of statements that competent speakers may assent to at the same time, and also specify the class of independently contentive statements sufficient to fix the content of any competently assertible theory. Since the sentences in the finite stock of content fixing independently

---

[22]Anxieties over this connective are considerable allayed by the fact that given independently contentive statements as arguments it produces only independently contentive statements.

contentive statements are the same in both of these two languages, and are associated with the same independent contents then any class of sentences that is contentive in both languages will have the same empirical content in each, and since the same logical apparatus receives the same syntactic embodiment in both languages any class of sentences that is contentive in the more restrictive language will also be contentive in the language whose semantic rules are exhausted by the specify no more than the logical apparatus and content mediating sentences. Thus any class of sentences which is contentive in the more restrictive language will have precisely the same content in the more liberal language (in which there are no non-logical analytic truths). Obviously there are some classes of sentences that are (non-trivially) contentive in the more liberal language that are not contentive in the more restrictive language at all viz., recursively axiomatisable classes of sentences which do not include the analytic truths of the restrictive language. Thus for any clearly learnable language in which there are analytic truths which are not (overtly) logically valid there is another clearly learnable language which maps all the theories expressible in the original to the same empirical contents as in the original language, which also maps some theories which do not contain the sentences expressing these non-logical analytic truths to empirical contents.

As we have already seen competent speakers of the more restrictive language are competent speakers of the more liberal language just as they stand, and competent speakers of the more liberal language who universally eschew theories in which the analyticities of the more restrictive do not feature are competent speakers of that more restrictive language. The upshot is that the assessment of which language a person is speaking is critically under-determined, and the choice of whether to take a statement as expressing an analytic truth or merely a belief that is held universally by the members of the relevant speech community floats free of any evidential connection remaining only as an artefact of the interpreting linguist's arbitrary preferences.

It has been suggested[23]that the epistemic inaccessibility, the freedom of choice linguists have in assessing the analyticities of foreign speakers, is simply another instance of the under-determination of empirical theories by evidence. Just as we do not dismiss the term "electron" simply because its extension is under-determined by the evidence, neither should "analytic" be rejected because its extension is under-determined. It is true that the epistemic inaccessibility does establish a kind of indeterminacy, but it is an extreme kind of indeterminacy — vacuity. Attributions of analyticity to any universally held true statement can be added to, and any attributions of analyticity deleted from, any empirical linguistic theory without any altering its empirical content in any

---

[23]See (George, 2000)

way. But still it is fair to wonder just what is wrong with vacuity in the first place. Philosophers who are confident in an absolute measure of simplicity, and that superfluities count against it, appear to have a ready answer, but I am as suspicious of simplicity here as elsewhere[24]. Such faith seems to me to amount to the re-introduction of faculty of a priori reasoning under the guise of a sense of simplicity. Mine is a more tolerant outlook and I do not suppose that there is anything so very dramatic to be said against the truth of theories with vacuous components, but they do carry some more accessible, if more minor, flaws on other points. The distinction between changes in belief and changes in linguistic/conceptual framework in particular, is apt to give conceptual truths undue strength against revision for it is easy to feel that the creation of a new linguistic framework is a momentous event, requiring its gifted creator to envisage a whole new conceptual scheme, whereas the mere tinkerer with more non-conceptual beliefs (practitioners of Kuhnian normal science) is free to rely upon the concepts supplied by his predecessors. But vacuity carries other flaws more generally. Analyticity is just as pointless an introduction to our theories as "neo-red" which denotes a hue of red indistinguishable from fire-engine red but instantiated by just the post box on the main street of Glenbrook, and Katoomba's two fire-engines. The proper response to theories attributing attributing neo-red to just these three objects is not to deny them as false but rather as simply pointless. Hold such theories if one will, it is obvious why they are not taught in schools and universities, a purely pragmatic drive to economy in the inculcation of theories is sufficient to drive us away from vacuity. Still if it is not outright false to attribute neo-red to objects, it is outright false to suppose that identifying neo-red things contributes to our understanding of colour or identifying analyticities contributes to our understanding of semantics, for these vacuous elements are the products of their inventors free imaginations.

4.5.1 Objections and loose ends

Our consideration of the argument for the epistemic inaccessibility of analyticity is nearly done, but there are still three loose ends that need to be secured.

Firstly, we have looked only at speakers not listeners and attended solely to theories language users are disposed to hold as a result of their experiences, rather than the range of experiences they are prepared to attribute to their peers as a result of their assertions. But linguistic competence requires more than the correct dispositions to assent or dissent from proffered sentences under stimulation, competent users must also be able to understand assertions made by their peers. By the lights of our holistic empiricism this amounts, roughly,

---

[24]cf. 6.2.1

to being able to determine which of those stimulus sequences their peers are apt to be exposed to are compatible with their assertions. In ignoring this second aspect of competence it appears that we may have overlooked critical evidence for analyticity. For though it is compatible with competence for a speaker be disposed to *hold* just one of several empirically equivalent theories in the face of an experience with which just those theories are compatible, competence requires him to *understand* all of them. A fully competent user is required to be able to determine which stimuli sequences are compatible with any contentive theory. It thus seems that competent speakers of restrictive (non-logical) analyticity containing languages can be distinguished from the speakers of more liberal languages by the fact that the former, but not the latter, are at a loss to make sense of a peer's (or investigative linguist's) denial of the analytic truths of the restrictive language in question. But already this is somewhat of an exaggeration, for analyticities are supposed to be analytically *true* and their negation analytically *false*, not lacking in truth value altogether. A speaker of the more restrictive language must be supposed to hold the visiting speaker of a foreign dialect as asserting (some) falsehoods, not as uttering nonsense. Still, the speaker of restrictese holds the class of statements to which the speaker of the more liberal language is disposed to give assent to be incompatible with every stimulus sequence, whereas the liberalese speaker does not, or such at least is the diagnosis proponents of analytic truth (and holism) invite us to share.

But there is an alternative diagnosis which squares equally well with all observable evidence. For according to holism, understanding a fellow speaker is not a matter (or at least not solely a matter) of understanding their individual assertions, but rather of being able to determine the empirical content of the whole theory that fellow speaker espouses. But of course, we are never exposed to one another's theories all in one go. Exposition of a theory is sequential, we are each presented with one another's theories one statement at a time, and only ever directly presented with a finite selection of the infinite number of statements (as our conceit has it) that comprise our fellow's theories. In order to understand one's fellows, to grasp which theory they hold, and thus the empirical content of their assertions (in the only sense in which most have any) one must in practice attribute unstated background beliefs to them. To get along with ordinary linguistic use one must take one's peers as stating the most salient or idiosyncratic elements of their theory, leaving a great wash of unstated commonplaces (often but not necessarily assumed to be shared by the asserter and her audience) to flesh out their theory in ways which affect the empirical content of the whole. There is at least the appearance of an exception to this principle in special cases where the statements of one's fellows imply all their beliefs and can serve as axioms of their theory, but even there one must rely

upon an assessment as to which other sentences one's peer will give assent to in order to judge that their assertions do axiomatise their beliefs, and in practice in any case no-one can give an axiomatisation of their beliefs.

The speaker of restrictese, in which sentence S is analytically true, who unknowingly stumbles across a liberalese speaker mid-oration will naturally enough take the speakers background beliefs to include S. For if S is to be even putatively analytic it must be held true by all the members of the restrictese speech community, and we may suppose that the speaker of restrictese takes his companion to be a fellow member of the community (for we are in effect investigating whether the two languages can be distinguished by the limits of competent language users comprehension, and hence may suppose that members of the two communities are indistinguishable in those features that are irrelevant to their linguistic competence). Indeed he may be so confident that the liberalese speaker's theory includes S as to insist that it does even in the face of the liberalese speaker's clear assertion of its negation, let alone statements that are less obviously inconsistent with S.[25] Even if this insistence that his supposed companion's theory includes S is due, in part, to a simple inability to elucidate a contentive theory that contains S, speakers of each of these two languages cannot be distinguished on this basis. For one may be a competent speaker of a language without being able to state a theory fulfilling any given constraint even though though there is such a contentive theory in that language After all, as the long-standing clash between quantum mechanics and relativity bears out, it is quite possible to be at a loss for a theory which has even a desired empirical content.

Thus the restrictese speaker's assessment that his liberal companion's theory is not compatible with any experience can thus be diagnosed as a misjudgement of the background assumptions included in that theory. Equally the correct assessment of those background beliefs can be recast as the discovery of a translation between distinct languages. Once he has mastered the putatively foreign language, the the newly open-minded auditor may continue steadfastly in

---

[25]To insist that a speaker is covertly committed to a sentence that she does not overtly assert is to no longer take dispositions to assent/dissent quite at face value, more strictly the restrictese speaker believes there is some chain of inferences which once attention is drawn to them will lead his apparently confused compatriot to assent to S after all. This is to surrender much of the behavioural immediacy of our account of holding a theory as being disposed to assent to its member statements, though the slack can be provided by altering either the conditions under which the relevant dispositions to assent are exercised (allowing for attention to be drawn to various syntactic features of the statements involved and other intermediaries, or in more flagrantly intentional terms taking the relevant dispositions to be those that obtain under reflective equilibrium) or by allowing a less restrictive account of holding a theory than being disposed to assent to the statements that compose it; but this much must be done to maintain the conceit that human are capable of performing recursive operations (and hence of mastering quantificational calculus) in any case. See the treatment of the third objection below.

his own adherence to theories containing S, he may even render the liberalese speaker's theory as including S when called upon to state the foreigner's theory in his own words; and yet still be treated as speaking the same language (rather than one in which S is analytic and another in which it is not) for there is no epistemically accessible line between non-homophonic translation between languages and indirect quotation within the same language. In relaying a linguistic peer's assertions to third parties one may turn to indirect quotation, recasting the original assertions so as to incorporate background assumptions shared only with those third parties without casting doubt on one's common linguistic allegiance. In both translation and indirect quotation the preservation of empirical content is the only essential requirement.

Thus the receptive linguistic capacities provide no more evidence for distinguishing between more and less liberal language, between analytic truths and commonly held beliefs, than the productive capacities. The distinction between liberalese and restrictese cannot be sustained on either dispositions to assert/dissent sentences or the limits of our abilities to make sense of our neighbours assertions and denials.

The second loose end is our tacit assumption that the the "conceptual structure" of the languages we have been contemplating does not preclude any stimuli sequences (a priori as it were). We considered only languages sufficiently flexible to express some theory compatible with any stimulus sequence. It might be thought that at least analyticity could be determined in such impoverished languages, in part by virtue of some distinctive response subjects might be supposed to produce upon being exposed to stimuli they are incapable of accommodating in their theories (such as being at a loss for words). But being at a loss for words is no guarantee of impending conceptual revision, being at a loss for a theory that matches the available evidence for any length of time does not (in itself) cast doubt on whether such a theory can be expressed. Incomprehension of one's surroundings is not a purely linguistic affair.

The third and perhaps most troubling loose end is the charge that the holistic standard of good translation relies upon an untenable idealisation of the evidence that is itself not empirically justified. In the name of simplicity and practical exposition, we have abstracted away from the mendacity and bias of the subjects upon which linguists' translations are based and most importantly from their fallibility. When these complicating factors are recognised, the charge goes, they reveal that the holistic standard of translation relies upon an overestimation of the empirical evidence. Borrowing Bertrand Russell's phrase, we have proceeded by theft rather than honest toil, imagining the world to be built to suit the needs of our philosophy rather than building with the materials supplied.

According to the holistic account of translation relied upon in the argument against the epistemic accessibility of analyticity, a good translation with English is required to map English sentences onto foreign sentences in a way that preserves the empirical contents of individually contentive statements and renders each competent foreigner speaker as disposed to assent[26] when prompted by just the translations of sentences that in English express some logically consistent theory. But, the charge continues, it is obvious that no good translation into English from a language spoken by humans satisfies these constraints. Human speakers are fallible, they misjudge both the stimuli to which they are exposed and the logical implications of their assertions, even worse they have only finite memories and attention spans (and in the limit of those finite lives) and there are perfectly meaningful sentences that are so long that no-one is disposed to assent to them, even if their other assertions logically imply them. For these reasons and many others speakers are not disposed to assent to sentences expressing a consistent theory in their own tongue, and any translation that renders them as disposed to assent to just the translations of sentences expressing a consistent theory in English is a mis-translation.

The objection's proponents claim that there are no empirically accessible properties distinguishing mistaken assertions, no empirical evidence that justifies excluding the dispositions to make such errors from the evidence upon which translation is based. Thus the challenge appears to be to find empirical evidence that justifies putting aside erroneous assertions(/assents). Many current philosophers have taken up this challenge and attempted to state the empirically accessible conditions distinguishing mistaken assertions, but thus far none of their proposals has been widely accepted. The suggestions have included putting aside dispositions to assert(/assent to) sentences that would not be rewarded in a learning situation[27], that are asymmetrically dependent on correct dispositions[28], and that do not survive reflection or the criticisms of peers.

If the version of holism we have been considering is correct, it is futile to seek a distinguishing feature of erroneous statements in isolation. To call a statement mistaken is to suggest a reform of its proponent's theory that will bring it one step closer to being consistent with the evidence or cure some other ailment of the whole theory, and there is no reason to believe that there is only one such reform that will do the job equally well. Holism implies there are many different theories having the same empirical contents, and a reform which results in any of them is equally acceptable. Just as theory is under-determined

---

[26] Once they have recanted their errors
[27] See (Dretske, 1981)
[28] See (Fodor, 1992b)

by the evidence so too is error.

Let us apply this insight to the main example given by the objection's proponents. Suppose Bertrand has been taught addition by being trained on pairs of numbers less than 57. When asked to add any such numbers Bertrand successfully gives their sum in response. When asked to sum larger numbers however he always produces the answer 5. There is a function that corresponds to Bertrand's dispositions, call it "quum". The question, as the objection's proponents set it up, is to determine what, if anything, determines that Bertrand is erroneously computing the sum of two numbers when asked to add numbers larger than 57 rather than correctly computing the sum. Or, to put it another way, what is it that determines whether "+" as Bertrand uses it should be translated as "sum of" or "quum of".

According to holism we should not expect there to be anything marking Bertrand's disposition to quum numbers when asked to add them as a mistake taken in isolation, it is only in the context of a whole theory that they can stand out as deserving revision. Suppose Bertrand goes to a restaurant and orders some dishes costing more than 57 dollars. When presented with the bill and asked to pay he calculates the quum of the prices (or miscalculates the sum and produces the answer 5) and offers the maitre'de 5 dollars. At least one component of Bertrand's error is easy to spot now, he predicts (presumably) that the maitre'de will be satisfied by his offer, that he will not be detained for failure to pay his bill, and that the police will not be called. His empirical theory turns out to be incompatible with the after-dinner evidence. One way to bring Bertrand's theory into line with the evidence (or at least a step closer to it) is to replace some of the sentences in which "plus" occurs in Bertrand's theory such as "five plus one hundred and two equals five" with sentences such as "five plus one hundred and two equals one hundred and seven" which in English express truths of arithmetic. The sentences eliminated in this revision are those in which Bertrand has made an error.

While this is a diagnosis of which of Bertrand's statements are in error it is not enough to satisfy the objection's proponents. They contend that Bertrand's assertions are not only mistakes but that they are mistakes in arithmetic; that (prior to any revisions of his theory) Bertrand is disposed to mistakenly assert the translation (from English) of "five plus one hundred and two equals five", rather than correctly assert the translation (from English) of "five quus one hundred and two equals one hundred and seven". The suggested modification of Bertrand's theory however can be equally regarded as a change in meaning of the term "plus" in his idiolect, or as a change in his beliefs concerning addition (or indeed as a change in his beliefs concerning quums, but let us leave that aside). On the one hand we might, prior to any modifications in his theory, translate

Bertrand's term "plus" into English as "quus" , and afterwards translate it as "plus", but on the other we might translate it as "plus" both before and after and take the modifications to be reflect a change in his beliefs concerning addition. On the former view the modifications in Bertrand's theory are taken to constitute a change of the meaning of "plus" in his idiolect, whereas on the latter they are taken as a correction of his disposition to assert arithmetic falsehoods. Since the offered explanation of Bertrand's errors does not settle between these two accounts it does not meet the original challenge, viz., to explain why it was wrong all along to translate Bertrand's term "plus" into English as "quus" despite his disposition to assert(/assent to) sentences such as "eight plus fifty nine equals five". The objection's proponents conclude that the explanation must leave something out. Even more strongly, they contend that if there is no empirical evidence that can justify our knowledge that the proper English translation of Bertrand's term "plus" is "plus" all along then empiricism is false and there are truths about semantics that no empirical investigation can reveal.

By and large the defenders of empiricism (in one form or another) have taken the objector's challenge at face value and sought to delineate empirically accessible properties which justify translating Bertrand's "plus" into English as "plus" and controvert its translation as "quus". Generally the defences they have offered have remained dispositional, but, instead of basing translation on the evidence of honest cooperative speakers' every disposition to immediately assert(/assent to) sentences they have thought to exclude some dispositions from consideration either on the grounds of the conditions exercising them, or of their relationships to speakers' other dispositions to assent to sentences. Defences of the first sort seek to delineate special *type one situations* in which the subject is more or less infallible and count only the dispositions exercised by such situations. One defence has it that *type one situations* are learning situations where "special care is taken to see that incoming signals have an intensity, a strength, sufficient unto delivering the required piece of information to the learning subject"[29], another that they are situations which give the subject time to reflect upon his responses and/or discuss them with critical peers. Defences of the second sort include suggestions for excluding dispositions that would not be re-inforced in learning situations should not be counted, or that are asymmetrically dependent upon dispositions to assert the sentence in question under other conditions.

In general the difficulty for such defences (apart from getting the intuitively right answer) is to distinguish subjects' dispositions to speak their language well from their dispositions to speak other languages. The suggestion that translation be based solely on (dispositions to make) assertions surviving the

---

[29]See (Fodor, 1992b, p. 40)

criticism of linguistic peers for instance, must be accompanied with a distinction between criticisms of subjects' use of the language and criticisms of the language and proposals for its reform. My suggestion is that the distinction cannot be drawn; that, like "species", "language community" is a term "arbitrarily given for the sake of convenience to a set of individuals closely resembling each other .... it does not essentially differ from the term variety [dialect], which is given to less distinct and more fluctuating forms. The term variety, again in comparison with mere individual difference [idiolect], is also applied arbitrarily, and for mere convenience sake."[30]. Like species, there is no hard and fast line to be drawn between those changes that alter a subject's linguistic affiliation and those that leave him speaking the same language

Rather than seeking to meet the objector's challenge and find empirical evidence that justifies translating Bertrand's "plus" into English homophonically and confutes its translation as "quus" I propose biting the bullet and maintaining that both translations are perfectly acceptable. This is not to object to one or another way of filtering subjects dispositions to assert sentences before basing translation of the evidence of them, but simply to drop the claim that any one such filter is uniquely correct. There are two main objections to this course, the first may be called the objection from reasoning and the second the objection from common-sense.

Sometimes speakers seek to change one another's verdicts upon sentences by presenting arguments rather than contrary evidence. Such arguments typically call upon their targets to assent to the argument's conclusions on the grounds that they already assent to the argument's premises. It is essential to our understanding of philosophy (and more generally of reasoning) that there are some good arguments; arguments that *require* (or show that it is required) competent speakers to assent to their conclusions if they assent to their premises — to fail to do so is to make a mistake. Essential to philosophy in particular because philosophical disputes are between interlocutors whose empirical expectations are the same, or at least not relevantly different; if either party concedes that the others argument is good and modifies her theory accordingly she does so without altering the empirical contents she uses it to express. Consider the current argument (and indeed the whole current text), no evidence is provided, it is doubtful that reading the current text exposes readers to stimuli confuting their expectations[31], and no alteration to the reader's empirical expectations is called for. Despite this I (defeasibly) expect the reader to assent to the conclusions or disagree with the premises and hold her to have made a mistake if she does not

---

[30]From *On the Origin of Species* in (Barrett and Freeman, 1987), p. 39

[31]At least not in any relevant way. The text may be surprisingly well or badly written, but the revisions motivated by that surprise do not include adopting its conclusions.

do so.

By contrast, general application of the tolerant attitude advocated towards the translation of Bertrand's assertions appears to produce an insipid relativism according to which no-one's assertions can ever be mistaken. Any subject who appears to be making an error, for instance asserting to the premises of a logically valid argument in English while denying its conclusion, can always be redeemed by treating him as speaking another language; indeed according to the holistic standard good translation based upon his unfiltered dispositions *must* render him as speaking a language other than English in which the argument in question is not logically valid (perhaps his "and" must be translated as "quand"). We thus appear to lose all right to criticise one another's assertions as false and one another's arguments as invalid (though there is still scope for the presentation of contrary evidence). Distinctions between good and bad reasoning, consistent and inconsistent assertions, appear at very best to depend upon brute conformity to arbitrary convention.

If these appearances are accurate the holistic dispositionalist standard of translation we have been considering (and perhaps empiricism) must be wrong. If that account of translation implied it was not an error to assert its negation there could be no good reason to hold it true. Fortunately there is yet hope of reconciling the tolerant attitude urged towards translating speakers such as Bertrand and the acknowledged truth that they are disposed to commit semantic mistakes (whether or not their theories are empirically adequate).

Provided independent and empirically accessible grounds can be given for holding certain whole languages to be defective; and we can give an independently motivated account of why such it is a mistake to speak such a language; and, finally, if to translate Bertrand's term "plus" into English as "quus" is to render him as speaking a defective language then translation based upon his immediate dispositions does not exonerate him of committing a linguistic error. Either he can be translated as a competent speaker of a defective language (going by his immediate dispositions) or an incompetent speaker of a non-defective (or at least relatively non-defective language). In either case he may be justifiably held not to have expressed himself well, and his dispositions to assent to sentences merit reform.

The nature of defective languages is taken up in part three of the current work, and for the moment we will continue to clear away loose ends in preparation for it. For now let us restrict ourselves to the weaker claim that *if* such a distinction can be drawn then we can objectively diagnose dispositions such as Bertrand's as flawed or mistaken.

The second objection to our tolerant attitude stems from common-sense. According to common-sense, English speakers express conjunction by "and"

and the result of flanking an instance of the term "and" with two statements
(or more precisely sentential phrases) is to produce a statement which is true
in English just in case both of the flanking component sentences are true in
English However, according to the standard of translation we are considering,
for a subject to express conjunction by use of "and" then (among other things)
he must be disposed to assent to each sentence of the form "... and ..." just
in case he is disposed to assents to each of the sentences taking the place of
the ellipses. However it is obvious that not all English speakers are disposed
to maintain such a pattern of assertions; they make mistakes and sometimes
assent to conjuncts of sentences that they do not assent to individually. Thus
according to the suggested standard of good translation any good translation
based upon subjects unfiltered dispositions to assent to sentences must render
them as not expressing conjunction by "and". If common-sense is right then
the suggested standard of translation must be incorrect.

According to its proponents this objection is even more pressing than the
last. Human subjects are computationally limited, they have finite attention
spans, finite memory and endure for only finite periods of time, while there is
no upper bound on the length of grammatical English statements. Thus there
are some English statements, and in particular statements that are (in English)
logical consequences of putative English speakers assertions, that are too long
for speakers to have any stable disposition to give assent/denial when prompted
by them. Prompting sentences that are too long or complicated for (unaided)
human comprehension might elicit only a confused "Huh ??", while speakers
cannot have any disposition at all to assent to or deny sentences taking longer
than a single human lifespan to enunciate (or read). Thus no filtering of human
subjects' dispositions can leave an evidential base mandating the translation of
their terms as expressing conjunction or disjunction.

The most obvious suggestion, and indeed the only one seriously considered
by both the objection's proponents and the defence, is to widen the range of
evidence through appeals to the dispositions of idealised subjects who have
unlimited computational resources.

The main problems with an appeal to ideal subjects are motivating the ideal
conditions and showing that the required dispositions would obtain under those
conditions. Unlike current computers[32] humans' computational architecture ap-
pears to have no neat division between memory and processing and there is no
ready sense in which 512K can be added to someone's memory. Haphazardly
connecting additional brain matter is overwhelmingly unlikely to increase com-
putational memory. Even worse, the only relevant measure of human memory
capacity is presumably their ability to perform increasingly memory intensive

---

[32]i.e. Computers based upon either the Harvard architecture or von Neumann's.

computational tasks, but to presume which tasks subjects are being asked to perform and thus to distinguish between those additions that allow the performance of more memory intense tasks and those that are disruptive is to beg the question of which language subjects speak under idealisation.

A more trivial (though to judge from my own limited observations more frequently mentioned) problem, is that even if the ideal conditions can justified independently rather than by virtue of producing the right answer, we have no idea what would be the case if they were fulfilled. Perhaps creatures with unlimited memories wouldn't speak languages like ours at all, or perhaps infinite memory capacity would require an infinite universe and ours is finite or perhaps it would require faster than light interaction — the list can be extended arbitrarily.

This aspect of the criticism of idealisation adds nothing of substance to the first. There are irreproachable idealisations to conditions under which our theories give no clear predictions. Our theories predict that there are no frictionless planes, mass points, or ideal gases, and there is no decided method for modifying them to predict what would be observed if there were. However the saving device is well known, statements about these ideal objects can be salvaged by paraphrasing them as statements about how objects tend to behave the more nearly they approach the ideal. But this presents a genuine problem (albeit only a more sophisticated version of the first problem mentioned above) for we have no measure of how far short subjects fall of the ideal at which their computational resources are unlimited (or where exactly this ideal is) which does not beg the question of which language they are speaking in the first place.

The objections to idealisation are not conclusive, and many philosophers continue to try to find independently motivated and epistemically accessible criteria for setting the limits of the ideals. None of them have yet been successful and the prospects of success are so dim only a firm conviction that there is no other alternative consistent with naturalistic empiricism warrants continued attempts. Since I propose to accept that good translation based upon subject's unfiltered dispositions will not in general render putative English speakers as expressing conjunction by "and", disjunction by "or", negation by "not", or existential quantification by "there are some", and live with the consequences, let us not spend further time contemplating the details of the various idealisations philosophers have proposed and the ways they have been found wanting, but instead consider some other attempts naturalistic empiricists have made to meet the objection.

In *Word and Object* Quine takes a fairly casual attitude towards extending inferential connections to sentences so long that subjects cannot comprehend them. He suggests linguists simply concentrate on dispositions to respond to

short sentences and generalise to cover long sentences — "the semantic criterion of negation is that it turns any short sentence to which one will assent into a sentence from which one will dissent, and vice versa."[33] and "when we find that a native construction fulfils one or another of these semantic criteria, we can ask no more toward an understanding of it"[34]. But as the objection's proponents point out, there are many mutually incompatible patterns of inferential connection between sentences compatible with subjects' dispositions to respond to the finitely many grammatically correct short sentences of their language. Doubtless linguists do generalise the pattern of assent and dissent they discover in the finite number of subjects' responses they actually observe to cover the infinite number of sentences both long and short which they do not observe subjects responding to, relying upon the similarity between their own and their subject's tendencies to generalise from their observations leads them to get the right answer, but in doing so (the objections proponent's point out) they are making an assumption which might be mistaken. Quine's complacent approach has the noteworthy virtue of being reasonably close to what linguists actually do, but it does not explain why they are right in doing it.

In his later work Quine shifted his position, defending taking subjects as assenting to the translations of logical truths so long as there was some chain of obvious steps starting from obvious truths by which subjects could be brought to assent to the sentences in question. His slogan is that "the truths of logic are all obvious, or at least potentially obvious, i.e., derivable from obvious truths by individually obvious steps."[35]. Just how this principle is to be brought to bear on translation is not clear. The most obvious proposal is to take speakers as being disposed to assent (in response to a given stimuli) to just those sentences to which they can be brought to assent by some sequence of statements related by obvious steps, and which no further sequence of statements can bring them to deny. There is however no guarantee that there is just one such class of statements for each speaker. The order in which such sequences are shown to subjects may affect the sentences they end up assenting to, it will not affect speakers dispositions to assent to potentially obvious statements, but the order in which a speaker modifies his theory to include these statements may affect which other statements he ends up assenting to; and indeed there is not a guarantee that there is any final result at all, the quest for logical consistency is not a process that is guaranteed to terminate. There is no obvious way for resolving such problems, but there is worse to come.

For Quine's slogan to be plausible the requisite steps must be obvious in two

---

[33](Quine, 1960, p. 57)
[34](Quine, 1960, p. 58)
[35](Quine, 1969, p.70)

separable senses. In the first and weakest sense it must be evident just which
objects are related by any one of these steps. A step is obvious in this sense just
in case there are some plausible conditions under which nearly all competent
speakers agree whether two objects are related as antecedent and consequent of
that step. But this is not enough by itself. Appending the phrase "and one plus
one equals three" to sentences is an obvious step in this sense, but that sentences
can be derived by this obvious step (to English speakers) from statements that
are obviously true (to English speakers) gives no reason to take English speaking
subjects as assenting to those sentences for the purposes of translation.

Presumably Quine's suggestion is that in addition to being obvious in the
sense just considered, the requisite steps obviously preserve truth. Obvious
steps in a novel tongue relate (ordered) pairs of sentences as antecedent and
consequent just in case nearly all competent speakers are nearly always disposed
to either deny the first or assent to the second (once their attention is drawn
to them). As the objection's proponents may point out however, there are no
obvious steps by which intolerably long sentences may be derived since human
subjects can have no disposition to assent to sentences that take longer than a
human lifetime to enunciate, and (presumably) no stable disposition to assent
to sentences so long that they cannot comprehend them. It may be obvious in
some sense that an intolerably long sentence composed of two other intolerably
long sentences bracketing the term "and" is true if its component sentences
are true, but it is not obvious that a sentence is a conjunct of two intolerably
long sentences and to determine it from direct inspection may require greater
computational resources than any unaided human can bring to bear.

On this point I think that the objection reveals a minor but significant flaw
in Quine's (and our) account of translation. In what is surely one of the most
incongruous positions in a subject full of bizarre and incongruous positions
Quine, though the chief and best known critic of reductive empiricism, yet
insisted that respectable psychological statements must be individually reducible
to statements about behaviour or jettisoned from strictly scientific talk. His
basing translation upon dispositions to assent to sentences was one aspect of
this predilection, and it runs into trouble in application to intolerably long
sentences. As good holists however we can admit psychological claims which
are not individually reducible to behavioural terms. In particular we can admit
the psychological state of holding a sentence true, and it is to the dispositions
of speakers to hold sentences true under various stimulations that we must turn
if we are to be able to fix the translation of intolerably long sentences and the
(putative) logical terms they contain.

Shifting the evidential basis from speakers' dispositions to assent to querying
sentences to their dispositions to hold sentences true (in their language) leaves

the resulting account of translation open to the charge that it is based upon empirically inaccessible foundations, for it may be argued that the psychological state of holding a sentence true is itself not empirically discernible. A lingering attachment to strong empiricism in both critics and empiricists easily leads to the assumption that since there is no behavioural (cum sensory) reduction of individual psychological claims there is no empirical evidence for them at all. No simple refutation of this charge is available if such psychological states are empirically accessible then, like other theoretical claims, the range of ways of determining whether or not a particular subject instantiates a given psychological state is open-ended; nor is there much hope of an easy reduction to other more obviously epistemically accessible properties, at best such a reduction would be horrendously complex and, if anomalous monism is true, there may be no such reduction to be had. A complete vindication of the claim that the state of holding sentences true (in one's own language) would require the translation of a large chunk of (folk) psychological theory, including the attributions of psychological states such as holding true, into the austere idiom described at the end part one. That this cannot be readily provided gives little reason to suppose that it cannot be done at all. Still it must be admitted that our only reason for supposing that it can be done is the apparent (empirical) contentiveness of (folk) psychological theory together with our general confidence in weak empiricism. For the time being at least I think we can do little better than to acknowledge that our account of translation depends upon the assumption that the attitude of holding a sentence true is epistemically accessible by empirical means (or at least not more under-determined than translation in general).

In effect we are assuming that we can determine which foreign predicates are translatable as "is true". It may appear that in assuming that a predicate imputes truth to both short and intolerably long sentences we are going beyond the empirical evidence, artificially restricting the range of good translations to fit our preconceptions. However at least some empirical constraints upon such translation can be noted. Obviously, though unhelpfully for our current purposes, for a predicate to be translatable as "is true" subjects must be disposed to assent to its application to just those sentences to which they are disposed to assent of those that are short and comprehensible enough to prompt a response at all. Translation of a predicate as "is true" coheres with the translation of logical terms iff the sentences to which subjects apply it are rendered as logically consistent. This second constraint implies that subjects must be disposed to assent to the short consequences of intolerably long sentences. The pattern of assent to such consequences requires the attribution of truth to such intolerably long sentences (barring other forms of intermediate storage) in much the same way that the empirical contents of sophisticated theories require the postulation

of objects, and both are subject to the same degree of under-determination by empirical evidence.

Modifying our account of translation in order to base translation upon dispositions to hold sentences true rather than to assent to prompting sentences does not solve the problems we are confronting in one fell swoop. While the relevance of subjects verdicts upon intolerably long sentences to the translation of their statements is recognised it is implausible that most human language users (a fortiori English speakers) generally hold true a sufficiently diverse range of intolerably long sentences to fix the translation of logical terms. Somewhat modified Quine's suggestion provides the saving grace. There are no steps that are obvious both in the sense that consensus as to their performance is easily achieved, and in the sense of obviously producing truth, but there are systems or devices for performing sequences of steps producing sentences about whose operation and outputs consensus may be achieved which all subjects[36] agree produce only truths.

Such devices are not mere philosophical fancies, the computer program used in Appel and Haken's proof of the four colour theorem is perhaps the most notorious modern example. That program may be regarded as producing as its output a sentence which goes, roughly, "If in map one region one is adjacent to region two and region one is adjacent to region four and . . . regions two is adjacent to region three . . . and in map two region one is adjacent to region four . . . and in map one region one receives colour one and region two receives colour two and region three receives . . . then no two adjacent regions receive the same colour" just in case it is true, and English speakers can be brought to hold that sentence true by consideration of the program which produced it. But this is not the only such system, there are theorem provers which do not produce just one but a stream of sentences all of which can be known to be true by virtue of being produced by such a program, and there are computational models which produce truths concerning particular places and times. Nor must such devices be computational, (distributed) social systems for deriving theorems also serve as systems for generating sentences to which speakers defer, and even oracles, horoscopes and other divinatory devices must be counted among the systems for producing sentences to which speakers defer. Any system for producing sentences, provided that it is deterministic (so there is a definite class of sentences it is disposed to produce as output) and speakers can bring one another to hold its outputs true will count. By means of such intellectual prosthetics the limits of human comprehension may be exceeded, even though the devices themselves are humanly comprehensible (though perhaps at more than one remove). On this point we may be concerned that we are tacitly relying

---

[36]or at least subjects to which others defer

upon some intentional/semantic notion, for it is those speakers who understand how such devices produce their outputs who hold them true in the first instance (though other speakers may be prepared to take their word for it), but we are not concerned with why subjects hold such outputs true but only with the fact that all speakers can be brought to hold them so. It is worth noting also that speakers do not simply specify the rules according to which systems operate, for those rules are subject to the same mis-interpretations as their other sentences, it is the dispositions to defer to actual devices for producing sentences and not any capacity for describing their operation which offers additional evidence upon which translation may be based.

If the contemplation of the outputs of such machines only added to the stock of sentences subjects held true and never lead them to revise their prior verdicts, then matters would be comparatively simple. Instead of taking a subject's theory to be composed of just the sentences to which he is disposed to assent, we might take it to be composed of the sentences he holds true together with the sentences produced by the devices to which he defers. But matters are not so simple, for subjects are apt to revise their pre-existing verdicts on sentences in the lights of the outputs of such devices.

We have (hopefully) sketched how we may expand the grounds of translation to include subjects' verdicts upon intolerably long sentences, but we have not yet attempted to state which of the various classes of statements that subjects can be brought to hold true should count as expressing their theory, nor to sketch a justification for putting aside some subjects' verdicts on sentences on the grounds that they are mistaken. The most naive attempt would be to try to delineate some definite method for rectifying errors. One example of such an approach is to suppose there is a class of sentences that a subject would hold true after a sequence of exposure to the outputs of the devices to which he defers and will continue to hold true through any further exposure and take the subject as being disposed to hold those sentences true. But such approaches are misconceived, there is no unique method for resolving errors for there is no unique method of expressing subjects empirical expectations, and no guarantee that errors can be resolved in any finite number of steps. Attempts to distinguish individual mistaken assertions/verdicts on sentences truth values combine an assessment of error together with one or another arbitrarily selected program for its resolution. While linguists are required to resolve such errors in formulating their translations it is a mistake to insist, as the objection's proponents do, that unless there is some definite criterion by which mistaken assertions/verdicts may be identified then something has been left out of our semantics (and hence that semantics relies upon an unempirical component).

Orthodox attempts to meet the objection from common-sense seek grounds

which justify translating putative English speakers as expressing conjunction by "and" and controvert alternative translations. But the holistic standard of translation, applied to subjects' unfiltered dispositions to hold sentences true, can never render speakers as holding true logically inconsistent sentences (in the target language). Inconsistent theories include every grammatically correct statement and may be viewed equally as either compatible with every course of experience or none. But all or nearly all putative English speakers hold true sentences that are logically inconsistent in English and do not hold true every every English statement, and the empirical expectations they express by asserting the statements they hold true are compatible with just some courses of experience. Such speakers must be translated as holding consistent theories and thus their idiolects (in particular English logical terms) cannot be translated homophonically into English Hence the need to justify some filtration/modification/idealisation of subjects dispositions.

By contrast I urge that there is nothing wrong with translations of putative English speakers which do not translate (English) logical terms from their idiolects into English homophonically. So long as those translations render the speakers as speaking defective languages, that they are competent speakers of their own idiolect does not preclude there being good reason to revise their verdicts on statements truth even if their theory fits the evidence. The familiar claim that no-one's beliefs are consistent goes over into the claim that no-one speaks a flawless language, and revisions are still called for in either case.

This is then a further constraint upon the required notion of a defective language — putative English speakers who are commonly held to have asserted (English) inconsistencies must in general be rendered as speakers of defective languages in any translation based upon their raw, unfiltered dispositions to hold sentences true. If the logical framework of English is flawless we can also account for linguists rendering putative English speakers as expressing conjunction by "and" etc., as due to a general practice of translating communities of speakers as speaking a language which contains a flawless framework and comes nearest and near enough to being a good translation of each of the individual speakers idiolects. Such mother tongues need not be flawless in their entirety, but it is desirable that the framework be sufficient to allow speakers to resolve the languages other defects by themselves. Desirable, but not strictly required, if the rumours of prelogical people speaking languages which have no equivalent of English's logical terminology are credible.

Explication of "defective language" deserves treatment in its own right, and we shall turn to that task in part three. For now, again, our purpose in discussing these objections is only to show that they are not as conclusive as they are generally taken to be. Provided the requisite notion of defective language can

be clarified (and we can explain just what is wrong with such languages) the holistic standard of translation and the indeterminacy of translation which it implies may yet be vindicated.

4.6 The End of a Dream

Empiricists have traditionally had a hard time explaining knowledge of necessary truths. Theories that differ only in their ascriptions of necessary truth have exactly the same empirical contents and thus, if the only test we have of whether theories are true is their fit with experience, we have no sound basis for determining which of them are true. Hume's advocacy of the doctrine that experience cannot reveal necessary truths (or implications) is well known, but the doctrine can be traced back as far as Aristotle and is explicit in Ockam's work in the thirteenth century. Ockham relied upon faith and God to deliver us from ignorance of necessary truths. In his age divine revelation was still a reputable source of knowledge. By Hume's time however the shift towards a linguistic/conceptual explanation for necessity was already in progress.

From Hume onwards explanations of necessity in terms of conceptual truth/analyticity gained (unsteadily) in popularity and clarity, reaching high points in Wittgenstein's *Tractatus Logico-Philosophicus* and Ayer's *Language Truth and Logic*. So it was that philosophy stood ready for the identification of concepts with psychological posits or linguistic meanings (though, among other complications, a strict Wittgensteinian would hold that the identification could only be shown not stated). Such an identification was required if conceptual analysis were to do its job of revealing necessary truths, refuting skeptics, and allowing its practitioners to correct the errors of their less analytic peers. The movement lives on in so called empirical philosophy whose practitioners seek to perform conceptual analysis by questionnaire The traditional view was that psychology could at best be philosophy's hand-maiden; knowledge of necessary truths could be justified only by conceptual analysis, but psychology (or linguistics) might be called upon to establish just which concepts (or propositional functions) others express in their assertions. Ironically the proffered justification for allowing empirical investigations to trump (more traditional) conceptual analysis is that there is no sharp distinction between philosophy and science, a principle for which in turn the only proffered justification is holism.

As the argument from holism against the epistemic accessibility of analyticity shows the same fundamental problem of empirical access applies to analyticity (mutatis mutandis conceptual truth) as obtains against necessity in the first place. Theories that differ only in their ascriptions of analyticity to statements (other than logical truths) have exactly the same empirical contents, just as

theories that differ only in their ascription of necessity.

The problem was obscured by two factors, a diagnosis of our intuition of 'analyticity', and reductionism, the strong empiricist doctrine that sentences can be translated into terms referring solely to sense data. Our unchecked intuitions made the theory compelling, but it was the second factor, reductionism, which made analyticity respectable. While it was held that the warranted assent conditions of a sentence were the product of meaning alone it was sensible to hold that such meanings were determinable, as determinable as the dispositions to verbal behaviour of competent speakers. Holism implies that the warrant for assent to most individual sentences is affected not only by the evidence but also by the truth values other statements are held to have. Once this is accepted it is also clear that a firmly held belief in "Either $\ulcorner B \urcorner$ or not $\ulcorner A \urcorner$" has the same affect as conceptual entailment of B by A. The observable consequences of conceptual entailment may be faultlessly mimicked by collateral information.

Even in the most liberal languages considered in the accessibility argument logical truths were immune to revision without change of meaning. From this some proponents of analyticity/conceptual truth have drawn the conclusion that at least the logical truths are determinately analytic. In a sense this is innocuous, for there is no harm in using "analytic" as an abbreviation of "logically true", but it carries the suggestion that logical truths enjoy this privileged role because they span the correct range of possibilities. Without a justification of logical truths' immunity to revision on transparently non-modal grounds the suspicion of some surreptitious reliance upon modal principles will remain. The logical truths admit to perfectly clear and accessible delineation, they are the valid statements of quantificational calculus and there are obvious procedures for deriving them; but an explanation of why it is just these statements that are immune to revision (in English) is still owed.

Whereas the ancients, and many moderns, conceived of logic as the study of the necessary preservation of truth, our goal in choosing a system of logic is not to capture some class of putatively necessary truths as theorems but to achieve an optimal combination of flexible articulation and rigid connection. Logic in this view is an inference engine to drive our scientific theories, a clarified medium in which to embed the connections between our observations and our predictions. The logician cum linguist's task is to distil such a constructive medium from the hodge-podge of ordinary language

Ideally such a medium would be finely enough articulated to capture any arbitrary class of independently contentive statements as theorems, and rigid enough to admit of an obvious procedure for determining of any pair of sentences whether one implies the other. It turns out that we will not achieve these goals. The first is unachievable outright, the best we can hope for is an

articulation fine enough to capture any recursively enumerable set of formulae (any set which a computer can be programmed to recognise); and even this is incompatible with the second desideratum. A wider range of theorems and predictions can be captured by obvious procedures for enumerating implications than by procedures which provide definite decisions.

Some objections to quantificational calculus's adequacy are immediately avoided by this approach. We are inured, for example, to the observation that there are models of ZFC (Zermelo-Frankl set-theory plus the axiom of choice) in which the continuum hypothesis is true and others in which it is false. The continuum hypothesis is generally held to be either necessarily true or necessarily false (if for no other reason than that it concerns only objects postulated in pure mathematics). Adrian Heathcote takes this indeterminacy to show that "one of the models of ZFC is necessarily false and therefore cannot represent a possible world"(p.79), and though we may agree with his conclusion it casts no doubt upon our choice of logic. The failure to capture some desirable necessities *may* reflect an undue flexibility in our theoretical framework, but it does not of itself constitute a failure to capture some fundamental form of reasoning.

Such allegations presume that the choice of logical terminology must be based upon modal grounds; that if statement A is true in every possible world in which B is true then any adequate system of logic must have "either not $\ulcorner B \urcorner$ or $\ulcorner A \urcorner$" as a theorem. Such an approach conflates models with possible worlds, and any mismatch is viewed as a defect. Fortunately we need not treat the rows of truth tables of propositional calculus as specifying statements truth values in other worlds, nor the models of quantificational theories as depicting alternate possibilities in which the same predicate (with the same meaning one might as well add) has different extensions. The former may be explicated as showing the truth values of complex sentences formed by substituting various sentences for the simple schematic place-holders heading (some of) the columns of such tables. Some substitutes are true and others false but each has but one determinate truth value which is not imagined to vary between rows. While the latter may be taken as the assignments of various classes of satisfactants to various schematic predicates (and tuples to the relational predicates).

Tarski made this interpretational approach famous in his semantic definitions of truth and logical consequence. As his critics have noted however[37], though the shift to re-interpretation may avoid modal dependencies it does not vindicate the choice of logical terms exempted from such re-interpretation. The Löwenheim-Hilbert-Bernays theorem assures us that there is a sufficient range of actual objects available to to provide a falsifying interpretation for each invalid schema and a satisfying interpretation for each consistent schema, and thus that

---

[37](Sher, 1996)

only the logical truths are true under every re-interpretation of their non-logical components; But satisfaction should not lead to complacency. The match between logical truth and truth under all interpretations can only vindicate the choice of logical terminology if the range of interpretations were independently justified. Thus justifications of the choice of logical terms in the following mould (mutatis mutandis justification of quantificational calculus's theorems immunity to revision) are worthless:

> Roughly speaking, the completeness theorem of first-order logic is the basis for thinking that our reasoning is adequately captured by the usual formal rules. That is, the completeness theorem tells us that if some statement S about the objects in question cannot be proved from the given axioms using the known rules, then we can exhibit something that satisfies the given axioms and not S; therefore, there is no valid principle of reasoning whose addition to the known ones would let us prove any new consequences of the axioms.[38]

The critical defect in such defences (smuggled in by hedging between the two attitudes towards truth tables and interpretations) is the assumption that every 'something' which satisfies the axioms and falsifies the conjecture S is a genuine possibility. Taking alternate models as depicting worlds in which the same property has various extensions [39] there is no guarantee that these depictions represent genuine possibilities. Taking alternative assignments of extensions to predicates as mere re-interpretation [40], the substitution of some actual predicates for others, then we may be sure that an object exists which satisfies the axioms *but only under re-interpretation*.

It is obvious that no completeness theorem can justify the adequacy of quantificational notion. Truth functional logic is well known to be complete and thus either the excess inferential structure provided by quantification is superfluous or completeness does not guarantee that there is not in turn some yet stronger logical system capturing more sophisticated patterns of inference and allowing for more varied predictions of experience. At best quantificational calculus's completeness guarantees only that justice is done to those inferences hinging solely upon truth functions and the interplay of quantifiers and variables.

Traditional defences of logic have sought to justify the choice of logical terminology on a priori grounds. Such an approach is natural if the problem is

---

[38]This example was caught live, culled from the from the comp.theory newsgroup

[39]with regard to truth functions this amounts to regarding different lines of truth tables as attributing different truth values to the same proposition

[40]with regard to truth functions, this amounts to viewing different truth table lines as representing the results of substituting differently truth valued sentences for propositional placeholders

conceived as demonstrating a match between the bounds of possibility and the implications captured by a given choice of logical terms. As we have seen however, the lack of epistemic access to modal truths appears to preclude any such vindication of the choice of logical terminology. Thus traditional defences have been caught between relying upon epistemically inscrutable assumptions (i.e. modal intuitions) on the one hand, or having nothing against which to measure the expressive power of logical terminology. By contrast, viewing logic as an inference engine allows both these problems to be avoided. For while, on this view, the power of our logic is still measured against an independent standard, that standard is not provided by a range of possibilities but by the sophistication and the subtlety of human subject's expectations of experience. The choice of logical framework is not justified by modal intuitions nor any completeness theorem, but by the delineation of that framework's limits, the empirical justification of claims concerning the limits on the sophistication of the expectations of experience that humans can harbour, and showing that two are equivalent; i.e., that just those pattern of expectation subjects can harbour are capturable as the content of a theory relying upon (just) that logic. The chief difference between this approach and more traditional approaches lies in its open reliance upon empirical evidence. In a "first philosophy" which sought to justify our knowledge of logical truths such reliance would be misplaced, but the new approach does not seek to provide any special justification for our knowledge of logical truths over and above that they accrue as an indispensable part of our scientific theories and the reliance upon the plausibility of those theories may be acknowledged unashamedly.

Limits of quantificational calculus

While the limits of QC are well known, a brief review may help us assess their fit with what we know of human reasoning powers. The positive virtues we seek in a logic have already been mentioned, articulation fine enough to capture any arbitrary set of sentences as theorems and connections rigid enough to provide an obvious procedure for determining implications. In settling for QC we give up on both objectives: instead of a decision procedure we settle for a proof procedure for both validity and invalidity but none for consistency, and we can capture only recursively enumerable sets of formulae as theorems.

A proof procedure for QC (of which many are well known) allows us to determine, sooner or later, of any given valid QC schema that it is valid. However for any such procedure there are some merely consistent schema, rendered false only by some interpretations in a infinite universe, upon which it will never issue a verdict. Upon some of these schemata we may wait forever, always wondering

if our chosen proof procedure is just about to show them to be valid. For others we may be insightful enough to come up with falsifying interpretations, but lacking a general method for producing demonstrably false interpretations for merely consistent schemata some schemata will never receive a justified verdict unless we are more insightful than our algorithms.

Relaxing the requirement of a decision procedure for theoremhood increases the range of classes of sentences that may be captured as theorems, though limits still remain. There are uncountably many classes of quantificational formulae, but only countably many that are captured as theorems by any acceptable quantificational theory; for we stick at requiring a finite specification of the axioms and inferential rules of any such theory and that each theorem be reached after some finite number of (decidable) inferential steps from such axioms. Gödel's theorem shows this restriction excludes not only patternless higgeldy-piggeldy classes of objects whose absence might not be of concern to pattern loving creatures like ourselves, but also classes in which we are keenly interested such as that of the formulae (of say ZFC) expressing mathematical truths.

Not only is there a proof procedure for recognising the theorems of any theory couched in QC, but QC stands at the very limits of what such procedures can achieve. Any set of (consistent) formulae that is accepted/recognised by some proof procedure can be captured as theorems of some theory couched in QC (a fortiori any recursively enumerable class of stimuli sequences captured as empirical contents) and any substantive addition to the inferential machinery of QC produces an inferential system lacking either a proof procedure for recognising validities or invalidities.

Limits of humanly co-optable computational capacity

The assurance that QC marks the limits of proof procedures does not yet guarantee that it also marks the limits of human reasoning powers. Traditionally phrases of the form "a proof procedure for S" have been explicated as "a step-by-step procedure performable without insight or understanding by a human executor which halts just when given a member[41] of S as input". Our goal however, is to determine whether QC does justice to all the patterns of empirical expectations that humans may harbour and/or enunciate including those that may only be formed with the exercise of considerable ingenuity and understanding. In any case, when pressed on the point of just which inferences/predictions require ingenuity, we turn from such vague talk and offer alternative explications in terms of Turing computability. A proof procedure in this sense is a turing machine that halts just when given a member of the requisite class as input.

---

[41]Or more generously a referent to a member

Under this more rigorous explication of proof procedure our claim that QC marks the limits of proof procedures is rendered as the claim that for any Turing machine T there is a recursively axiomatisable quantificational theory whose axioms imply that one of that theory's predicates is true of just those inputs on which T halts, and, for any quantificational theory there is a turing machine that halts on just that theory's theorems; and it was the conclusion of the previous section that this claim is true. The remaining question, to which the current section is devoted, is whether our minds, or some other co-optable natural processes, outstrip the capacity of Turing machines and enable us to be disposed to make predictions that no Turing machine could enumerate.

The thesis that our computational powers do not exceed those of Turing machines is one of the theses commonly known as "the Church-Turing thesis". The weakest claim known under this ambiguous title is that no human can compute a Turing incomputable function (i.e. be disposed to respond distinctively to just the members of some non-recursively enumerable class). Unfortunately this claim is too weak for our purposes, for humans are incapable of computing even recursive functions and achieve the degree of sophistication in our linguistic performance required for quantification only by relying upon external processes (including groups of humans) to guide them. If we could rely upon a hypercomputer, capable of computing functions beyond the power of Turing machines to guide us we may be capable of speaking languages more sophisticated than quantificational calculus.

The strongest claim known as "the Church-Turing thesis" is that any physical system whatsoever can be accurately simulated by a Turing machine. Though this doctrine would obviously guarantee that quantificational calculus can do justice to the sophistication of our predictions it is most likely false. If physical space is continuous then our own solar system is too chaotic for any Turing machine to simulate precisely. Gravitational systems of more than three bodies are too sensitive to their initial conditions, vary too widely in their evolutions, and can too closely resemble one another for arbitrarily long periods of time for any device to generate accurate predictions by performing a finite sequence of digital operations. For such systems of such sensitivity to be co-opted into the project of generating predictions however, would require total and implausible precision in setting initial conditions, performing the computation, and reading off the results. Thus though such systems falsify the strongest Church-Turing thesis, they cannot be co-opted into the project of decisively determining membership in non-recursive classes, a fortiori that of deciding which sentences are theorems of logics substantially stronger than QC.

The claim upon which the defence of QC's adequacy rests is that any process that humans can co-opt into producing predictions could be replaced by some

turing machine without altering the sentences which subjects relying upon that process are disposed to hold true. This claim could only be fully vindicated by a accurate and complete scientific theory, or at least an in principle sketch of one, couched in quantificational notation. Given a quantificational theory allowing us to accurately and precisely predict our observations, excluding only those vagaries (determinably) due to chaotically sensitive systems, we could be as confident in the lack of any hyper-computational processes as we were in the theory itself; needless to say we do not have such a theory at the moment. In the absence of such a theory we face the difficulty known colloquially as "proving a negative" and must turn to more circumstantial evidence. To begin with many proposed hyper-computers, infinite speed machines, fair turing machines, analog turning machines, and such like are obviously physically unrealisable, and those that are not obviously unrealisable are generally based upon the most extremely optimistic assessment of what quantum computing can achieve, and all attempts to describe a quantum hyper-computer have thus far been found wanting. Secondly there is the incredible difference in computational power between turing machines and hyper-computers. If there were hyper-computers it would turn out that all of traditional mathematics, and indeed all human intellectual history was trivial, never having addressed a problem for which there was not a determinate method of producing the correct answer in a finite period of time; the time taken to produce a proof of Fermat's last theorem would be cause for black humour indeed. Thirdly, if hyper-computational processes were common it would be surprising that our limited quantificational theories succeed as well as they do, for the advantages in processing speed of such processes would lead one to expect that all the non-trivial generalisations capturable in quantificational notation would be drowned out by what, from our limited perspective, would be unpredictable noise. At very least this suggests that hyper-computation is rare or so (bizarrely) organised as to cancel itself out and let those generalisations that can be captured in quantification notation shine through. The final point, of which all the others are mere aspects, is simply that as far as we can currently tell our current best fundamental scientific theories imply that this moderate version of the church-turing thesis is true.

Logic, on the view espoused here, is the minimal toolkit sufficient to do justice to the connections between observation and prediction insofar as we can grasp them. All other inferential connections between statements can be mimicked or supplanted by logical implications, either under the guise of conceptual/logical analysis or by relying upon additional axioms which may be cited as collateral

information, but further reductions in the primitive inferential rules (a.k.a. conceptual implications) produce systems incapable of doing justice to the empirical contents of our sophisticated theories. Logical truths are unrevisable because any language that doesn't contain statements translatable as logical truths is incapable of expressing our expectations, and only statements all speakers hold true (or are generated by obvious procedures to which they defer) can be so translated. In a sense logical truths may yet be said to be true by virtue of meaning alone, for they are indispensable to the meanings of the theories we profess; in much the same vein they are also self-evident, for the capacity to assert logical truths guarantees that there are true theories in which they are required.

The claim that quantificational calculus is logic, that the English statements that remain true under re-interpretation of their terms other than "and", "or", "identical to", "everything", "something" etc., are the logical truths, hinges upon the moderate version of the Church-Turing thesis mooted in the previous section. If that thesis is false and our scientific theories or patterns of argument may be based upon inferential patterns captured only by some stronger system, such as Hintikka's independence-friendly logic, then that system would count as logic after all. In such case the instances of valid quantificational schema would yet be logical truths, and quantificational calculus would survive as a fragment of a larger whole just as truth-functional logic or propositional calculus is a fragment of quantificational calculus; but this need not be so, for if the moderate version of the Church-Turing thesis is false then there there may a more flexible logic, which does not have any aspect of quantificational calculus as a fragment. In computational terms there would be an irreducibly distinct alternative gate (presumably equivalent to an infinite number of logic gates based upon current technology) upon which our information processing devices could be based. Though such hyper-computational gates would be able to emulate the outputs of and/or/nand gates, but no gates performing the equivalents of such quantificational derivations would be required. Quantificational calculus would not be a fragment of this more powerful logic but a special case, standing to the superior logic as Euclidean geometry stands to geometry and not as propositional calculus stands to quantificational calculus.

If the moderate Church-Turing thesis is true however, then any chain of reasoning we care to engage in has its corresponding codification in quantificational calculus. The advocates of so-called logics for which there is no proof procedure tend to emphasise the superiority of their notation's expressive power, but the value of descriptive powers which are not under-written by determinable inferential machinery is doubtful. Indeed it is counter-productive to the goal of making the machinery by which our expectations are generated manifest in distinctive

inferential connections. If additional inferential connections are forged, at the price of giving up upon a recursively decidable proof procedure and enumerable implication, and the moderate Church-Turing thesis is true then only a recursive subset of such connections (isomorphic to some fragment of QC) will be of use to human theorists. Hintikka's independence-friendly logic, for instance, may be viewed as being based upon games of verification and falsification, but the strategies required for successful play of some of these games (and in terms of which 'true' is defined) are beyond the capacity of human players to execute (or defer to). If attention is restricted to those relations of independence-friendly implication that humans can determine, and the sentences that humans can derive according to its inferential rules the result is quantificational calculus; the extraneous super-structure is thus revealed as an invidious obscufication of our reasoning abilities.

To answer the question which instigated our investigation of logic and logical truths, if the moderate Church-Turing thesis is true then there is no language in which classes of sentences including the negations of the translations of instances of valid quantificational schema are assigned any empirical content. Hence, in particular, the liberalese imagined in the argument against the epistemic accessibility of analyticity, in which only translations of instances of quantificational schema are unrevisable, is the least restrictive language which speakers holding sophisticated expectations can be rendered as speaking.

Part III

Philosophy

Purpose

The major historical effect of Quine's attacks upon the analytic-synthetic distinction has been to popularise the belief that there is no sharp line between science and philosophy[42]. The etiology of this belief is not hard to trace; the rejection of analyticity(/conceptual truth) and externalist accounts of modality appears to leave philosophers neither a distinctive subject matter, nor a distinctive methodology for contributing to human knowledge. Thus, if they are making any useful contribution at all, it must be as part of a larger project within which philosophical contributions cannot be sharply distinguished. Taking science broadly as the pursuit of knowledge justified on empirical grounds, the empiricist doctrine of justification implies that all knowledge may be obtained by scientific means, and hence if philosophy produces any knowledge at all, it may be viewed as a part of the overall scientific project of producing knowledge. Currently the most popular view among philosophers is that philosophy and science are complementary aspects of the pursuit of the simplest and strongest theories fitting with experience, differing from each other only in degree and not in kind.

In a final bizarre twist the belief that there is no sharp distinction between science and philosophy has played a critical role in the re-acceptance of externalist accounts of modality in current philosophy. For there really is no sharp distinction between science and common-sense, and thus philosophy came (under David Lewis' influence) to be regarded as the project of simplifying and systematising the stock of existing beliefs that held prior to engaging in philosophy. The goal of philosophy has thus come to be widely taken as that of finding simple and minimal system(s) of principles implying maximal swathes of pre-philosophical beliefs. On occasion simplicity mandates revisions or excisions but the purpose of these modifications is to maximise the range of pre-philosophical beliefs captured rather than, as empiricists hold, to produce the simplest overall system that squares with experience. Havoc in libraries is to be avoided on principle. This stock of pre-philosophical beliefs includes modal beliefs (such as that nine is necessarily prime and only contingently the number of planets in the solar system) and these beliefs too are to be simplified and systematised as best they can, hence the attempts to develop modal logic rather than live with modal skepticism.

The purpose of part three is to show that there is a perfectly sharp distinction between philosophy and science, and to deliniate the scope for serious and productive work on the side of philosophy compatible with holistic empiricism.

---

[42]This was pointed out to me by Micheal McDermott in private correspondence

Chapter 5

Distinguishing philosophy and science

Quine's argument from the lack of an analytic-synthetic distinction to the lack of a sharp distinction between philosophy and science is fairly straightforward and is presented most explicitly in the second and third last paragraphs of *Two Dogmas*:

> Ontological questions, under this view, are on a par with questions of natural science. Consider the question whether to countenance classes as entities ... Carnap has maintained that this is a question not of matters of fact but of choosing a convenient language form, a convenient conceptual scheme or framework for science. With this I agree but only on the proviso that the same be conceded regarding scientific hypotheses generally. Carnap has recognised that he is able to preserve a double standard for ontological questions and scientific hypotheses only by assuming an absolute distinction between the analytic and the synthetic, and I need not say again that this is a distinction which I reject.
>
> The issue over there being classes seems more of a question of convienient conceptual scheme, the issue over there being centaurs, or brick houses on Elm Street, seems more a question of fact. But I have been urging that this difference is only one of degree, and that it turns upon our vaguely pragmatic inclination to adjust one strand in the fabric of science rather than another in accomodating some particular recalcitrant experience ...

The initial premise of this argument must be granted. The rejection of the analytic-synthetic distinction is a rejection of the claim that there are two

sharply distinguished types of sentences: those whose truth can only be determined with the aid of empirical investigation and those whose truth can be determined (by competent speakers) purely by unempirical considerations such as simplicity or form or meaning. Holism does indeed imply that whole theories have a "double dependence upon language and experience, but this duality is not significantly traceable into the statements of science taken one by one"[1]. From this premise the course of the argument is fairly clear: Just as individual sentences cannot be sharply divided into factual and non-factual, so too questions over the truth of individual sentences cannot be sharply divided into factual and non-factual, and hence (on the presumption that philosophical questions are non-[empirically]-factual) there is no sharp distinction between philosophical questions and questions that can be decided by empirical investigation.

Despite this the argument is fallacious. While holism does imply that there are not distinctively non-factual philosophical truths — that there is no distinctive realm of statements that can be known to be true on purely conceptual/linguistic grounds to serve as the subject matter of philosophy — it does not imply that there is no distinction between philosophical and scientific questions. While it is true that disputes over the truth of individual statements cannot be divided into scientific and philosophical disputes, this is simply because holism implies that disputes are (except perhaps for those over individually contentive statements) never over the truth of an individual statement, but always over which whole theory to hold; and although it makes no sense to say of two statements that they have the same empirical content it *does* make sense to say of two theories that they have the same empirical content. At the level of whole theories (the minimal contentive semantic unit according to holism) the distinction between scientific and philosophical disputes is easily drawn.

When all that is known of a dispute is one of the statements in contention it is impossible to determine decisively whether the dispute is philosophical or scientific. By itself, the knowledge that two (English speaking) parties disagree over the truth of "There are brick houses of Elm Street" justifies the suspicion that their dispute is empirical (i.e. that there is some stimuli sequence which is compatible with just one parties theory), but the suspicion is defeasible. If the dispute also concerns the existence of physical objects in general, or the legitimacy of street naming authorities, then it may well be that the two parties hold empirically equivalent theories and their dispute is purely philosophical.

Likewise, in the other direction, knowledge that two (English speaking) parties disagree over the truth of "There are classes", can justify the suspicion that their dispute is unempirical, but again only defeasibly. We who know of the plenititude of alternative constructions rendering the same inferential articulation

---

[1](Quine, 1953, p. 42)

as sets, tend to suppose that a denial of the existence of classes will be accompanied with some alternate suggestion rendering an overall theory with the same empirical content. However, the discovery that the dispute also concerns the computational complexity of the world's dispositions to provide stimuli, for instance, might show the dispute to be amenable to empirical resolution.

In each case so long as disputes are described only as being over a particular sentence, no clear distinction can be drawn between empirical and unempirical disputes. But the apparent imprecision of the distinction is due to the partiality of such descriptions. No argument can be solely over the existence of brick houses on Elm St, other claims must also be revised, at very least either logical truths or the sentences logically connected with "There are brick houses on Elm St". When such inclinations to revision are fully spelt out, when the protagonists of a dispute lay out their full theories, the distinction between empirical and unempirical disputes (or, to use Quine's terminology, factual and non-factual disputes) is perfectly clear even by Quine's standards. Scientific disagreements occur between theories differing in their empirical content and are amenable to being settled by empirical investigation, philosophical disputes occur between empirically equivalent theories and no empirical investigation can resolve them.

5.0.1 Causes of Quine's oversight and its acceptance

The popular acceptance of Quine's argument against a sharp distinction between philosophy and science is due, in large part, to the freedom from intellectual constraints that it easily lends itself to justifying; but the rationale for such laxity is produced by accepting the holistic account of empirical content while persisting in a non-holistic account of disputes and dispute resolution. The origin of Quine's mistake however, lies in a different quarter.

Micheal McDermott suggests that "Quine's failure to see a distinction between factual and non-factual questions may be a result of his focusing on a certain task of theory choice in which the competing theories are *assumed* to be factually equivalent. This is the task of the imaginary theorist who is somehow *given* the full truth about experience, and seeks the theory which best systematises the data."[2]. I agree with McDermott's diagnosis but go a little further in suggesting the motive for Quine's misplaced focus. The doctrine of the inscrutability of analyticity (and the indeterminacy of translation) is trivial if taken to apply only to partial evidence. It is no surprise to anyone that some fragment of the available evidence leaves analyticity, scientific theories and translation under-determined. Quine's striking point was that even the totality

---

[2]McDermott cites the third and second last paragraphs of "Two Dogmas" as evidence – see (McDermott, 1996)

of the evidence is not enough to determine the meanings of individual sentences. Since this is the key point on which the argument against the analytic-synthetic distinction hinges a continued focus on the slack left by the total evidence is natural, but taken too far it led Quine to recognise only philosophical disputes and ignore scientific disputes altogether.

Chapter 6

Resolving philosophical disputes

6.1 Philosophy of the don't cares

A trite summary of the state of philosophy described in the previous chapter is
that there are philosophical questions but no philosophical answers. But this
quip suggests that philosophical questions have no resolution, whereas the truth
is that answers to philosophical questions are not distinctly philosophical, but
may arise equally by a different route in response to scientific disputes. Quine
was right in claiming that there is no distinct line between philosophical and
scientific doctrines but vastly mistaken in when he said ontological *questions*
are on a par with questions a of natural science.

The advocated distinction between science and philosophy and science is
not only wholistic but dialectical. The truths discovered through philosophising
stand upon just the same epistemic foundations as those discovered by scientific
investigation. There are neither distinctively philosophical individual state-
ments nor even distinctively philosophical whole theories. This has lead many
looking at such theories and statements in isolation to suppsoe that there is no
distinction between science and philosophy and, in isolation, this is correct. It
is only when we consider disputes between proponents of disparate theories that
disputes which cannot be resolved by empirical means can be distinguished and
identified as distinctively philosophical.

This immediately raises the question of just how philosophical disputes can
be resolved and, at the outset at least, the propects look poor. The distinc-
tion between philosophy and science as drawn is only sufficient to guarantee a
philosophy of the don't cares. According to Quine philosophical analysis is a

matter of "supplying lacks", i.e. of assigning truth values to sentences that are inessential to our theories empirical contents. Having identified "the particular functions of the explicated expression that makes it worth troubling about" one devises a substitute fulfilling that function which is "clear and couched in terms to our liking"[1]. Quine's paradigmatic example is the analysis of the ordered pair, successful analysis preserves the principle that two ordered pairs are identical just if they have the same first member and the same second member while settling other less momentous truths concerning them in one convienient and clear way or another.

Being relegated to deciding the don't cares is a fate many philosophers will not accept. Among other things it makes applying for funding rather difficult. It is not simply that philosophers take themselves to be discovering important and fundamental truths, a don't care, in the sense Quine has adapted from digital logic, may be a statement whose truth value we are deeply interested in, but philosophers often take their opponents to be advocating theories that are not merely inelegant, but outright mistaken.

Before dealing with the decisive resolution of philosophical disputes directly let us look at just how much falls within the purview of a philosophy of the don't cares, and the range of dispute reolutions which even such meagre philosophy allows.

Merely Verbal

The most concilatory resolution of putative philosophical disputes, and one which is allowed even by the most minimalistic accounts of philosophy compatible with the science-philosophy distinction just drawn, is to show that they are a product of needlessly inflammatory translation.

At least sometimes this diagnosis is correct. Advocates differing in the verdict on some statements while holding the same expectations are sometimes brought into alignment with each other when some retranslation, at its simplest a reconstrual of predicates, produces agreement. A simple dispute over whether electrons are positively or negatively charged (vice versa protons) is of this kind. More seemingly substantial issues resolvable in like manner include questions of whether physical space is unbounded or bounded but increasingly compressed near the boundaries (a la Poincare), and whether the universe began at a definite moment in time or has existed forever.

Wittgenstein's treatment of solipsism is an example of such an approach being applied in full seriousness to substantial philosophical problems. "When I made my solipsist statement, I pointed, but I robbed the pointing of its sense

---

[1] This is quoted from p.520 of (Passmore, 1968) in which the quotations appear uncited

by inseparably connecting that which points and that to which it points. I constructed a clock with all its wheels, etc., and in the end fastened the dial to the pointer and made it go round with it."[2], of course one *can* tell time by such a clock if the face is free to turn with the hour hand by looking at the number that is uppermost on the face and compensating for the position of the minute hand (relative to the face) accordingly. However, if one continues to read such a clock in the traditional way one comes to the conclusion that it is always between twelve and one o'clock, without re-translation it appears that there is dispute over the time such clocks display.

It is unsurprising that there is scope for such minimalistic resolutions of philosophical disputes even in a philosophy of the don't cares, since in essence the resolution consists of showing that such disputes are illusory in the first place.

Mutual Incomprehension

Another resolution which superficially appears to be less symmetric is to explain away philosophical disputes as holding between the proponents of mutually untranslatable theories. Quine put this ploy to work in *Logic and the Reification of Universals* in considering a protagonist who "professes to repudiate universals but still uses without scruple any and all of the discursive apparatus which the most unrestrained of platonists might allow himself". If such a protagonist should reject the quantification standard of existence then "Legalistically his position is unassailable" but the caveat is that the resulting position is also incomprehensible. "It is scarcely a cause for wonder that we should be at a loss to say what objects a given discourse presupposes that there are, failing all notion of how to translate that discourse into the sort of language to which 'there is' belongs"(Quine, 1953, p.105).

In giving up so thoroughly on translation one also gives up on disagreement. The result however is symmetrical; just as the pseudo-antiplatonist's discourse is untranslatable into the language to which "there is" belongs, so too (at least some) statements in that language are untranslatable into the psudo-antiplatonist's terms.

Wittgenstein took this approach when he denied that there can be philosophical theses[3]. While it is strikingly odd to see this agreement between Quine and one of his favourite scapegoats, this consonance is the product of diametrically opposite reasons. (Early) Wittgenstein takes concepts so seriously that conceptual error is impossible[4], we cannot say of an illogical world how it would

---

[2](Wittgenstein, 1958, p. 71)

[3]See (Wittgenstein, 1922, 4.112)

[4]Later Wittgenstein takes it that *anything* impossible is a conceptual or "grammatical"

look and nor can we think it; whereas Quine is so skeptical of concepts that he has no standard of conceptual error by which to judge certain inferences as illicit, a variation in usage which is not accompanied with empirical disagreement is grounds for preferring an alternative translation. For both Quine and Wittgenstein illogicality is a myth.

6.1.1   The lack of decisive resolutions

While both the approaches to resolving philosophical disputes just outlined are compatible with the most minimalist philosophy they give no insight into how philosophical disputes may be decisively resolved in favour of one party and against the other. The results of the resolutions just considered are symmetrical. While the existence of such disputes is implied by holism (via the under-determination of scientific theories) it would be better to speak of them as philosophical differences rather than philosophical disputes.

Though these two approaches do provide a satisfactory resolution of many philosophical differences, they do not provide resolutions in favour of one side or the other. An account of determinate criteria for deciding between empirically equivalent theories compatible with holism is still lacking. The identification and settlement of substantive philosophical disputes (such as over the existence of an analytic-synthetic distinction) remains beyond us and the nature of reasoning (or rather of being unreasonable) and how it is that philosophical disputes can be settled by arguments from common ground is yet mysterious.

Even without such resolutions there are still some substantive issues falling within the purview of philosophers that can be settled. Philosophers may demonstrate that theories are empirically equivalent, or, more strongly that two theories can be rendered as each other by a suitable reconstrual of terminology[5]. Once interest turns to theories in their own right, independently of their truth, a range of interesting questions come to light. Whether or not a theory requires the full power of quantificational calculus or can be expressed with the more limited resources of truth functional logic (as can theories which can be viewed as positing only a finite number of objects), whether a theory involves the positing of sets (as Boole's predicate calculus is sometimes mistakenly thought to), and what the postulation of hidden objects contributes to a theory. All these are questions philosophers quite properly pursue even though they are not guaranteed to be without empirical import. Such issues do tend to be philosophical, it is unusual for the logical requirements of a theory to

---

truth, hence for instance his interest in Moore's paradox.

[5]Both these relationships hold between scientific theories treating of standard Newtonian physics in an unbounded uniform space and those treating of Poincare's bounded but distorted space which are otherwise alike.

mark an empirical difference between protagonist's theories, though it can happen (perhaps mainly in electrical/computational engineering). Even when such questions carry empirical import they retain an unusually close connection with philosophy, for they are what remains of philosophical concerns once interest in the truth of a theory is put aside.

Even of the tepid issues that remain once concern over the (first-order) truth of theories is excluded, many are beyond the capacity of any mechanical proof procedure to resolve. The empirical equivalence of some pairs of theories can be proven, but there is no guaranteed method of producing proofs.

The connection between philosophy and the study of theories is particularly close because the contribution someone knowledgeable in the study of theories would be specially suited to making when concerned solely with *true* theories is just that of resolving philosophical disputes. Such a theorist would have no special expertise in discerning the correct empirical content required, for that she would be best off deferring to empirical scientists who might in turn defer to the theorist in the development and choice of a suitable logical framework to express that content. In the matter of selecting among those theories having the same empirical content however, the theorist's talents in discerning the relationships among theories may come to the fore. She might attempt to discern the adequate theory which had the fewest primitive predicates or the fewest types of primitive objects that must be recognised to realise the scientist's aims. Of course this is not likely to happen in such a cut and dried fashion, interaction between theorist and scientist is likely to be considerably more messy, in the best case subject to ongoing feedback between the two. As more sophisticated theoretical tools are mastered, they are apt to be used in capturing more sophisticated scientific expectations, and as expectations grow more complex and detailed they are apt to call upon a theorists ingenuity in devising new tools to suit.

6.1.2 The need for decisive resolutions

Though some of the contributions philosophers might make to scientific theories without the capacity to decisively resolve philosophical disputes have been outlined, it is not yet clear how it is that philosophers manage, even if only rarely, to resolve their disputes. Criticism of one's own and each other's work is part and parcel of philosophical practice, and can be traced as far back at least as far as Anaximander's criticism of Thales' doctrine that water is the source of all things. The nub of his criticism was that the elements are "in opposition to one another — air is cold, water is moist, and fire hot — and therefore, if any one of them was unbounded, the rest would have ceased to be by this time"[6]. While

---

[6]*Physics* 204b 204 quoted from (Allen, 1985)

that criticism may appear, by our standards, to be scientific (i.e. reliant upon the observation that elements other than water yet exist) recall that our standard of a philosophical dispute is dialectical. If, as we may reasonably presume, Thales did not believe that air and fire had ceased to be then it is reasonable to suppose that the disagreement between Anaximander and Thales was not a scientific dispute.

Perhaps Thales did not agree with Anaximander in holding that the elements were in opposition to each other; if so Anaximander's argument proceedes from premises which Thales would not grant. But surely it is obvious this need not have been the case. It is entirely plausible that Thales did agree to all the premises of Anaximander's argument, and hence that Anaximander's argument is both valid and based upon commonly held premises. But by the standards of meaning that have guided the approach to standards of good translation relied upon in parts one and two of this thesis such arguments are deeply puzzling. Without the analytic-synthetic distinction we have no sense of conceptual error, and it appears that theories can only be marred by empirical error — a failure to remain squared with experience, can deface theories. In the case of empirically equivalent theories however, there is of course no space for empirical error. The same point can be put in terms of translation; the evidence upon which such translations are based are peoples' dispositions to verbal behaviour, the tendency to assert or deny some sentences in concert with each other and to assert or deny sentences under the influence of different stimuli. By the strict standards compatible with that approach to meaning Anaximander's disagreement with Thales is evidence of a failure of homophonic translation, difference of inferential connection implying difference of meaning (given the same empicial content overall), and there appears to be no scope for a decisive resolution in favour of one or the other.

A sense of this problem manifests in the complaint that Quine leaves no room for philosophy. Generally this complaint is dismissed on the grounds that there is no sharp distinction between philosophy and science, but we have already seen that that rationalisation is false. Without scope for decisive resolutions of philosophical disputes however not only does much of the tradition of philosophical practice make no sense at all but there is not even scope for reasoning (i.e. for good reasons to change one's mind other than empirical evidence). In particular there is not scope for arguments for empiricism, nor for holistic empiricism nor, most particularly, for the lack of an analytic-synthetic distinction. Putnam's original charge (cf. Outline) appears to be well founded and the positivistic approach that has been urged throughout this work appears to be self-refuting. This is all the more devastating as it was the extremely good arguments of the British empiricists and their analytic inheritors that lead to the adoption of such

positive views in the first place.

6.2 Some traditional suggestions for resolving philosophical disputes

6.2.1 Simplicity

The most widely suggested ground for settling philosophical disputes and choosing between empirically equivalent theories is simplicity. Indeed it is likely that this proposed solution has been so forward in some readers' minds (if I may be generous and use the plural) that they have dismissed much of the concern over the resolution of philosophical disputes as wilful rabble rousing.

Not only is simplicity currently widely supposed to provide (unempirical) grounds for choosing between theories but the scope for such appeals and the range of disputes resolvable by them has in recent times been taken to be increasingly wide. Even at his strictest Quine allowed appeals to simplicity to settle more than the don't cares. Following Popper's lead, Quine suggested that simple theories carry a richer range of empirical consequences and acknowledged this as just one of the virtues which made simplicity one of the criteria by which empirical investigators chose between empirically distinct theories. Under David Lewis' influence however the appeal to simplicity has been extended even further to allow for the determination of claims which are at once supposed to be worth caring about while supporting no empirical weight (such as, in particular, modal claims).

But this recourse to simplicity must offend against our empiricist mores. To be sure there is a pragmatic reason for preferring simpler systems — in them the derivation of predictions is an easier matter requiring less time, memory and calculation. But the supposition that simpler theories are (more likely to be) true and complex theories (more likely to be) false is a last a priori principle of reasoning which empiricism leads us to consider for rejection.

Before turning directly to the question of whether simplicity can properly play a guiding role in settling philosophical disputes, let us distinguish two subsidiary senses of simplicity which must be acknowledged as playing a proper role in theory selection and which, I suggest, have mislead many into overestimating simplicity's philosophical importance.

Scientific simplicity

Perhaps the major illegitimate cause of the wide spread of appeals to simplicity in philosophy is that appeals to simplicity are rife in the practice (if not the

publications) of practising scientists; and philosophers, keenly aware of their inferiority in producing knowledge, have sought to model their practice upon that of scientists.

In scientists' hands however simplicity is appealed to as a guide to choosing between empirically distinct theories. Attempts to codify this sense of simplicity have met with mixed success at best. Among the best known are Popper's account of simplicity as a sign or byproduct of empirical strength or contentiveness, Nelson Goodman's program for assessing the relative complexity of sets of extra-logical predicates[7], Harold Jeffreys' attempts to lay down governing principles of curve fitting [8], and Elliot Sober's application of Aikaike's theorem to problems of curve-fitting, the unification of theories and the positing of multi-factor etiologies.

The chief merit of Popper's account of the role of simplicity by empiricist standards is that it does not rely upon any *a priori* ranking of the simplicity of theories but rather explains the pursuit and production of simple theories as a reult of the attempt to produce theories rich in empirical content (i.e. compatible with the least range of stimuli sequences). If Popper's account were correct simplicity could be dismissed as a guide to choice betwen empirically distinct theories, for while simpler theories would be stronger (ceteris paribus), and the pursuit of simplicity a useful heuristic for producing strong theories, simplicity would cease to be an end in itself and could be dismissed as an in principle superfluous to the purposes of scientific theory production (though in practice it would remain a useful stepping stone).

Unfortunately such a happy reconciliation of empiricism and the pursuit of simplicity is not to be. Goodman has presented [9] compelling instances of pairs of theories (or rather hypotheses but we may imagine the background theory remaining fixed) of which the simpler theory is obviously weaker (i.e. compatible with a wider range of stimuli sequences).

Goodman's response was to develop techniques for assessing the simplicity of bases of theories extra-logical predicates. I shall not attempt to outline Goodman's sophisticated techniques here, but it is quite easy to state the empiricist misgivings regarding such project. Even though Goodman's techniques essentially depend upon the entrenchment or established prior use of extra-logical predicates to assess the simplicity of newly proposed theories, and are in this sense based upon the results of empirical science, they still produce an unempirically motivated ordering of theories on the basis of their simplicity. Simplicity, on Goodman's account, is a further unempirical constraint upon theory choice,

---

[7]See (Goodman, 1955) and (Goodman, 1959)
[8]See (Wrinch and Jeffreys, 1921)
[9]in (Goodman, 1961) and (Goodman, 1983)

justified (insofar as it is justified at all) on *a priori* grounds or taken as a brute fact of scientific practice. Such unempirical grounds for assessing the plausibility of theories are fundamentally incompatible with a thorough-going empiricism.

Much as Goodman's account involves an *a priori* ordering of theories (or their extra-logical predicate bases) on the point of their simplicity and a preference for simpler theories, Jeffreys' account of scientist's practice of fitting curves or functions to data points involves an *a priori* ordering of laws (or forms of laws) on the grounds of their simplicity (as determined by such factors as the number of adjustable parameters they contain) and assigns simpler curves a higher initial probability. Conditionalising on evidence may shift our confidence towards more complex laws, but the initial probabilities set the preference for simplicity. Jeffreys' account is considerably less sophisticated that Goodman's but again we need not concern ourselves with the fine details for such initial preferences must be taken either as brute and unjustified aspects of scientific practice or as justified upon unempirical grounds.

It is not difficult to formulate a rough and ready account of the empirical justification for preferring the simplest among empirically distinct theories and curves. As we noted in §2.2.3 complex mechanisms are more likely to come apart and hence offer more chancy explanations of the regularities of the systems they compose. Furthermore each component is liable to leave other traces of its existence than its contribution to the complex system of which it is a part (at a time) and one would expect those collateral effects to be independently detectable. In terms of curve fitting, the more adjustable parameters a function has it is (intuitively) less likely that they should combine to produce any one particular curve, and the more likely that different parameters should dominate (or at least be combined in different degree) given different values as the functions arguments.

While this rough and ready justification for preferring the simplest among empircally distinct alternatives is clearly plausible and empirically based, it is also clearly less than fully explicit. It appears that "simplicity" is a place-holder term, like "similarity" or "stability", and that just as it is "a very special mark of the maturity of a branch of science that it no longer needs an irreducible notion of similarity and kind"[10] so too a science in its full maturity will dispense with the notion of simplicity. Likewise, until recently it has seemed that justifications for preferring simpler theories of the kind offered in the previous paragraph are mere templates for fully explicit justifications which must be fleshed out with (commonly accepted) details of the particular theory empirical investigators propose to reform in the name of simplicity.

To come across a watch on a beach does provide a strong justification for be-

---

[10](Quine, 1969, pp. 137-138)

lieving in a watchmaker. But that belief is less justified if the beach in question is littered with watch parts (provided there is a natural non-intentional explanation for their presence) and still less if there are also (naturally occuring) watch constructing machines on the beach which mindlessly process whatever grist falls into their maw. Our assessment of the simplest explanation for the discovery of a watch on a beach, and the consequent justification for believing in a watchmaker, is critically dependent upon our background knowledge of the forces at play on beaches. Until recently it seemd that such specificity was inevitable — that empirical justifications for a preference for simplicity must be inextricably tied to the particular nature of the subject in hand. To confute simplicity with parsimony for the sake of a good quote, it appears that "The legitimacy of parsimony stands or falls, in a particular research context, on subject matter specific (and a posteriori) considerations. ...What makes parsimony reasonable in one context may have nothing in common with why it matters in another"[11].

In the last fifteen years however, Elliot Sober has brought a new, widely applicable, and thoroughly empirical justification for the preference for simplicity to the attention of philosophers. At its core Sober's proposal relies upon a framework for evaluating scientific models of which Akaike's theorem is the centerpiece. Aikaikie's theorem shows, given some very plausible and general assumptions[12], that the natural logarithm of the probability of the observed data being gathered given that a hypothesised curve is correct minus *the number of adjustable parameters of the curves family* provides an unbiased estimate of that curves (relative) predictive accuracy (i.e. how well it will fit further observations). Taking the number of adjustable parameters as a measure of a (family of) curves simplicity, Aikaike's theorem provides a striking, widely applicable, but unremittingly empirical justification for preferring simplicity.

Further elucidation of the fine points of Akaike's theorem and its application can be found in Sober's publications[13]. But some points call for special attention. Firstly, Aikaike's theorem frees scientific theorising from the need for an extra-empirical preference for simplicity, showing instead that the preference for simpler curves over those passing through the collected data points perfectly is justified by the evidence alone (i.e. that "the data tell you more than you may have thought"). Secondly, Aikaike's theorem does not only apply to curve fitting simpliciter but extends also to cover such matters as model selection and theoretical unification. Thirdly, the curve with the best estimated predictive

---

[11](Sober, 2006)

[12]That new and old observations are generated by the same underlying processes (i.e. that the curve being fitted does not change between data sets), that observational errors are asymptotically normal, and that the data sets are sufficiently large.

[13]including (Sober, 2002), (Sober, 1981), (Forster and Sober, 1994), and (Sober, 2008)

accuracy, as determined by Aikaike's methods, is to be preferred because there is good reason to believe that it lies closest to the true curve which reflects the underlying processes by which data points (or more generally observations) are generated. There is thus no reason for preferring simpler formulations of the same function, apart from the ease of applying Aikaike's theorem. (This point is easily overlooked since Aikaike's thereom applies in the first instance to comparisons of the predictive accuracy of the best fitting members of distinct families of curves. More complex curves, even if they are acurate, are members of families which contain better fitting curves because they over-fit the data.)

The scientific sense of simplicity is thus justified on purely empirical grounds, both in the sense that it is the data alone that justifies the preference for simpler curves with fewer adjustable parameters rather than an *a priori* principle, and in the sense that the preference for simpler curves is justified on the grounds that it leads to the selection of curves generating more reliable predictions of future observations. Obviously different formulas describing the same curve (say a polynomial and the corresponding fourier series in which the co-efficients are treated as adjustable parameters) will both be equally reliable predictors of future observations and Aikaike's theorem will provide no reason to prefer one formulation over the other other than the ease of application of the theorem.

The key point is that scientific simplicity is ground for choosing between empirically distinct theories (or models or curves). The empirical justifications for preferring simplicity simply do not apply to empirically equivalent systems for generating predictions.

The Marketplace of Terminology

Another sense of simplicity which does provide reasonable grounds, albeit superficial and altogether pragmatic, for choosing between empirically equivalent theories guides such choices through the almost entirely unconscious process of language drift.

One aspect of language drift is that commonly used words tend to be abbreviated. As Quine notes "the form of a word is affected by its frequency. The more frequent a word, the more readily it is expected; the more readily expected, the more erosion it is apt to tolerate and still be recognised for what is intended. But the shortening is apt to increase its similarity to other words, which, being only runners-up in the popularity contest, then get decked out with redundant suffixes to distinguish them; for people like to make themselves understood without excessive enunciation or repetition"[14]. This aspect of language drift can be viewed as being due to the operation of a market for terminology. The cost of

---

[14](Quine, 1990b, pp. 111-112)

long terms is extra time and effort spent pronouncing the sentences that contain them, but brevity is constrained by the human capacity to distinguish strings of phonemes. Given the doubtless idealistic assumption that speakers wish to communicate while minimising the time and effort they spend in doing so, the market for terminology will (in the long run, ceteris paribus) allocate brevity so as to minimise the overall costs of communication. Reallocation of brevity in response to social and environmental changes does not always occur by steady erosion and accretion and in some cases is so sudden that it produces a shift in the felt meanings of pre-existing terms. So it is that one of the most attention catching instances of the reallocation of brief English terminology in modern times is the shift in usage of the term "gay", though in this case the shift was (initially) accelerated by the corrosive forces of euphemism.

Another more sophisticated form of competition for vocabulary obtains between theories. Commonly used terms are not only apt to be shorter, but native speakers are already inculcated with dispositions to hold many statements containing them true or false under various stimuli and accept or reject inferences between those statements. In addition to a rich and fairly uniformly range of dispositions to hold sentences true extensive training from early childhood equips native speakers with a ready facility in making and assessing inferences between them. Pre-existing words in common use are thus hot property in the market of terminology and as sought after as keenly as apposite domain names for internet servers.

This aspect of language drift can also be viewed as a market which, given (another equally idealistic assumption) that speakers wish to communicate and justify their empirical expectations with a minimum of effort allocates vocabulary (in the long run, ceteris paribus) so as to maximise the ease and brevity with which useful predictions may be derived. In such a market, as in many others, the rich will get richer — inferential apparatus that is simpler, in the vague sense of offering the fastest, least taxing and least error prone derivation of commonly sought after predictions, will have its value compounded by being given the use of the most entrenched terminology and inferential connections. The most commonly useful inferential apparatus will thus come to be expressed in the shortest and best known terminology.

Simplicity, is thus one facet of pragmatic utility which in part determines which sentences are true in any particular speech community's language. The winners in the marketplace of terminology do not merely acquire the use of previously accepted infertial patterns but reshape them; the general adoption of a new theory is signified by change in the sentences that competent speakers are disposed to hold true and the inferences they accept as preserving truth. Still the claim that simple theories turn out to be true and cumbersome theories false

is only superficially true, for though cumbersome theories will turn out to be false in their own terms, they can always be resurrected by substituting new jargon for the commonly used terms they fail to capture. Such jargon will lack the ready associations of commonly used terminology but this is no reason to impugn the reformulated theory's truth.

While the high demand for the use of established words goes some way to explaining why philosophers are apt to emphasise the simplicity of their favoured theories, in the end all that is at stake is which inferential patterns get to be expressed by brief and well entrenched terminology and which are relegated to expression in esoteric jargon. Not only is this a matter beneath philosophical concern it is also better decided by the operation of the marketplace of terminology than by trusting to philosophers' intuitions as to which theories will turn out to be useful.

Recapitulation

Appeals to simplicity are thus subject to three apparently exhaustive objections. Either such appeals rely upon empirical contents of the theories they are used to evaluate, and so are of no use in resolving disputes between proponents of empirically equivalent theories, or they do not rely upon empirical content and are thus unacceptable by empiricist standards, or they render only a superficial resolution based upon the pragmatics of language use. Much the same objection applies, mutatis mutandis, to any criteria that may be suggested for selecting between empirically equivalent theories and we seem to be left without any hope for resolving philosophical disputes.

It should not be surprising that simplicity is so unsatisfactory a guide to resolving philosophical disputes, for it was required to serve as a smeared out vestigial analytic-synthetic distinction. As a guide to translation it did not guarantee a uniquely correct result, nor that there is in each particular language a unique best form of expression for any given empirical content. Where the analytic-synthetic distinction was supposed to neatly divide statements into factual and purely linguistic, simplicity was a matter of degree. But despite these concessions it was still to serve as a unempirical guide to choosing and modifying theories. Just as attempts to base accounts of concepts and translation upon the similarity of concepts rather than the identity of concepts do nothing to avoid the problem of finding empirical grounds for distinguishing concepts in the first place[15] neither does this more humble graded version of the analytic-synthetic distinction avoid the problems of epistemic access to analyticity. Of course epistemically accessible criteria for assessing (relative) simplicity

---

[15]See (Fodor, 1992a), (Fodor and Society, 1987) and (Fodor, 1998) among others

can be formulated but something stronger is required, viz., epistemically accessible criteria for simplicity capable of supporting the claim that less simple theories (among empirical equivalents) are false.

6.2.2   Inconsistency

Half-hearted versions of analyticity such as simplicity are unsatisfying in any case. Philosophers may settle for calling one anothers' theories inelegant or cumbersome, but it is allegations of outright errors such as inconsistency, absurdity or circularity that give the sport its purpose and joy. Indeed unless such charges can be be vindicated it is difficult to see what is left of philosophy at all. Protagoras once announced that he could argue convincingly against even an opponent who did not grant the principle of non-contradiction, but even if we accepted his sophistical bravado surely our sights are set higher than merely swaying the audience.

In particular, though Quine tended to cast his rejection of analyticity as a matter of eliminating needless complexity by excising a useless fragment of theories concerning language, that gain was only tangentially related to our argument against the analytic-synthetic distinction. The good argument of part two raised no objection to taking subjects as speaking languages in which the statements they generally and stably hold true are analytic. The point was that doing so gave putative analyticities no greater credence, nor any special resiliance to revision, beyond that conferred by the fact that they are generally and stably accepted by competent speakers. The rejection of such sentences and the shift to a language that is less (or at least differently) restrictive is of no greater moment than changing one's opinion on a matter of fact. The charge is that analyticity does not serve as a guide to truth (or empirical adequacy) and that conceptual analysis is worthless as a method of producing knowledge. The contrary claim is supposed to be inconsistent with the principles of holistic empiricism and it is this that enables us to argue from empiricism to the conclusion that flawed reasoning is not a product of conceptual error and philosophy, insofar as it is a source of knowledge, is not conceptual analysis.

However, the power of a logical argument to justify speakers in altering their verdict on its conclusion appears to depend upon speakers having previously held the argument's premises to be true and its conclusion false i.e., having held logically inconsistent beliefs. Such inconsistency is precluded by the account of translation our empiricism led us to and upon which the argument against the epistemic accessibility of analyticity was based.

The sole guide to translation we found compatible with holistic empiricism was the preservation of the empirical contents of whole theories and indepen-

dently contentive statements. Any theory couched in quantificational calculus that contains a contradiction contains every statement, and hence may be viewed equally as either compatible or incompatible with every experience. Speakers however quite obviously have distinct expectations about the course of their experiences; some may expect sugar in their tea but no one expects tea that is both jasmine and earl grey but not a mixture of the two. Hence translation of any speaker's assertions into a language built upon quantificational calculus must render them as espousing a logically consistent theory.

Insistence upon translating others as espousing consistent theories appears to leave no scope for poor (deductive) reasoning or its correction. Many anti-naturalists have eagerly leapt to the conclusion that no naturalistic explanation can do justice to rationality. In such accusations a semi-conscious arrogance can be sensed, a desperate hope that the power of reasoning will forever be a mysterious realm discoverable only through the deep cogitations of those who are specially wise. Philosophers have sought to pander to this conceit in a variety of ways, Sellars talks of a space of reason, Husserl insists upon the absolute validity of the principle of non-contradiction, and Kant postulated the vast and intricate machinery of intuitions and categories in order to support both analytic and synthetic a priori truths.

Even philosophers sympathetic towards naturalism have sought to account for conceptual error by providing an additional standard of correct translation beyond the preservation of empirical content or verification conditions. So it is that Ayer has proposed linguistic rules[16], and Brandom normative statuses[17], while Dennett presumes upon a prior standard of rationality produced from we know not where. Each can be viewed as an attempt to provide criteria of good translation capable of over-riding the preservation of empirical content.

Such over-riding principles, rules of proper linguistic usage or adherence to (or evaluation by) social linguistic norms, which obtain independently of subjects' dispositions to hold sentences true (in their idiolect) in the face of various stimuli, would allow the attribution of inconsistency. Homophonic translation of putative English speakers' terms "or", "and", "not", "there exists" and "for all" into a language in which these terms express the connectives and quantifiers of quantificational calculus (together with unvarying translations of their other terms) would doubtless render them as holding inconsistent statements true. But we lack a justification for preferring such translations to those that render subjects as holding consistent theories expressing the empirical contents that match their empirical expectations (as manifested in their dispositions to hold independently contentive statements true).

---

[16]See (Ayer, 1936)
[17]See (Brandom, 1994)

One of the most apparently plausible justifications for insisting upon translation that does not preserve empirical content is orthodoxy. Communication is facilitated where homophonic translation succeeds; so it is that we inculcate a uniform framework of patterns of assent and rejection of sentences into those with whom we wish to engage in regular communication, and correct those of our peers whose dispositions to assent to sentences vary so widely from the mainstream that homophonic translation fails. But hope of requiring speakers to hold certain sentences true or to use certain terms to express logical connectives and quantifiers on pain of breaking the rules upon which communication supposedly depends is forlorn, for communication is equally facilitated by expressing empirical content by the same theory among those that can express it in a shared language and again any line drawn between sentences that linguistic propriety requires to be held true and those are are merely known to be true by all competent speakers will be abitrary.

Even if such a distinction between analytic and synthetic truths could be drawn on such foundations the resulting view of reasoning and philosophy is appalling. Philosophers would be rendered as grammarians of inference, insisting upon adherence to a convention maintained solely for the sake of uniformity' and philosophical disputes merely the product of blow-hards trying to impose their own extensions to the orthodoxy by pomposity and force of personality. Reasoning rendered as nothing more than chanting in key or at best the imposition of brute authority. The defence that reason has been supposed to offer from the dictates of the powerful and the judgements of the mob would be just more of the same, and the consolation of philosophy a mere conceit.

We began by seeking for a method of resolving philosophical disputes, and have been lead to seek a criterion of defectiveness applicable to theories in isolation. It may appear that the goal has shifted, but the suggestion that a criterion such as inconsistency would allow theories to be judged purely on their own merits and not in comparison with one another is misleading. Gödel's theorem assures us that no consistent theory possessing the minimal complexity required to express the empirical contents of sophisticated scientific theories contains a statement asserting its own consistency. The consistency of such a theory may be a theorem of other theories, but these other theories' consistency may be doubted in turn. The best that can be established is relative consistency (i.e. that if one theory is consistent then so is another) and even this may elude demonstration for indefintely long. In general the best that can be shown is that one of the ways in which an inconsistency may be derived in one theory is blocked in another. The choice of one theory over another might be justified on the grounds that it is free from the particular defect besetting the alternative, but if we lack the capacity to conclusively demonstrate that the chosen theory

is not defective in some other way the justification for confidence in it can only be relative.

In order to leave scope for the resolution of philosophical disputes without relying upon imputations of inconsistency, and yet abide by our insistence that translation preserves empirical content it seems we must find some defect other than inconsistency to do the job. To face the question squarely then, what is wrong with defective theories?

6.3 Looking for a defect other than inconsistency

Analytical empiricists accompanied their rejection of externalist accounts of modality and other forms of mysticism with an explanation of our supposed knowledge of necessary truths based upon internal aspects of meaning such as concepts or linguistic proprieties and viewed philosophy as the project of clarifying our concepts or linguistic rules largely by determining the statements that must be held true for subjects' concepts to compose beliefs or for their beliefs to be expressed by their assertions.

Empiricism led analytic philosophy's founders to propose linguistic/conceptual foundations for necessity. In section two we argued, following Quine, that those proposals also lack empirical foundations and must be rejected accordingly. Still, the project of clarifying our assertions rather than seeking knowledge of some unempirical realm retains its appeal as the proper task of philosophy, for as empiricists we are driven to hold that there is nothing else for unempirical philosophical investigations to do. The question is whether we can make sense of this task, whether we can make sense of the project of clarification without postulating concepts and taking whole theories rather than individual statements as the fundamental content bearing linguistic unit.

6.3.1 Theories

In setting the problem up as one of distinguishing clear from unclear *theories* it appears we risk overlooking the defects of expression that we are seeking. "Theory" is a term of philosophical art, referring to deductively closed classes of fully interpreted sentences. Perhaps inexplicit expressions of empirical content are inexplicit by virtue of failing to be theories, either because they are not logically closed or because their interpretation is not fixed and hence restricting our attention to theories precludes the recognition of the defects we are seeking.

The immediate question is whether the class of sentences that a subject holds true can fail to be a theory. Let us deal with the requirement of logical closure

first. Even to talk of logical closure is to rely upon a translation of a subject's utterances into a familiar logical notation. As should be clear the requirement that good translation preserve empirical content will ensure that translation imposes our logic upon a subject's assertions since only classes of statements closed under logical implication are contentive at all in such purified languages. The myth of proponents of sentences that are not theories on this score (i.e. who use terms translatable as logical but do not defer to systems for producing sentences according to rules of logical inference) is simply a more complicated version of the myth of pre-logical people.

While the contraints upon good translation guarantee that the requirement of logical closure excludes no class of sentences a speaker holds true from counting as a theory, the requirement of full interpretation may be taken in so strong a sense as to exclude any class of sentences held true from being a theory. If a speaker's dispositions to hold sentences true are required to fix a unique interpretation of their sentences for the sentences they hold true to count as a theory then no such class counts as a theory, since holism implies the indeterminacy of translation and hence indeterminacy of interpretation.

To specify the objects over which a theory's variables range, and the tuples of objects that satify its predicates is to give a translation of the theory into the language in which the interpretation is specified, and the constraints the subject's dispositions to hold sentences true place upon acceptable interpretations are the same as those upon translation (i.e. the preservation of empirical content). The closest we can come to requiring full interpretation while avoiding recourse to fixing arbitrarily upon a particular translation into a particular background theory is that the empirical contents of the contentive classes is determinate.

A theory, on this view, is a class of sentences together with a language in which that class is contentive, and a language is a mapping from classes of sentences to classes of stimuli sequences. We have already seen that holistic languages will map more than one class of sentences to some contents, and the question that confronts us is whether there can be languages in which there is a decisive reason for rejecting some among empirically equivalent theories in favour of others. In order not to be distracted by philosophical fancies we shall restrict the question even further, for we shall only consider languages in which the empirical contents of each contentive class is just the composite of all the independently contentive classes it contains (i.e. the class of stimulus sequences compatible with each of its independently contentive members).

6.3.2 Defective Theories: First Impressions

Let us begin our search for an account of defective theories by considering a description and an example of clear intellectual malpractice.

A first sense of what we are looking for can be gained by contemplating the following confession by Marx to having an each way bet on the outcome of the India Mutiny of 1847/8:

> "It's possible that I shall make an ass of myself. But in that case
> one can always get out of it with a little dialectic. I have, or course,
> so worded my proposition so as to be right either way."[18]

Putting the matter in the unrefined terms of meanings and senses Marx has admitted to making an ambiguous statement (or statements) which in one sense implies (in combination with sundry other commonly known or easily established truths) that the India mutiny is a national uprising that the British forces will not be able to quell and in another sense implies (again in combination with other background beliefs) that the India mutiny will be successfully put down.

There is room for dispute as to the tone of Marx's admission but none as to its content. Perhaps he spoke in jest, was mocking other philosophers, or the misuse of dialectic by his fellow travellers. To demonstrate his statement was in fact an admission of his own intellectual malfeasance is not to our point. Whether or not he committed the admitted misdeed is debatable, but that it would have been a misdeed for him to act as he says is perfectly clear. For Marx confessed not to producing a sentence that was merely compatible with both the failure or the success of the Indian mutiny, he did not assert some irrelevant ambiguity, such as "Our mothers bore us", but instead was careful to word his proposition so as to be "right either way". His declarations failed to make clear the sequence of inferences by which he generates his predictions as to the course of future events and experience.

Though Marx's malfeasance (as we have supposed it) was a deliberate attempt at deception. The real concern is not with deliberate deceptions such as Marx's, but with unwitting reliance upon ambiguous statements. A subject who genuinely predicted the success of the Indian mutiny ahead of time, but who could only justify his opinion by means of an argument which turned upon ambiguities, and hence could be read as being "right either way", would have failed to express himself clearly. Either the means by which he generated his prediction lack a corresponding derivation in his theory, or though his theory does contain such a derivation he is unable to provide it upon request.

---

[18](Marx and Engels, 1975)

Ambiguities are common in everyday speech and writing. Sometimes, as in Marx's case, they are asserted semi-consciously to allow the speaker to pretend to a knowledge that he does not possess, or to present the appearance of agreement with parties on both sides of a dispute, but more commonly they are inadvertent failures in honest attempts to make clear the chain of reasoning by which speakers have arrived at their beliefs from their observations. Indeed ordinary speech about matters of any complexity often tends to be so unclear, the overt arguments so dependent upon suppressed premises and charitable interpretation, that it is difficult to isolate just where the defect lies. Clear cut and succinct ambiguities of more than passing significance to their proponents' theories tend to be produced by professional philosophers.

One philosopher in whose work ambiguities play a key role is Kuhn. Typically the ambiguities he relies upon hinge upon conflating beliefs with facts. For example he says that after chemists had adjusted their theories to reflect John Dalton's beliefs about the atomic structure of chemical compounds[19] "even the percentage composition of well-known compounds was different. The data themselves had changed."[20]. To suppose that chemists could, by the act of adopting or rejecting some chemical principle, alter the percentage composition of chemical compounds has widespread empirical ramifications (in combination with the rest of our theory of course). One would expect chemical research to become considerably more lucrative given that it could be used to affect the prevalence of precious metals and other valuable substances. But of course it is fairly clear that the empirical content of Kuhn's theory does not differ so radically from our own (or at very least, that if it does it is wildly implausible). The situation is saved by commonsense, on any careful reading of this passage the literal claim is so fanciful, so plainly false, that we (and likely Kuhn himself, when he is not in the grip of his philosophy) takes his claim to be that chemists' beliefs regarding the percentage composition of well-known compounds was different.

There is nothing wrong with using language in this way, dropping out "believes that" clauses, provided it is done consistently. One suspects of course that the tremendous interest in Kuhn's account of science stems from taking him as claiming that chemist's beliefs really do affect the chemical composition of compounds (and presumably physicists beliefs really do affect the speed of light), and the plausibility of his view comes from taking him as claiming that chemists' reviews of their theories affect their beliefs regarding the chemical composition

---

[19]In particular after they had mistakenly adopted as a law the principle that in distinct compounds composed of the same elements the ratio by weight of the elements to each other in one must be a small integer multiple of the ratio in the other. The idea being that a definite number of molecules of one element must be associated with each atom of the other element in any pure compound.

[20]Kuhn, 1970, p.135 The structure of scientific revolutions, 2nd ed university of chicago press

of compounds. Situations of this sort are not uncommon and have given rise to the philosophical quip that it is hard to increase clarity and maintain interest. The problem with Kuhn's formulation is not that he uses phrases of the form "X changed" to mean that "beliefs about X changed", in itself such alternative usage would simply mean that homophonic translation failed; that Kuhn was speaking a language misleadingly like our own. To take that as a philosophical defect in Kuhn's formulation would be to view philosophers as nothing more than enforcers of linguistic conformity. No, the problem is that Kuhn does on occasion want to talk about certain states of affairs and not people's beliefs about them and lacks a symbol to indicate which it is that he is talking about.

Just before the sentences quoted above Kuhn writes "Chemists could not, therefore, simply accept Dalton's theory on the evidence, for much of that was still negative."[21]. This is utterly uninteresting if if taken as the claim that some group of people (perhaps sociologists of science) *believed* that chemists could not simply accept Dalton's theory .... Non-scientists believe all sorts of odd things about scientists. But of course we do not take Kuhn as saying this, in guiding our translation we rely upon subsidiary guidelines about Kuhn's motivations and one of them is that Kuhn thinks what he has to say is interesting, and/or significantly affects our understanding of his topic[22]. Since statements about some sociologists' beliefs about chemists' reactions to Dalton are utterly irrelevant to our understanding of chemistry (unless accompanied by some claim as to the truth value of those beliefs) we do not interpret Kuhn in that way.

The problem with Kuhn's account of scientific practice is not his use of an aberrant form of expression, that homophonic translation fails, but his reliance upon an ambiguity. On occasion we translate statements of his of a form such as "X changed" into English as "X changed" and at other times we translate it as something like "Members of group Y believe that X changed". The guidelines for our preferring one or the other translation are, to a first approximation, to always prefer the former interpretation except when it would result in taking Kuhn as advocating a theory which differs wildly from our own in point of empirical content.

The key point concerning ambiguous statements is that they leave the connection between their proponent's observations and predictions unclear. An argument which relies upon an ambiguous statement which is, in one of its senses, compatible (according to its proponent's theory) with the falsehood of

---

[21]*ibid.* p. 135

[22]This is not to say that such subsidiary guidelines are some sort of adjunct to the behavioural evidence for translations we have considered thus far. It is simply that for want of more direct methods we must speculate as to the behavioural evidence we would elicit from Kuhn given that he thinks it worthwhile to devote his time and energy to writing these sentences down and then having his work published.

the argument's conclusion leaves it unclear just which subset of the combinations of statements from which the proponent is disposed to infer the ambiguity also support the arguments conclusion. Even if Marx had been sincere and openly declared a prediction on the outcome of the Indian mutiny, the reliance upon statements that would turn out to be "right either way" in the argument for his prediction would have left us uncertain as to the grounds by which Marx inferred that uprisings against imperialist forces would succeed or fail in general.

6.3.3   Ambiguities without senses

Our suspicion is that defective theories fail to make clear the steps by which their proponents generate their predictions. That while such theories contain the independently contentive predictions which their proponents espouse because of their observations, the derivation of those predictions from the independently contentive sentences they hold true because of exposure to a member of their empirical content (i.e. which they have directly observed to be true) is left unclear. In Marx's case we suspect that such obscurity was deliberate, the product of an attempt to appear to be able to predict more than he really could, in Kuhn's case it was the result of an unhealthy addiction to a philosophical doctrine of the incommensurability of scientific paradigms which would not survive clear exposition. But surely the more common instances are where well intentioned subjects fail to make their predictive abilities clear not through malfeasance or love of a cherished doctrine but simply through a failure to understand the nuances upon which their own ability depends.

The examples we considered were of particular arguments rather than whole theories, but the defects of such arguments would extend to the whole theories of which their statements are a part if they were the best derivations of their conclusions that can be offered in their containing theory. We have attributed such defects to reliance upon ambiguities in the derivation of predictions, but the term "ambiguity" is embedded in a theory of meanings and senses which is incompatible with our empirical holism. Naively put, a statement is ambiguous if it is unclear which sense it is used to express. The question thus appears to be whether we can make sense of ambiguity without the machinery of statement meanings or senses (reliance upon which is forbidden to us by our rejection of the analytic-synthetic distinction and synonymy).

Quine suggested that the distinguishing feature of ambiguous statements was that they could command both assent and dissent at the same time; "Well, yes and no" the subject sensitive to an ambiguity replies. But such a criterion does not avail us in considering those unwitting of their ambiguities. A person who advances some argument without being aware of their own inconstancy in

assenting to its conclusion when they are disposed to assent to its premises, may well not distinguish between, as it were, the sense in which an intermediate sentence in their argument follows from its premises and the sense in which the conclusion follows from it. While they may sometimes offer their argument as a justification for their belief in its conclusion and at other times assent to all its premises while yet denying its conclusion, their argument may be so lengthy or convoluted that they never adopt an openly ambivalent attitude towards any of its intermediary steps.

In any case our emphasis on holism and the rejection of individual sentence meanings should put as on guard against any individualistic diagnosis of ambiguity. Rather we are seeking an account of defective theory formulations for which a first step to their rectification can be the diagnosis of a particular statement as ambiguous. Just as there is a choice as to how to modify a theory in the face of recalcitrant evidence so too we would expect there to be more than one way to reform a defective theory. There may be a free choice as to which of several statements are to be treated as in need of disambiguation; or rather than treating a given statement as ambiguous some of the arguments in which it features may be treated as having a suppressed premise (a point we will return to later). What we required is not a criterion for identifying ambiguities, but an account of the defects which diagnoses of ambiguity (or enthymemesis) are the first step towards rectifying.

Observational inadequacy

We are seeking for an account of the distinction between those theories in which the connection between observations (as reflected in the independently contentive sentences held true/false because of exposure to stimuli in their positive/negative content) and predictions (in particular the independently contentive sentences held true/false whose content the theory's proponent has not yet been exposed to) is clear and those in which it is obscure. One might naturally suppose that one clear-cut manifestation of such obscurity at its extreme is under-determination. Where a language's independently contentive sentences are not finely enough attuned to register distinctions between observations that affect predictions of future observations, the grounds upon which those predictions are based will obviously be left unclear in the theories formulated in that language.

A language fails to reflect the distinctions upon which its speakers base their predictions just if it contains a statement S and (at least) a pair of stimuli sequences such that neither sequence is a member of the (positive or negative) independent content of S (i.e. whether a subject is disposed to assent to S in

response to exposure to either sequence depends upon the stimuli to which he has previously been exposed and the background beliefs they have induced), and for some range of background beliefs/prior stimuli just one of the stimuli sequences disposes subjects to assent to S and yet both stimuli sequences and each of their corresponding temporal parts are members of the empirical contents of just the same independently contentive statements.

Hopefully an example will make these conditions clearer. Suppose there are junglese speakers whose language includes a general term "red-crested-whatsit" which may be translated as true of birds of a particular kind, viz., red-crested-whatsits. Suppose also that there is some stimuli sequence SS_1 which defeasibly disposes a competent speaker to assent to "Red-crested-whatsit?"; for example a stimuli sequence occasioned by trekking through the jungle for a day in Red-crested-whatsit territory and then being confronted by a flash of something red moving in the distance. Only defeasibly for if the subject has recently had experiences of being through this patch of wood and spotting a clump of red-crested-whatsit feathers caught on a bush at just the right location exposure to SS_1 will not dispose him to assent to "Red-crested-whatsit". SS_1 is thus *not* a member of the positive independent content of "Red-crested-whatsit" in the subject's language, as whether it disposes the speaker to assent to "Red-crested-whatsit" depends upon the subject's background theory. Suppose also that there is another stimulus sequence SS_2 which is exactly like SS_1 except that the flash of red is of slightly (but of course discernably, for these are supposed to be distinct stimuli sequences) different hue[23] and while SS_1 and SS_2 (and their corresponding temporal parts) each dispose the subject to assent to the same independently contentive statements, exposure to SS_1 (given the right background beliefs/prior stimulations) disposes the subject to assent to "Red-crested-whatsit" while exposure to SS_2 does not (given the same background beliefs/prior stimulations). Perhaps exposure to SS_2 disposes the subject to assent to "Ever-so-slightly-off-red-booby" instead, but in any case it does not lead the subject to expect or report the presence of a red-crested-whatsit.

The defect in the subject's fanciful junglese is that it lacks terms for precise hues of red, in particular it lacks a colour term for the precise hue of red-crested-whatsits, and has (at best) only a more general term "Red" true of both red-crested-whatsits and other slightly but distinguishably differently coloured objects. Thus while the distinction between precise hues of red affects the formation of junglese speakers expectations and the predictions they are prepared to make, there is no term which has just those stimuli in which a subject is exposed to one of these hues among its independent empirical contents. This may

---

[23]This can be cashed out in terms of the impingement of photons of slightly different wavelength upon the external surface of the subjects eyes.

be diagnosed as an ambiguity in the junglese term that translates as "red", and such a diagnosis would be supported if junglese speakers were disposed to both assent to and deny "Red-crested-whatsits are red", or more plausibly to assent to both (the putative junglese translations of) "All red-crested-whatsits are red" and "All ever-so-slightly-off-red-boobies are red" and yet deny "All red-crested-whatsits and ever-so-slightly-off-red-boobies are red"; but failing such oddities it is more likely that this defect would be treated as the lack of precise terms for colour hues (i.e. red-crested-whatsit-red) and the arguments from "I saw a flash of red in red-crested-whatsit territory" (and sundry other supporting claims) to "There was a red-crested-whatsit present" treated as enthymemes.

One apparent shortcoming of the example is that it focuses upon stimuli sequences (SS_1 and SS_2) to which only a single subject may be exposed (for we have no standard of inter-subjective identity of stimuli) and one might suppose that this allows us only to diagnose a defect in that speaker's idiolect. Still, once we have undertaken to provide a translation (into English) for all junglese speakers the existence of even a single speaker to whom the distinction between red-crested-whatsit-red and other hues of red is significant is sufficent to warrant the charge that junglese is defective. While it is true that if only one or few speakers find the distinction between hues of red relevant to their prediction that will support the diagnosis of the defect as the lack of a statement to serve as an additional premise in their arguments rather than as an ambiguity of the term provisionally translated as "red", that the distinction is not relevant to some speakers (who do not have the same sensitivity to evidence of red-crested-whatsits' presence of their more jungle-wise linguistic peers) and hence the lack of a independently contentive term reflecting it does not constitute a defect in their idiolect does not free the linguist from having to treat junglese as defective in order to account for the verbal behaviour (in particular the predictions) of those who do find the distinction in hue significant.

Another concern is that it is implausible that the language of a community whose members attached importance to the presence of red-crested-whatsits would lack a independently contentive colour term whose (positive) contents included just those stimuli occasioned by the conspicuous presence of objects which are obviously of the same hue as red-crested-whatsits and those in (all the members of) which light of the same hue as that reflected from red-crested-whatsits under conditions which may, in hindsight, be called normal, impinges upon their eyes. Still, defective languages are no philosophers fancy, it is widely recognised that there are a range of subtle cues upon which humans base their estimation of one another's intentions and emotional states which they are sometimes sensitive to but unable to specify in words. Even though these cues can be described in English (and other sophisticated natural languages) the inde-

pendent empirical content of the attributions of these cues to fellow humans does not include all those stimuli in which those cues affect our predictions of one anothers' behaviour and hence, due to its speakers' lack of self-awareness of their sensitivities, English is observationally defective.

While we can make sense of a language that is observationally inadequate to expressing a subject's expectations, and it is plausible that there are such languages, a far more serious problem is that this is not the defect we set out to find. The problem is not that we set out to find a defect in theories and have instead found a defect in languages, for if a language is observationally inadequate to expressing a subject's empirical expectations so too are all the theories couched in that language. Rather the problem is that we are seeking grounds for choosing between empirically equivalent theories, and in particular for defects theories may possess that would warrant choosing other empirically equivalent theories in favour of them, while the flaw in observationally defective theories is that their empirical content fails to match their proponents' expectations.

A theory in a language which is observationally inadequate to expressing a subject's expectations will not merely fail to register some features of the observations the subject has already made, and hence fail to elucidate the grounds upon which her current expectations have been formed, it will also fail to reflect the expectations observations may lead to in future. Though a junglese speaking red-whatsit spotter may expect a flash of red-whatsit-red in red-whatsit territory to be followed up with clear sightings of a red-whatsit her observationally inadequate theory is compatible with a subsequent palpable lack of red-whatsits.

We were seeking an account of theories in which the connection between observations and predictions was obscure, but have described theories in which such connections are omitted altogether. The ground for finding observationally inadequate theories wanting is that they fail to express the empirical content of their proponent's expectations wheras we are seeking grounds other than empirical content for choosing between theories. To find such grounds we must turn elsewhere than to the range of independently contentive sentences a theory contains.

Inferential inadequacy

A second kind of defect a theory may suffer from is a lack of inferential articulation. Even if a proponent's language contains sentences with independent contents graded finely enough to reflect all the distinctions between observations upon which her predictions turn, that does not guarantee that the method by which she generates predictions from her observations is explicitly expressed in her theory. The steps by which she derives her predictions from the indepen-

dently contentive sentences reflecting her observations may lack any corresponding sequence of inferences in her theory.

The orthodox standard of explicitness requires derivations or proofs to proceed by recursive steps between sentences. The traditional justification for this standard is that it allows derivations to be checked decisively. Another justification, more explicitly reliant upon the Church-Turing thesis, is that the computation of recursive functions marks an absolute and non-arbitrary bound upon our computational capacities. To fix upon steps more elementary than the recursive would be too arbitrary, fixing irrelevant matters of how the requisite computational capacity must be implemented which might vary from person (or computational aid) to person. While to allow in inferential steps of greater complexity would guarantee that any theorist with our computational limitations must rely upon some form of storage which is not reflected in the sentences they hold true. For in that case much the same argument that was advanced in part one for the indispensibility of external non-sensory objects could be applied internally to the theory's proponents. Just as quantificational theories must postulate unobserved external objects to capture recursively enumerable classes of stimuli sequences as their empirical contents, so too for theorists to hold such theories they must contain internal objects to serve as storage for the derivation of recursively enumerable predictions. Provided subjects are capable of performing recursive operations upon sentences (or rather whatever internal objects and mechanisms underlie the state of holding those sentences true in their language), the inferences of a subject's language are recursively decidable, and they hold a theory in which their predictions may be derived by a sequence of such inferences (via intermediate sentences) from the independently contentive sentences reflecting their observations and the members of a recursive class of sentences they stably hold true (so long as their theory remains compatible with their observations) then no internal objects are required as storage of intermediate states beyond those required to implement the attitude of holding sentences true.

A theory is inferentially adequate just in case it fulfills the following four requirements:

First, the sentences of the language in which the theory is couched must each be composed of a finite number of distinguishable components and there must be a finite number of recognisable relationships these components can stand in to each other, such that fixing these relationships between such components fixes which sentence (if any) they compose. Obviously this is not required just for inferential adequacy, sentences must be distinguishable for speakers to use them as units of communication as well the ability to identify sentences is required for them to serve as the units of comunication their use both in communication

and the derivation of predictions. One way of satisfying this requirement is for a language's sentences to be composed of distinguishable words arranged in linear order. For the purposes of communication, it is sufficient that these elementary components and relationships be distinguishable by some means or other; to serve as the sentences of an inferentially adequate theory however their presence must be discernable without recourse to sophisticated theoretical assumptions or computations (lest the individuation of sentences itself provide scope for smuggling in complex means of deriving predictions which may themselves be left inexplicit). This condition is satisfied for a given subject if she can be taught to speak a language in which the reports of the presence of such components and their elementary relationships are independently contentive (i.e. if for each type of component, and each of the fundamental relationships determining sentence identity, there are stimulus sequences which are sufficient to assure the subject that a component of that type is present, and that components stand in that relationship).

Second, the combinations of these elementary components in these elementary relationships that are significant in the theory's language (i.e. a member of some contentive class of sentences) must be recursively decidable.

Third, the inferences in the language (i.e. the class of ordered triples[24] of sentences such that every contentive class of sentences containing the first two also contains the third) must be the closure of a recursively decidable subset of those inferences.

Fourth, the theory must be the closure of a recursively decidable class of sentences under its language's inferences containing some independently contentive statements, and its empirical content must be a simple composite of the contents of the independently contentive sentences it contains (i.e. compatible with just those stimulus sequences that are compatible with all of the independently contentive sentences it contains).

These requirements are not, in large part, novel. Barring the stipulations concerning the capacity to distinguish sentential components and the relationships between them determining sentential (type) identity, the first two requirements are familiar requirements upon recursive formal grammars; and barring the stipulations upon empirical content the last two are familiar requirements upon formal systems. This is of course no accident, for the recourse to formality was motivated by the same desire for explicitness.

Together these requirements guarantee that the derivation of predictions from independently contentive statements and statement previously held true requires no extra memory beyond that required to store an arbitrary number of

---

[24]The specification of triples presumes the language containing a term for conjunction, but can be recast in terms of tuples of sentences at the cost of some extra technical detail.

sentences, and no computational power beyond that to perform recursive combinations of elementary operations upon sentences corresponding to the language's inferences. A subject need rely upon no abilities (apart from those of being capable of holding sentences true and asserting them) beyond the recognition of the elementary components of sentences and the relationships determining the sentences they compose, performing recursive operations upon those sentences, and distinguishing their exposure to stimuli sequences in the content of the independently contentive sentences to realise the requisite dispositions to hold true each of the sentences of such a theory. In particular they would be capable of enumerating the independently contentive sentences of such a theory and hence determine if a stimulus sequence was incompatible with it.

Historically, the main purpose for which formality was sought was the clarification of mathematical proofs, and the theories proposed to capture mathematical truths as theorems were unchanging. The empirical theories of the sort we have been considering are subject to modification in the light of experience, and their proponents are exposed to new stimuli are thus hold new independently contentive sentences true. One might thus wonder if the constraints already mentioned are sufficient to ensure that the connections between compatible future observations and the consequent predictions are also clear. But this concern may be allayed by noting some further aspects of the requirement that the contents of any class of sentences are a composite of the contents of its independently contentive members.

Languages fulfilling the requirement of composite content will possess an inferential structure supporting proof by contradiction. If the inferential closure of a theory (contentive class) and a compatible independently contentive sentence S was not itself a theory of that language then clearly the original theory is incompatible with the stimuli sequences in S's positive content. But if the original theory did not include S's negation then its content was not the simple composite of its independently contentive members. In a similar spirit the inferential structure must support a deduction theorem, for if the inferential closure of a theory and an independently contentive sentence A includes a further inferential sentence B, while the original theory does not contain A x B, where x is a sentential connective such that the language supports every inference from sentences of the forms A and A x B to B then again the content of the original is not guaranteed to be the composite of its independently contentive members.

Apart from the constraints upon assignments of content to a language's contentive classes, the requirements upon adequate theories amount to recursive axiomatisability under (their language's) recursive rules of inference. What is outlawed are theories composed of a recursively enumerable smattering of sentences which are not tied together by chains of recursively decisable inferences

from a recursively decidable class of fundamental principles and observations.

Ambiguities revisited

Ambiguities, on this view are key points at which the derivation of predictions is marred by an indissolubly non-recursive pattern of inference, or rather points at which such connections may be restored. For while inferential inadequacy can (or could be, there is a caveat to follow) be identified outright without relativisation to any other language or translation, the localisation of such inadequacy into problems with this or that particular sentence (and the sentences in which it appears as a sentential clause) cannot. Just as, roughly speaking, meanings are indeterminate except relative to a manual of translation so too ambiguity being an over-loading of such meanings is likewise indeterminate.

Before proceeding let us refresh our minds as to the constraints upon good translation. A good translation, by our holistic standards, is a mapping from sentences of a language L to sentences of a language L' which preserves the empirical content of whole theories and (hence, for languages of the sort we have been considering where the empirical contents of each whole theory is the composite of the contents of the independently contentive sentences it contains) the independently contentive sentences it contains; i.e. a mapping from sentences of L to sentences of L' such that the image of each contentive class of $L(/L')$ under that mapping is a class of sentences of $L'(/L)$ with the same content. As a consequence of preserving content good translation must also preserve inferential role, for if every contentive class of L that contains sentences A and B also contains C, then every contentive class of L' that contains the translations of A and B will also contain the translation of C and vice versa.

To diagnose the inferential inadequacy of a theory T in language L as being due to the ambiguity of a number of statements $A_1, A_2 \ldots A_n$ and propose a resolution of these ambiguities is to propose a mapping from the sentences of L to the sentences of a language L' which fulfills the following conditions:

First, this mapping preserves the content of all the contentive classes of L which do not contain any of $A_1 \ldots A_n$. (For these theories the mapping serves as a translation.)

Second, the sentences $A_1 \ldots A_n$ are mapped to disjuctive sentences of L', "⌜B_1⌝ or ⌜C_1⌝ or ⌜D_1⌝ or ...", "⌜B_2⌝ or ⌜C_2⌝ ...", ..., "⌜B_n⌝ or ⌜C_n⌝ ..." respectively, where "or" is a sentential connective supporting (in L') the standard rules of (inclusive) disjunction (i.e. translatable into regimented English as "or").

Third, the mapping preserves the content of the inferentially adequate classes of L which do contain some of statements $A_1 \ldots A_n$, and the images of those

classes are inferentially adequate contentive classes of L' which contain the corresponding disjunctions but none of their disjuncts in isolation.

Fourth, unlike strict translations, the images of the inferentially inadequate sentences of L containing any of the sentences $A_1 \ldots A_n$ (including T) are not required to be contentive classes of L'. Instead the image of each contentive class C of L which contains some of the sentences $A_1 \ldots A_n$ is required to be the subclass of a contentive class of L' with the same empirical content in L' as C in L which, in addition to the members of C's image, some of the disjuncts of the translations of the sentences $A_1 \ldots A_n$ that C contains, some sentences that may be formed from members of C's image containing translations of $A_1 \ldots A_n$ as subsentences by repeatedly replacing any instance of a subsentence A_i with one (or a disjunct of several) of the sentences $B_i, C_i, D_i \ldots$, and all the sentences that may be derived from the combination of the newly introduced sentences and those in the image of C.

Fifth, at least one of the contentive classes of L' which contains the image of T and may be produced from it by the method specified in the previous requirement is inferentially adequate.

Inferential adequacy as a criterion for theory choice

It seems that this account of ambiguity is sufficient to account for the striking flaws of Kuhn's philosophy and Marx's political punditry with which we began our investigation. More generally inferential inadequacy looks to be precisely the kind of defect we sought as a basis for resolving philosophical disputes. It is independent of empirical content and hence, unlike observattional inadequacy which gives rise to inadequate attempts to express independently given empirical content, just one of a pair of empirically equivalent theories may be inferentially inadequate. Furthermore, though inferential inadequacy does not give grounds for supposing that the statements of inadequate theories are false, there is a real sense in which they are only a partial representation of the truth, for in them only part of the process by which we generate our predictions is made explicit.

The claim that inferential inadequacy is the sole defect which provides grounds for resolving philosophical disputes is obviously stronger, but still a good case can be made. Perhaps the simplest and most compelling point is simply that there appear to be no remaining defects to mandate a choice between theories that are both empirically and inferentially adequate. But of course this falls short of being an argument that there are no such grounds to be had. More positively, if philosophical practice is to result in the discovery of previously unheralded truths without reliance upon empirical investigation it must proceed by making the claims of empirically supported theories explicit. Like

the traditional understanding of philosophy as conceptual analysis, philosophy is on this view a matter of making our theories clear and explicit. Unlike conceptual analysis however there is no presumption of a unique and determinate "logical structure" implicit in our assertions. Philosophers have a wide freedom of construction in their attempts to provide explicit connections between the sentences of inferentially inadequate theories.

The defect in an inferentially inadequate theory formulation may be viewed as an excess of teleology over structure. The syntactic structure of such a formulation is capable of bearing its load, systematising experience, only if the total course of experience is already known. One must know the ends to which the structure is built, the empirical content it expresses, before one can adopt the appropriate means of translation. It is striking, and perhaps somewhat confusing, that teleology, the adaptation of means to ends, the very hallmark of intelligence also marks poor or unclear reasoning. But the situation can be understood thus — we seek in our theories to condense our intelligent foresight into a physical structure, to embody our insight into the world's workings into a system whose inferential connections do not require the further application of intelligence to determine.

Before moving on to consider the critical defect which renders inferential adequacy unfit to be a criterion for resolving philosophical disputes, there are three further points that deserve mention.

The demand of a proof procedure for deriving predictions from independently contentive statements observed to be true is not to be confused with the demand for a proof of those predictions. We rejected this latter requirement back when we considered the very strongest form of strong empiricism. The most elementary generalisation even if couched in a mixture of purely logical and observational terms does not follow deductively from particular observations. But a proof procedure is not a proof. We are not demanding that the desired conclusion must follow logically from the reports of particular observations but merely that the procedure be partially computable, and the path marked by steps that are (decidably) computable. Such derivations will rely upon a decidable class of laws in addition to the observed truths as the axioms from which derivations may proceed.

Next, though we have treated inferential inadequacy as being due to ambiguities and resolvable by their disambiguation, this is merely one of the ways in which inferential inadequacy might be rectified. Much the same scope for choice (and more) presents itself in the rectification of defects as is present in translation. We call a statement ambiguous if we wish to add supporting statements by splitting homonyms, and the arguments in which those putative ambiguities enthymematic when we wish to rectify defects by adding supporting statements

in some other fashion. The supposition that there is some fact as to whether an argument is enthemematic or reliant upon an ambiguity is of a piece with the myth of concepts. Having abandoned that myth our attitude must be that such attributions are not diagnoses of some underlying fact but alternative courses of treatment, alternative ways of constructing a chain of statements by which the statements of a defective theory may be derived.

Finally, if inferential adequacy is the criterion by which philosophical disputes are resolved then the well-known charge against Quine that he leaves no scope for philosophy is vindicated. We have already noted that Quine's presumption that theorists (and their interpreters) have access to the total evidence leaves no scope for scientific disputes, but the bizarre linguistic behaviourism of his works leaves no room for philosophy either. In *Word and Object* Quine claimed that sentences are solely connected to one another by the mechanism of conditioned response. "There are so-called logical connections, and there are so-called causal ones; but any such inter-connections of sentences must finally be due to the conditioning of sentences as responses to sentences as stimuli."[25]. Even unspoken intermediate steps are conceded only grudgingly; (transitivity of conditioning is supposed to result in such intermediate steps being skipped even though their omission deprives theories of the inferential articulation required to do justice to our sophisticated theories) and even where they are conceded Quine's bizarre adherence to *reductive* (!) behaviourism requires him to suppose they are inaudibly spoken. By contrast, inferential inadequacy requires there to be inferential connections between sentences which are an order of complexity beyond simple stimulus-response conditioning. On Quine's scheme there simply is no scope for the non-linguistic generation of predictions, no scope for anything to be left inexplicit in our theories, and hence no scope for philosophy at all as we have conceived it. It is striking that while many have complained that Quine's theory leaves no scope for philosophy, this has generally been attributed to his naturalism whereas the real cause of this occlusion, Quine's reductive behaviourism (and its amazing incongruity with his anti-reductive empiricism), has gone generally unnoted.

6.3.4 The fly in the ointment: Craig's theorem revisited

Despite all the appeal of inferential adequacy as the criterion for resolving philosophical disputes, there is an absolutely knock-down objection to its holding that role.

Trivial variations on Craig's theorem assure us that any inferentially closed and recursively enumerable class of statements in any language with the most

---

[25](Quine, 1960, p. 11)

minimal inferential sophistication contains among its members a recursive class of statements from which the rest may be derived. The members of any such class can, of course, be enumerated and provided there is, for each sentence S of the language in which they are couched, an orderable sequence of further sentences with which S is inter-derivable (each of which occurs in just one such sequence) the position which any member of the class occupies in that enumeration can be surreptitiously encoded by one of the inter-derivable equivalents in its corresponding sequence, the class of all such double serving equivalents is recursively decidable and from them the other members may be derived.

In particular any language which includes devices for expressing any one of conjunction, disjunction, negation or quantification is sophisticated enough to guarantee that all of its recursively enumerable theories are recursively axiomatisable. The problem is thus that none but theories in the most implausibly inferentially impoverished languages are inferentially inadequate. While inferential inadequacy remains a good reason to reject a theory, it does not help one whit in resolving realistic philosophical disputes since those are always between inferentially adequate theories.

In general the response to Craig's theorem has been to try to find a stronger criterion than mere recursive axiomatisability by which to exclude systems of axioms such as are produced by his technique. Systems built upon such axioms are said to lack coherence, or economy, or unity, or they fail to integrate the theorems they capture. For my part I suggest that whether we regard the arguments by which the members of a theory are derived as sequences of inferences or convoluted encodings of the number of repetitions of the conclusion in the axiom from which that conclusion may be drawn by repeated simplification is of no moment whatsoever. Provided a theory is completely explicit (i.e. there are recursive syntactic operations for determining its axioms and deriving its theorems) and empirically adequate there is nothing more to be desired of it. As a weak empiricist I have no truck with the replacement/elimination programs of which Craig's technique provides an extreme example provided the inferences hinging on the supposedly suspect terms are clear enough to allow Craig's technique to be applied at all.

I suggest that the moral to be drawn from Craig's theorem is that we must give up looking for an intrinsic property of theories to guide us in choosing between empirical equivalents. Between them Craig's theorem and the Church-Turing thesis assure us that every humanly believable theory in all but the most trivial of languages harbours among its members a recursive class of statements from which the rest may be derived by repeated application of recursively decidable inferences. The syntactic/inferential structure of our theories is thus always sufficient to support their empirical load one way or another. We are

not however, guaranteed to know how the load is supported — which recursive classes of sentences and patterns of inference are sufficient to generate just the members of a theory. Craig's technique is a method for converting one axiomatised theory into another (with the same empirical content and more restrictive terminology), but it gives no hint as to how to produce an axiomatisation from scratch.

If the only grounds for choosing between theories concern their empirical content and whether the means of expressing that content is explicit and we accept the explication of theoretical explicitness in terms of decidable syntactic operations by which all the sentences of a theory may be derived from a decidable class of axioms (a fortiori by which predictions may be derived from observations) as proposed in the last two sections, then the only defect that justifies rejecting just one of a pair of empirically equivalent theories is our ignorance of the (decidable classes of) axioms and inferences from and by which all its members can be derived.

It is too stringent to demand that theories contain terms true of just the (members of one class of) principles and patterns of inference by which its statements may be derived, and too stringent to demand that their proponents, such as the hapless slave boy from the introduction, be able to describe principles and inferences by which they would be able to derive the statements they hold true. But we can ask that those proponents be able to give arguments for the statements they hold true which rely upon only recursively decidable forms of inference and take only the members of a recursive class of fundamental principles as absolute premises.

The ideal remains to produce a recursively decidable class of statements and a recursively decidable class of inferences such that subjects whose abilities were limited to producing those principles and performing those inferences (or at least the syntactic operations corresponding to them) would be capable of deriving all the independently contentive statements of our theory. We are assured that any reasonable language contains raw materials sufficient to the task, but the question of just how it is to be done remains open.

It is striking that we can hold all the statements of our theories to be true without being able to axiomatise them, but all of us, with regard to some sentences or terms or other, are in the same position Justice Potter Stewart declared himself to be in with regard to "hardcore pornography" — "I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it ..."[26] (and presumably when it is accurately described to him too), but the philosophers among us do

---

[26]concurring opinion in Jacobellis v. Ohio 378 U.S. 184 (1964)

not share his pessimism or at least do not let it dissuade them the attempt to further define their terms. So too all of us know that the Gettier counter-examples are not instances of knowledge, and that some other instances of true justified belief are, but we are unable to provide the principles by which we make the distinction.

Potter's declaration of his inability also suggests by contrast what it is we want from definitions — to make clear the inferential connections of sentences that hinge upon their containing their definienda. A successful definition, whether regarded as a program for eliminating its definiendum or as an additional claim of mutual (material) implication between it and the definens, brings us closer to making the inferential connections with sentences containing the definienda explicit (i.e. of purely syntactic operations). We know that Potter relied upon his inchoate sense of propriety to guide him in his application of "hardcore pornography"[27], but where this sense is shared by the ordinary folk of his linguistic community (as one would hope) we may wonder why it does not count as fully explicit already.

One response is to rely upon quantificational calculus as our touchstone and insist by fiat that only the fundamental inference patterns hinging upon its terms count as explicit. As a concession to proponents of alternative logics we may even allow in those patterns of inference which may be described in a quantificational theory which takes letters and tuples of letters and tuples of those tuples as its domain of discourse, provided such a description is guaranteed to be provably true or provably false of any given tuple of tuples of letters, for that boundary is one not even the boldest advocates of alternative logics never overstep. It is clear after all that the basic patterns of inference upon which quantification calculus is[28] based are explicit, i.e. that they are recursively decidable upon syntactic grounds, and the assurance that quantificational calculus stands at the limits of computable proof procedures (cf. §4.6) together with the Church-Turing thesis assures us that any humanly performable pattern of inferences can be captured by some quantificational theory.

One flaw with this approach is that it leaves the requirement that the axioms upon which quantificational theories are based must be recursively decidable. Since we must thus rely upon the notion of recursive decidability in any case, why not omit the middleman and simply adopt as our standard of explicitness derivability from recursive classes of axioms by recursive inferences.

In any case, success in either form would show that the dubious terms were in fine shape to begin with. Here as elsewhere definition delivers both elimi-

---

[27]My own sense of propriety rails against using this term as an example, I had originally thought the key term was "obscenity", but let us not be thrown off a good example by prudery.

[28]or rather may be

nation and vindication. It is not the lack of a recursive class of axioms among the sentences we hold true from which the sentences containing dubious terms may be derived by recursively decidable inferences that marks out such terms as needing explication, but our inability to specify those axioms and inferences. The most commonly attempted way to provide such a specification is to provide a (recursive) function from the sentences of the theory to be explicated onto a recursively axiomatized theory whose basic inferences are already known to be decidable which maps all the sentences of the original theory to theorems of the target. Such reductions serve to convert our ability to specify the recursive inferences and axioms of the target theory into an ability to do the same for the source. A more direct approach is to specify the turing machines (or recursive functions) which accept just axioms and compute just inferences sufficient to derive all the members of the original theory. Even more directly we may produce (finite but obviously extensible) devices computing those functions themselves. Any of these approaches serves to show a method by which the sentences of our theories can be derived by purely recursive syntactic operations, and in that sense makes our theory explicit.

Philosophy, on this view, is the attempt to elucidate fundamental principles and forms of inference from and by which we can derive the independently contentive statements we hold true. To this end philosophers disambiguate statements, invent logical structure, accept novel forms of recursively decidable inferences, announce suppressed premises of enthymemes and otherwise flesh out our theories with various supporting statements; not because our theories lack the requisite inferential/syntactic structure as they stand, but because we do not know how to derive our theories members (a fortiori their independently contentive members) without incorporating such additions (whether as sentences outright, or as factors of the Gödel numbers encoded in our axioms) and the arguments which are found to be in need of such supplementation mark our best attempts to do so to date.

One virtue of this view is that it makes sense of the curious combination of close argumentation and unsettled conclusions that is characteristic of philosophical arguments. If the premises of such arguments were conceptual truths, discernable by suitable scrutiny of propositions or guaranteed by competent use of language, and the inferences by which they proceeded similarly guaranteed to preserve truth, this combination would be inexplicable; but the conclusions of good philosophical arguments are the offshoots of the premises and patterns of inference from and by which their proponents propose to derive the (independently contentive) sentences we hold true. The credibility of these principles and patterns of inference thus relies upon their capacity to support our theories. Not only is the demonstration that they provide such support literally incom-

putable (though this is not to say that there cannot be good reason to believe they succeed in doing so in particular cases) but the development of such systems is not a steadily incremental process, any partial attempt no matter how successful at supporting some fragment of a theory is liable to be overthrown upon attempting to widen its support or combine it with the support of other fragments.

6.4   Conclusion

Philosophy is the art of making our reasons manifest in language — making explicit which features of the world guide our actions and expectations and the steps by which they give rise to them. This task is not to be distinguished from that of systematising our opinions — determining the most fundamental principles and forms of inference by which they may be derived — nor from that of resolving disputes between empirically equivalent theories, for the system.

This account of philosophy shares a common emphasis on systematisation with David Lewis' account, but standing on firm empirical ground. Where Lewis viewed the business of philosophy as the systematisation of all (or nearly all) of our pre-philosophical beliefs, on this view the target of systematisation is the empirical content of our pre-philosophical theories (more particularly the independently contentive statements connecting them with their contents). As in other forms of empiricist philosophy the project is to save the appearances. By contrast Lewis advocated respect for all our pre-philosophical beliefs, tempered only by the pursuit of simplicity or systematic beauty. On the view advocated here the uneasy and unprincipled trade-off between these two counter-vailing forces is avoided since neither is given any weight at all. Any system capable of supporting the empirical content of our beliefs is acceptable, no matter how ugly or combersome, or how much it is at variance with common-sense, though if it does not pay its way in expediting the production of pragmatically useful predictions it may be relegated to expression in esoteric jargon. This is not to say that no heed at all is paid to our pre-philosophical theoretical beliefs. While no value is attached to the capturing of intuitions per se, our intuitions represent the best way we have of systematising our empirical expectations to date, indeed we may have no idea how to do so without them. Capturing such intuitions is thus a good heuristic for developing a fully explicit system capturing the same contents. If, however, in the attempt to develop such a system the rejection of intuitively plausible claims allows a wider range of our independently contentive claims to be supported, or obscured connections between them to be brought out, we may reject our intuitions without regret no matter how cumbersome or ugly the result. Furthermore intuitively plausible claims which play no role in

supporting our theories' empirical contents are to be excised, not merely on the grounds that they are superfluous, but because they cannot be systematically combined with the fundamental principles of empiricist epistemology which do provide such support.

Pure reason — the capacity to resolve disputes on unempirical grounds — is not, on this view, a mystical ability to perceive necessary connections, nor a sensitivity to social norms, nor a sensitivity to meanings, nor an unempirical sense of simplicity. Doubtless some of these abilities are what some people have meant to attribute to those they call rational, and in translating them we face the indeterminacy that error (or more generously, disagreement) gives rise to, for error can be distributed in multifarious ways. Still if we fix upon pure reason as a capacity that we actually do have then the only remaining candidate compatible with empiricism is the capacity to make one's thoughts manifest — to mimic the structure of internal belief fixation with a structure of external proclamations. This ability is not a form of perception nor a gift of divine revelation and is not developed either by turning towards the light of reason, or by the analysis of concepts.

This view of philosophy leaves no room for the criticism of theories on the grounds of reliance upon pre-suppositions for we can make no sense of a background range of absolute possibilities against which pre-suppositions are to be judged. However, our consideration of consideration of the role of ambiguity shows that such criticism need not be completely off the mark. The difference between ambiguity or enthymemesis, on the account offered here, and pre-supposition is that ambiguities are calibrated against those stimuli which produce a decisive verdict upon predictions which hinge upon those ambiguities. Against a pre-supposition made uniformly and consistently we have no objection unless it leads to empirical error. In general critics harp upon just such pre-suppositions as being the most entrenched and unquestioned. In doing so such critics confuse a scientific mode of criticism (applicable to scientific disputes) with philosophical criticism. If our current physical theories were a special case, say for instance the speed of light depended upon the strength of the gravitation field through which was passing and physicists had "pre-supposed" that estimates of the speed of light traveling within the earth's gravitational field held universally, then we might after discovering the true more general theory decry the pre-suppositions of its more limited predecessor. In doing so we would be relying upon the range of nomological possibilities supported by the principles of the new theory. Even before such a theory was discovered, some related discovery, say that magnetic fields were distorted by gravity, might give reason to question the "orthodox pre-supposition" that the speed of light was constant. We would not however have any grounds to object that the original

theory was defective, instead the sole objection would be empirical, i.e., that it made certain predictions that were not borne out by experience.

To some the loss of the illusion of a sensitivity to necessity or to meanings or concepts is distasteful, their addiction to strong reason preventing them from living happily on the comparatively bland diet of naturalism. Compared to their elevated hopes for the nature of reason the ability to make the grounds for one's beliefs and actions explicit may seem mundane, and indeed in the best sense it is, no super-natural equipment is required.

Though empirical holism has lead us to reject concepts from our ontology, the traditional account of philosophy most closely resembles the view advocated here is conceptual analysis. The main difference is that the target of analysis is shifted from the supposed conceptual structure of our assertions and beliefs to the inferential structure of our theories. Another difference is that the term "analysis" is misplaced, for the results of philosophical reform of a theory are not required to be unique nor somehow implicit in the original; instead it is openly admitted that such reform is a mtter of constructing an original contruction from clear materials supporting the inferential connections of the original.

Leibniz' dream of a *characteristica universalis* is a more distant antecedent of this understanding of philosopy. Gödel's theorem has put paid to hopes of producing a system in which all truths can be derived, and even for those truths that can be derived we must give up hopes of a decision procedure. Still the objective of producing a system in which statements can be derived, and the derivations checked, by a mechanical *calculus ratiocinator* without reliance upon intelligent insight remains. Unlike Leibniz we do not hope for a language in which "the characters and the words themselves would direct the mind"[29]. Translation into such a language would be a trivial task, wheras the whole difficult business of philosophy, as we have conceived it, is to translate our theories into languages for which we can supply adequate mechanical proof procedures.

Looking even further back in philosophical tradition there is a more than passing similarity to one aspect of Socratic elenchus. For like Socrates we seek to compare "the man and his words, and note the harmony and correspondence of them"[30]. However it is plausible to interpret Socrates as having an objective completely opposite to that served by philosophical reform on our account advocated here. Clearly the doctrine that knowledge of truths of reason is recalled from previous lives is a primitive form of rationalism[31]. But even if this

---

[29] (Leibniz and Loemker, 1969, p. 654)

[30] *Laches*

[31] The natural question is "Recollections of what?". How was experience in a previous life so different from what we have now that it could justify knowledge of necessity?

doctrine is put aside as a Platonic interpolation[32], the negativity of Socratic elenchus the relentless attempt to induce his interlocutors into being aware of their own ignorance is at odds with the positive philosophy advocated here. It is very tempting to read Socrates as holding to a latent mysticism according to which human language can never express the truth concerning such divine matters as the nature of justice or virtue. On this view Laches' bluff stolidity and Euthyphro's incipient sycophancy are not merely reflected by the doctrines they espouse but are consequences of such espousal, deformations of character from which Socrates is free, in part, becasue of his dissatisfaction with any attempt to express divine truths in language.

In both modern and ancient times many philosophers who have attempted to apply their technique outside academic circles have encountered the following obstacle: having argued their case convincingly and left their interlocutor without any plausible justification for their view on some disputed matter, their argument is dismissed, without refutation, without any atempt to engage with the substance of what has been said, but simply on the grounds that ."You're just good at arguing". Making the worse appear the better argument is how it was put in Socrates' day. On the traditional view of reasoning, such willful obstinacy is particularly egregious. Well brought up academics are taught to give up, or at least soften, their allegiance to any thesis they cannot justify. Of course their thesis may still be true for all that but without justification they have no reason for thinking so. Justify or abandon is the standard that applies in academic discussions, and it is easy to see everymans' willingness to flout this standard as intellectual laziness.

Academic practice is (ideally) motivated by a desire to get the underlying theory right, and not merely to inculcate the ability to make correct predictions. If our understanding of philosophy is correct, the academic interest is not in matching some super-empirical facts, but in making sure that the fit with the ordinary empirical evidence is achieved by explicit means. When such a project is tacitly shared by all parties, the 'justify or surrender' standard serves a common purpose. It is a luxury other professions often cannot afford. We all want surgeons at work to be more concerned with getting their patients to survive than trying to state the principles by which they operate. If some part of those principle remains cloaked in ambiguity, it may still be taught by example and learnt through practice. A practicioner who learnt her craft largely through hands-on experience may be wise to stick to his practices even when she cannot justify why they work.

Thus laymen are sometimes rightfully scornful of academic criticism of their beliefs, feeling their occasional inability to do justice to their understanding

---

[32]See (Vlastos, 1991)

in words should not be held against the soundness of their predictions. The
feeling of having one's words twisted, familiar to any who engage in academic
disputes (and particularly in philosophy) is on our view an awareness of some
ambiguous sentences in one's theory being subject to a deliberately contrary
equivocation so as to produce a patently false system. The layman feels that he
has been cheated of his conclusion, the soundness of his judgement impugned
merely because he cannot play the academicians' word game, while for her part
the academic feels a distrust for the inexplicit. A surgeon may know that a cut
just so deep will save the patients life just as well as anything may be known
despite being unable to account for that knowledge. If his opinion cannot be
justified that should not, of itself, reduce our confidence in his surgical skills,
though we will not trust his students so much until they too have a comparable
range of experience, for there is something to his surgical practice that he has
been unable to express.

To return at last to the slave boy from the introduction and answer the
question of when arguments which are neither based upon contrary observation
nor warrant revision of his empirical expectations nonetheless justify him his
adoption of their conclusions — When arguments proceed from principles he
without which he is unable to derive his empirical predictions by forms of in-
ference without which he is unable to derive his empirical predictions he ought
to adopt their conclusions on pain of failing to make the process by which he
forms his beliefs manifest in language.

Bibliography

Ackermann, R., 1963, 'A Neglected Proposal concerning Simplicity', *Philosophy of Science* **30(3)**, 228

Allen, R. E., 1985, *Greek Philosophy: Thales to Aristotle*, The Free Press

Armstrong, D. (ed.), 1965, *Berkeley's philosophical writings*, Macmillan Publishing New York

Ayer, A., 1936, *Language, Truth and Logic*, Penguin, 1971 edition

Ayer, A., 1940, *The Foundations of Empirical Knowledge*, Macmillan, 1964 edition

Ayer, A., 1956, *The Problem of Knowledge*, Penguin books

Ayer, A., 1973, *The Central Questions of Philosophy*, Penguin books

Barcan Marcus, R., 1968, 'Essential Attribution', *The Journal of Philosophy* 7

Barrett, P. and Freeman, R., 1987, *The Works of Charles Darwin.*, New York University Press

Berson, D., 2003, 'Strange vision: ganglion cells as circadian photoreceptors', *Trends in Neurosciences* **26(6)**, 314–320

Bertrand, R., 1959, *The Problems of Philosophy*, Oxford University Press

Brandom, R. B., 1994, *Making It Explicit: reasoning, representing, and discursive commitment*, Harvard University Press, 1998 edition

Burgess, J., 1998, 'Quinus ab omni naevo vindicatus', in A. Kazmi (ed.), *Meaning and Reference*, 25–66, University of Calgary Press, Canadian Journal of Philosophy Supplementary volumne 23

Carnap, R., 1954, *Introduction to Symbolic Logic and Its Applications*, Dover Publications, 1958 edition

Chalmers, A., 1976, *What is this thing called science?*, University of Queensland Press

Chomsky, N., 1959, 'A review of B.F. Skinner's Verbal Behavior', *Language* **35(1)**, 26–58

Craig, W., 1956, 'Replacement of Auxiliary Expressions', *The POhilosophical Review* **65(1)**, 38–55

Davidson, D., 1984a, *Inquiries into Truth and Interpretation*, Clarendon Press

Davidson, D., 1984b, 'On the Very Idea of a Conceptual Scheme', in *Inquiries into Truth, and Interpretation*, Clarendon Press, Oxford

Davidson, D., 1990, 'Meaning Truth and Evidence', in R. Garret, R. & Gibson (ed.), *Perspectives on Quine*, 68–79, Oxford: Blackwell

Dennett, D., 1987, *The Intentional Stance*, MIT Press

Dretske, F., 1981, *Knowledge and the flow of information*, MIT Press Mass

Einstein, A., 1956, *The Meaning of Relativity, 5 e ed*

Elder, C., 1998, 'Essential Properties and Coinciding Objects', *Philosophy and Phenomenological Research* **58**, 317–332

Elton, M., 2003, *Daniel Dennett: reconciling science and our self-conception*, Key Contemporary Thinkers, Polity Press

Etchemendy, J., 1990, *The concept of logical consequence*, Harvard University Press

Field, H., 1994a, 'Deflationist Views of Meaning and Content', *Mind* **103(411)**, 250–285

Field, H., 1994b, 'Disquotational Truth and Factually Defective Discourse', *Philosophical Review* **103(3)**, 405–462

Fodor, J. A. & Chihara, C., 1965, 'Operationalism and Ordinary Language: A critique of Wittgenstein', *American Philosophical Quarterly* **2(4)**, 281–295, reprinted in (Fodor, 1986)

Fodor, J. & Lepore, E., 1992a, *Holism a Shoppers Guide*, T.J. Press Ltd.

Fodor, J., 1975, *The Language of Thought*, The Language and Thought series, Harvard University Press

Fodor, J., 1986, *Representations*, MIT Press

Fodor, J., 1987, *Psychosemantics*, MIT Press

Fodor, J. and Society, B. P., 1987, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, MIT Press

Fodor, J. A., 1992b, *A Theory of Content and other essays*, MIT Press

Fodor, J. A., 1998, *Concepts: where cognitive science went wrong*, Clarendon Press: Oxford

Forster, M. and Sober, E., 1994, 'How to Tell When Simpler, More Unified, or Less Ad Hoc Theories will Provide More Accurate Predictions', *The British Journal for the Philosophy of Science* **45(1)**, 1–35

Gemes, K., 1998, 'Hypothetico-deductivism: The current state of play; the criterion of empirical significance: endgame', *Erkenntnis* **49(1)**, 1–20

George, A., 2000, 'On Washing the Fur Without Wetting It: Quine, Carnap, and Analyticity', *Mind* **109(433)**, 1–24

Gomez-Torrente, M., 1996, 'Tarski on Logical Consequence', *Notre Dame Journal of Formal Logic* 37(1)

Goodman, N., 1955, 'Axiomatic Measurement of Simplicity', *The Journal of*

Philosophy **52(24)**, 702–722

Goodman, N., 1959, 'Recent Developments inthe Theory of Simplicity', *Philosophy and Phenomenological Research* **19(4)**, 429–446

Goodman, N., 1961, 'Safety, Strength, Simplicity', *Philosophy of Science* **28(2)**, 150

Goodman, N., 1983, *Fact, fiction, and Forecast, 4th edn.*

Grice, H., 1957, 'Meaning', *Philosophical Review* **66**, 251–259

Gulick, R. V., Spring 2008, 'Consciousness', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*

Harman, G., 1967, 'Quine on Meaning and Existence', *Review of Metaphysics* **21**, 124–151

Harman, G., 1993, 'Meaning Holism Defended', in E. Fodor, J. & Lepore (ed.), *Holism: A Consumer Update*, Rodopi

Heathcote, A., 2002, *The Art of Necessity*, A PDF version of this work was available online when Dr. Heathcote was a member of faculty at the University of Sydney, but it has since disappeared

Hempel, C., 1950, 'Problems and Changes in the Empiricist Criterion of Meaning', in A. Ayer (ed.), *Logical Positivism*, London: Allen and Unwin, 1959 edition

Higgs, H., Palgrave, R., et al., 1963, *Palgrave's Dictionary of Political Economy*, AM Kelley

Hill, C., 2002, *Thought and World*, Cambridge University Press

Hillman, D., 1962, 'The Measurement of Simplicity', *Philosophy of Science* **29(3)**, 225

Hintikka, J., 1999, 'Quine's ultimate presuppositions', *Theoria* **65**, 3–24

Hintikka, J., 2002, 'Hyperclassical logic (aka IF logic) and its implications for logical theory', *Bulletin of Symbolic Logic* **8(3)**, 404–423

Hintikka, J., 2004, 'Independence-friendly logic and axiomatic set theory', *Annals of Pure and Applied Logic* **126(1-3)**, 313–333

Horwich, P., 1990, *Truth*, Basil Blackwell

Hume, D., 1739, *A Treatise of Human Nature*, Oxford Press, 1990 edition, ed., Selby-Bigge, L.A.

Husserl, E., 1964, *The Idea of Phenomenology*, Martinus Nijhoff, Translated by William P. Alston and George Nakhnikian

Kment, B., 2006, 'Counterfactuals and Explanation', *Mind* **115(458)**, 261–310

Kripke, S., 1972, *Naming and Necessity*, Harvard University Press, 1980 edition

Kripke, S., 1982, *Wittgenstein on rules and private language: an elementary exposition*, Blackwell Oxford

Laing, R., 1967, *The Politics of Experience*, Routledge & Kegan Paul

Leibniz, G. and Loemker, L., 1969, 'Gottfried Wilhelm Leibniz: Philosophical

Papers and Letters', *Edited and translated by Leroy E. Loemker. Dordrecht: Reidel*

Lewis, D., 1970, 'How to Define Theoretical Terms', *The Journal of Philosophy* **67(13)**, 427–446

Lewis, D., 1972, 'Psychophysical and Theoretical Identifications', *Australasion Journal of Philosophy* **50(3)**, 249–258

Lewis, D., 1998, 'Possible Worlds', *Contemporary Readings in the Foundations of Metaphysics*

Lewis, D., 2001, *Counterfactuals*, Blackwell Publishers

Locke, J., 1690, *An Essay Concerning Human Understanding*, Dent: London, 1961 everyman edition

MacFarlane, N., 2003, 'Future Contingents and Relative Truth', *The Philosophical Quarterly* **53(212)**, 321–36

Margolis, E. & Laurence, S. (ed.), 1999, *Concepts: Core Readings*, MIT Press

Marx, K. and Engels, F., 1975, *Karl Marx, Frederick Engels: Collected Works*, Lawrence & Wishart

McDermott, M., 1992, *From a Psychological Point of View*, unpublished manuscript

McDermott, M., 1996, *Quine on Science and Metaphysics*, unpublished manuscript

McDermott, M., 1999, *Quine's Holism*, printed copy

Millikan, R., 1984, *Language, Thought, and Other Biological Categories*, MIT Press

Nagel, T., 1997, *The Last Word*, Oxford University Press

Passmore, J., 1968, *A Hundred Years of Philosophy*, Penguin Books

Philips, M., 1986, 'The Justification of Punishment and the Justification of Political Authority', *Law and Philosophy* **5(3)**, 393–416

Putnam, H., 1967, 'Psychological Predicates', *Art, Mind and Religion* 37–48

Putnam, H., 1979, 'The Meaning of 'Meaning'', in K. Gunderson (ed.), *Language, Mind, and Knowledge*, Vol. 7 of *Minnesota Studies in the Philosophy of Science*, University of Minnesota Press

Putnam, H., 1981, 'Two Conceptions of Rationality', in *Reason, Truth and History*, Chapt. 5, Cambridge University Press

Putnam, H., 1983a, 'Computational psychology and interpretation theory', in *Realism and Reason*, Vol. 3, Chapt. 8, 139–154, Cambridge University Press

Putnam, H., 1983b, 'Reference and Truth', in *Realism and Reason*, Vol. 3 of *Philosophical Papers*, Chapt. 4, 69–86, Cambridge University Press

Putnam, H., 1983c, ''Two Dogmas' revisited', in *Realism and Reason*, Vol. 3 of *Philosophical Papers*, Chapt. 5, 87–97, Cambridge University Press

Putnam, H., 1983d, 'Why Reason can't be naturalized', in *Realism and Reason*,

Vol. 3 of *Philosophical Papers*, Chapt. 13, 229–247, Cambridge University Press

Quine, W., 1952, 'On omega-Inconsistency and a So-called Axiom of Infinity', in *Selected Logic Papers*, Chapt. 9, 114–120, Harvard University Press

Quine, W., 1953, *From a Logical Point of View*, Harvard University Press, 1980 edition

Quine, W., 1954, 'Logic, Symbolic', in *Selected Logic Papers*, Chapt. 2, 37–51, Harvard University Press

Quine, W., 1960, *Word and Object*, MIT Press

Quine, W., 1966, *The Ways of Paradox and other essays*, Random House

Quine, W., 1969, *Ontological Relativity and other essays*, Columbia University Press

Quine, W., 1975, 'On Empirically Equivalent Systems of the World', *Erkenntnis* **9**, 313–328

Quine, W., 1990a, *Pursuit of Truth*, Harvard University Press

Quine, W., 1990b, *Quiddities*, Penguin

Quine, W., 1990c, *Roots of Reference*, Open Court

Quine, W., 1991, 'Immanence and Validity', in *Selected Logic Papers*, Chapt. 25, 244–250, Harvard University Press

Quine, W., 1995, 'Whitehead and the Rise of Modern Logic', in *Selected Logic Papers*, Chapt. 1, 3–36, Harvard University Press

Reibetanz, S., 1998, 'A Problem for the Doctrine of Double Effect', *Proceedings of the Aristotelian Society* **98**, 217–223

Richardson, A., 1997, 'Two Dogmas about Logical Empiricism: Carnap and Quine on Logic, Epistemology, and Empiricism', *Philosophical Topics* **25(2)**, 145–167

Rosch, E., 1978, 'Principles of Categorization', in B. Rosch, E. & Lloyd (ed.), *Cognition and Castegorization*, Lawrence Erlbaum, Reprinted in Margolis & Laurence 1999

Rudner, R., 1961, 'An Introduction to Simplicity', *Philosophy of Science* **28(2)**, 109

Russell, B., 1972, *The Philosophy of Logical Atomism*, Open Court, 1985 edition

Shapiro, S., 2003, 'The Guru, the Logician, and the Deflationist: Truth and Logical Consequence', *Nous* **37(1)**, 113–132

Sher, G., 1996, 'Did Tarski commit "Tarski's fallacy"', *The Journal of Symbolic Logic* **61(2)**, 653–686

Sher, G., 1999, 'Is there a place for philosophy in Quine's theory', *The Journal of Philosophy* **96(10)**, 491–524

Sipser, M., 1997, *Introduction to the Theory of Computation*, PWS Publishing Company

Sober, E., 1981, 'The Principle of Parsimony', *The British Journal for the Philosophy of Science* **32(2)**, 145–156

Sober, E., 2002, 'Instrumentalism, Parsimony, and the Akaike Framework', *Philosophy of Science* **69(S3)**, 112–123

Sober, E., 2006, 'Let's razor Ockham's Razor', *Philosophy of Science: An Anthology*

Sober, E., 2008, *Evidence and Evolution: The Logic Behind the Science*, Cambridge University Press

Stroud, S., 2006, 'Epistemic Partiality in Friendship', *Ethics* **116(3)**, 498–524

Thomason, R. and Stalnaker, R., 1968, 'Modality and reference', *Nous* **2(4)**, 359–372

Vlastos, G., 1991, *Socrates, Ironist and Moral Philosopher*, Cornell University Press

Wasserman, R., 2004, 'The Constitution Question', *Nous* **38(4)**, 693–710

Williams, M., 1999, 'Meaning and Deflationary Truth', *The Journal of Philosophy* **96(11)**, 545–564

Winnie, J., 1967, 'The Implicit Definition of Theoretical Terms', *British Journal for the Philosophy of Science* **18**, 223–229

Wittgenstein, L., 1922, *Tractatus Logico-Philosophicus*, Routledge, 1974 edition

Wittgenstein, L., 1953, *Philosophical Investigations*, Blackwell, 1958 edition

Wittgenstein, L., 1958, *The Blue and Brown Books*, Basil Blackwell Oxford, 1964 edition

Wolenski, J., 2001, 'In Defense of the Semantic Definition of Truth', *Synthese* **126**, 67–90

Wrinch, D. and Jeffreys, H., 1921, 'On certain fundamental principles of scientific inquiry', *Phil. Mag*